



---

# Independent agents in branching time:

Towards a unified framework for reasoning about multiagent systems

---

presented and publicly defended on the 20th of July 2007 at Toulouse  
as partial fulfillment of the requirements for the degrees of

Dottore di Ricerca in Informatica e Telecomunicazioni  
dell'Università degli studi di Trento

*and*

Docteur en Informatique et Télécommunications  
de l'Université Paul Sabatier – Toulouse III

*by* Nicolas Troquard

Reviewers

John  
Pierre  
Wiebe

Horty  
Marquis  
van der Hoek

Examiners

Jean-Paul  
Jan  
Robert  
Claudio

Bodeveix  
Broersen  
Demolombe  
Masolo

Thesis advisors

Andreas  
Laure

Herzig  
Vieu



---

# Independent agents in branching time:

Towards a unified framework for reasoning about multiagent systems

---

*Agents indépendants dans le temps ramifié : vers un cadre unifié de raisonnement sur les systèmes multi-agents*

*Agenti indipendenti nel tempo ramificato : verso un ambiente unificato per il ragionamento sui sistemi multi-agenti*



---

# Abstracts

## Independent agents in branching time

The work presented in this thesis is a multidisciplinary study of the notion of *agency*. We build new formal approaches starting on the literature of agency in philosophy of action, game theory or computer science.

Belnap and Perloff's STIT theory is our frame of experimentation. This is a logic that stems from philosophy of action based on the observation that an action can be identified with what it brings about. In this tradition, the sentence "Ishmael sails on board the Pequod" will be paraphrased by "Ishmael *sees to it that* Ishmael sails on board the Pequod".

Our first contribution is to simplify the axiomatics of a version of the logic restrained to individual agency and without temporal aspects. This allows us to simplify the semantics of STIT as well as to discover a link with product logics. We establish the NEXPTIME-completeness of the problem of satisfiability. We capitalize on the simplifications and extend the axiomatization to coalitional actions. We show that we can embed Coalition Logic in the resulting logic. We also provide an epistemic extension and use it to tackle the problem of epistemically uniform strategies.

Then we study further the temporal aspects of agency. We first do it by way of a logic combining STIT with a dynamic logic providing actions with duration, that can be deliberately continued or aborted along time. We then give an embedding of Alternating-time Temporal Logic in a slightly adapted strategic STIT logic. Having developed a neat understanding of relevant structures of agency, we propose a fine-grained ontology of action and agency.

## Agents indépendants dans le temps ramifié

Le travail présenté dans cette thèse est une étude multidisciplinaire de la notion de *réalisation*. Nous construisons de nouvelles approches formelles à partir de la littérature de la réalisation en philosophie de l'action, théorie des jeux ou informatique.

La théorie du STIT de Belnap et Perloff est notre cadre d'expérimentation. C'est une logique issue de la philosophie de l'action basée sur l'observation qu'une action peut être identifiée avec les effets qu'elle cause. Selon cette tradition, la phrase "Ishmael navigue à bord de la Pequod" sera paraphrasée par "Ishmael *fait en sorte que* Ishmael navigue à bord de la Pequod".

Notre première contribution est une simplification de l'axiomatique d'une version de la logique restreinte aux actions d'individus et sans aspects temporels. Cela nous permet de simplifier la sémantique du STIT ainsi que de découvrir un lien avec les logiques produits. Nous établissons que le problème de satisfiabilité est NEXPTIME-complet. Nous tirons parti des simplifications et étendons l'axiomatisation aux actions de coalitions. Nous montrons que nous pouvons simuler Coalition Logic dans la logique résultante. Nous fournissons également une extension épistémique et l'utilisons pour traiter le problème des stratégies épistémiquement uniformes.

Ensuite, nous étudions plus en profondeur les aspects temporels de la réalisation. Nous faisons d'abord cela en combinant STIT avec une logique dynamique avec des actions avec durée, qui peuvent être délibérément continuées ou abandonnées durant leur exécution. Nous donnons ensuite une simulation de Alternating-time Temporal Logic dans une version adaptée de la logique du STIT stratégique. Ayant alors une nette compréhension des structures de la notion de réalisation, nous proposons une ontologie précise de l'action et de la réalisation.

## Agenti indipendenti nel tempo ramificato

Il lavoro presentato in questa tesi è uno studio pluridisciplinare della nozione di *realizzazione*. Costruiamo nuovi approcci formali a partire dalla letteratura della realizzazione in filosofia dell'azione, teoria dei giochi o informatica.

La teoria STIT di Belnap e Perloff è il nostro quadro di sperimentazione. È una logica sviluppata in filosofia dell'azione. Tale logica è basata sull'osservazione che un'azione può essere identificata con gli effetti che causa. Secondo questa tradizione, la frase "Ishmael naviga a bordo del Pequod" sarà parafrasata da "Ishmael *fa in modo che* Ishmael naviga a bordo del Pequod".

Il nostro primo contributo è un semplificazione dell'assiomatica di una versione della logica limitata alle azioni individuali e senza aspetti temporali. Cioè possiamo semplificare la semantica STIT e scoprire un legame

con le logiche dei prodotti. Stabiliamo che il problema di soddisfiabilità è NEXPTIME-completo. Traiamo vantaggio dalle semplificazioni ed estendiamo l'assiomatizzazione alle azioni di coalizioni. Mostriamo che possiamo simulare Coalition Logic nella logica risultante. Forniamo anche un'estensione epistemica e la utilizziamo per trattare il problema delle strategie epistemicamente uniformi.

In seguito, studiamo più a fondo gli aspetti temporali della realizzazione. Facciamo inizialmente ciò combinando STIT con una logica dinamica con azioni con durata, che possono essere deliberatamente continuate o abbandonate durante la loro esecuzione. Diamo in seguito una simulazione di Alternating-time Temporal Logic in una versione adeguata della logica STIT strategica. Avendo allora una netta comprensione delle strutture della nozione di realizzazione, proponiamo una ontologia precisa dell'azione e della realizzazione.



---

# Acknowledgments

First of all, I am grateful to the “Fondo provinciale per i progetti di ricerca” of the Provincia Autonoma of Trento for funding my research. Thanks are also due to the University of Toulouse and the organizing committees of AAMAS’06 and KR’06 for offering me travel grants.

My acknowledgment then goes to Andreas Herzig and Laure Vieu for their friendship and awesome supervision. They offered me the opportunity to do this PhD, giving me a large amount of autonomy and liberty while being always ready for discussions. Thank you for those three years and all we learned together in philosophy of action and logic. Thank you for the exciting travels and in some sense for my new skills in English and Italian.

I also owe special thanks to Jan Broersen who tried to teach me how to tell stories. Thank you for the invitation at Utrecht University where after almost two years of working over emails, we finally have had the opportunity to stand together in front of the same whiteboard.

I am indebted to John Horty, Pierre Marquis and Wiebe van der Hoek for examining my thesis and providing both very kind comments and crucial remarks on it.

I thank Jean-Paul Bodeveix, Jan Broersen, Robert Demolombe, Andreas Herzig, John Horty, Claudio Masolo, Wiebe van der Hoek and Laure Vieu for their entertaining questions during the defense of this thesis.

This thesis is based on several communications published or not, most of them coauthored. It turns out that some paragraphs are even not written by me, but of course the true authors could not be responsible for the mistakes committed throughout this thesis. (I cannot imagine there are none.) I warmly thank them for letting me use our material here: Philippe Balbiani, Jan Broersen, Cristiano Castelfranchi, Olivier Gasquet, Andreas Herzig, Emiliano Lorini, François Schwarzentruher, Robert Trypuz and Laure Vieu. I am glad and proud to have you in my circle.

I also thank those who made my work richer in one way or another. People with who I have had discussions in person or over email. They

gave me insights or simply encouragements on some aspects related either directly to my research or to the academic world in general.

I also thank those who made my work easier and enjoyable in more ways than one. Those with whom I shared lunches and coffees at Toulouse, Trento or Utrecht. They are those with whom I shared beer, wine and parties at Trento, Toulouse, Hakodate, Windermere, Amsterdam, Rome, Brussels...

I could not be exhaustive in their enumeration, hence, instead of committing an irrevocable omission, I prefer to refrain my tentative of naming them all.

To the 'Lémuriens' that used to live at the *258 avenue de Muret* during those last three years, you are easy to name by chronological order: Mathilde, Gaël, Virginie, Pere, Gaëlle, Markos, Vincent, Chiska<sup>1</sup>, Françoise, Hélène, Penny<sup>2</sup>, Élodie and Emelyne. You made my after-work cheerful with all the cooking, the dinners and drinks on our marvelous balcony on *la Garonne*. You also somewhat contributed in my training at managing a team.

My constant thought goes to my friends and family for having been there, even if sometimes I was not completely with you. Again, I will refrain from listing all your names but be certain, *this* is for you.

Toulouse, September 2007

---

<sup>1</sup>Who is actually a dog...

<sup>2</sup>Penny also corrected my fuzzy English on some parts of this memoir.

---

# Contents

<b>Abstracts</b>	<b>i</b>
Independent agents in branching time . . . . .	i
Agents indépendants dans le temps ramifié . . . . .	i
Agenti indipendenti nel tempo ramificato . . . . .	ii
<b>Acknowledgments</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Agency as a modality . . . . .	1
1.2 Agency in philosophy of action . . . . .	3
1.3 A pinch of Game Theory . . . . .	6
1.4 Logics of action in computer science . . . . .	8
1.5 Outline . . . . .	9
<b>2 The logic of “Seeing To It That”</b>	<b>13</b>
2.1 Generalities . . . . .	13
2.2 Deliberative STIT theories . . . . .	14
2.2.1 Models: rudiments . . . . .	14
2.2.2 Constraints on <i>Choice</i> . . . . .	15
2.2.2.1 Individual choice . . . . .	15
2.2.2.2 Group choice . . . . .	16
2.2.3 Truth conditions of operators . . . . .	16
2.3 Strategic ability . . . . .	18
2.4 Achievement stit . . . . .	20
<b>3 Meta-logical aspects of individual choice</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 Xu’s axioms for the individual Chellas STIT (CSTIT) . . . . .	24
3.2.1 Language . . . . .	24
3.2.2 Axiomatics . . . . .	25
3.3 An alternative axiomatics . . . . .	26
3.4 Historic necessity is superfluous in presence of two agents or more . . . . .	28
3.5 A simpler semantics . . . . .	30

3.6	Complexity . . . . .	32
3.6.1	Complexity of CSTIT . . . . .	32
3.6.2	Complexity of the deliberative STIT logic . . . . .	33
3.7	Conclusion . . . . .	34
3.8	Annex: Proofs . . . . .	35
<b>4</b>	<b>Logics of collective choice</b>	<b>41</b>
4.1	Introduction . . . . .	41
4.2	Coalition Logic . . . . .	42
4.2.1	Coalition model semantics . . . . .	43
4.2.2	Game semantics . . . . .	44
4.2.3	Axiomatization . . . . .	44
4.3	Extension of CSTIT to groups of agents ( $\mathcal{G}$ STIT) . . . . .	45
4.4	Coalitional choice plus discrete time . . . . .	48
4.4.1	Motivations . . . . .	48
4.4.2	Normal Simulation of Coalition Logic (NCL) . . . . .	49
4.5	STIT embraces Coalition Logic in the realm of normal modal logics . . . . .	51
4.6	Expressiveness . . . . .	54
4.7	Seeing to it under imperfect knowledge . . . . .	56
4.7.1	Epistemic NCL (ENCL) . . . . .	57
4.7.2	Reasoning about uniform strategies . . . . .	58
4.7.3	Discussion . . . . .	61
4.8	Concluding remarks . . . . .	61
<b>5</b>	<b>Agency in branching time</b>	<b>63</b>
5.1	Introduction . . . . .	63
5.2	Time in CSTIT . . . . .	64
5.2.1	Some temporal order remains... . . . .	64
5.2.2	... but does not capture fine-grained time . . . . .	65
5.2.3	Chellas's stit is the brute choice component of agency . . . . .	66
5.2.4	Causality in agency . . . . .	67
5.3	Measuring the length of an action . . . . .	68
5.3.1	A discrete STIT framework . . . . .	69
5.3.2	Duration of an activity. . . . .	72
5.3.3	Comments on Chellas's $\Delta_a\varphi$ . . . . .	72
5.3.3.1	Semantics of time and actional alternatives . . . . .	72
5.3.3.2	Chellas's stit is not $\Delta_a\varphi$ . . . . .	73
5.3.4	Choice vs. causality in branching time . . . . .	74
5.4	A modal view of actions with duration . . . . .	75
5.4.1	Motivation . . . . .	75

5.4.2	A modal logic for actions with duration . . . . .	76
5.4.3	Ontological justification and some validities . . . . .	78
5.4.3.1	Time . . . . .	78
5.4.3.2	Actions . . . . .	78
5.4.3.3	Continuation of an action, completed actions . . . . .	79
5.4.3.4	Not doing anything . . . . .	80
5.4.4	Choices and group agency: a new characterization of independence of agents . . . . .	80
5.5	Discussion . . . . .	82
<b>6</b>	<b>Alternating-time Temporal Logic vs. STIT</b>	<b>85</b>
6.1	Introduction . . . . .	85
6.2	Alternating-time Temporal Logic . . . . .	86
6.3	Formal properties of ATL . . . . .	89
6.3.1	Complexity . . . . .	89
6.3.2	Semantic equivalence results for ATL . . . . .	90
6.4	A convenient strategic stit operator of ability . . . . .	92
6.5	From ATL to STIT logic . . . . .	95
6.6	Discussion . . . . .	99
<b>7</b>	<b>Ontology of Agency and Action</b>	<b>105</b>
7.1	Introduction . . . . .	105
7.2	Motivations . . . . .	106
7.2.1	Why going first-order? . . . . .	106
7.2.2	Summary of formal aspects of STIT . . . . .	106
7.3	STIT Ontology of Agency – OntoSTIT . . . . .	108
7.3.1	A modal or an ontological approach? . . . . .	108
7.3.2	From STIT to OntoSTIT . . . . .	109
7.3.2.1	Language . . . . .	109
7.3.2.2	Characterization of primitive relations and categories . . . . .	111
7.3.3	Agency in OntoSTIT . . . . .	113
7.3.3.1	Agentive and causal gaps . . . . .	114
7.4	Towards an Ontology of Action – OntoSTIT+ . . . . .	117
7.4.1	Language. . . . .	118
7.4.2	Characterization of new universals . . . . .	119
7.4.3	Agency in OntoSTIT+ . . . . .	125
7.4.4	Understanding <i>PO</i> . . . . .	125
7.4.5	Expressivity . . . . .	128
7.4.5.1	Responsibility – filling the causal gap . . . . .	129
7.4.5.2	Filling the agentive gap . . . . .	129

---

<b>8 Conclusion and perspectives</b>	<b>133</b>
8.1 Summary . . . . .	133
8.2 Towards rationality . . . . .	134
8.3 Towards extensive games . . . . .	136
<b>References</b>	<b>139</b>

# 1

---

## Introduction

This dissertation is about the formal structures related to agency, rationality left aside. More specifically we concentrate our analysis on the concept of choice and its manifestation over time. We essentially aim at pushing towards a uniform framework for agency.

The approach is multidisciplinary, building new formal approaches upon the literature of agency in philosophy of action, game theory or computer science. Our first intention is to study rich frameworks, capable of balancing temporal reasoning with notions of agency. Such structures are generally interesting for being powerful and complex. Their complexity nevertheless often leads to a partial understanding about them and constitutes an obstacle for sagacity. Authors are less likely to establish connections between complex frameworks and thus to transfer results. As a consequence it prevents two theories that are intuitively concerned by the same notions to evolve in symbiose. For this reason, we aim to offer a contribution to the coherence of the field, establishing several intra- and inter-area parallels between formalisms. Pushing further, we are able to propose new formal tools that we believe useful and elegant.

In this introduction, we give the landscape where this dissertation takes place. We present informally the models relevant to our agenda.

### 1.1 Agency as a modality

In this dissertation we mean by *agent*, an individual that makes choices or acts over time. It is not restricted to persons or intentional agents and could equally be applied to processes making random choices. Actions are thus idealized in a way that ignores any mental state. If  $\mathcal{Agt}$  is the collection of all individual agents, we call a *coalition* any subset of  $\mathcal{Agt}$ .

Origins of agency considered as a modality go back to St Anselm of Canterbury. He suggested that acting was adequately captured by what an agent brings about. He suggested that a verbal group like “killing directly” could be reformulated as “directly bringing it about that the victim is dead”. We direct an interested reader to [Hen67] and [Dou76] for more details on St Anselm’s mediaeval logic. We preferably overview more recent accounts that without a doubt will fit better in the scope of a work in computer science. Indeed, this approach gave birth to a variety of logics of action over the past fifty years that have inherited the particularity that they do not refer to the action itself but rather to its resulting state of affairs.

They are logics whose main operator reads “the agent  $a$  brings it about that  $\varphi$ ”. Several principles that such an operator may verify have been discussed in the literature. We present some of those principles in Figure 1.1. To encompass modalities of agency in general, we use  $\nabla_a\varphi$  as a general notation to present them.

(M)	$\nabla_a(\varphi \wedge \psi) \rightarrow (\nabla_a\varphi \wedge \nabla_a\psi)$
(C)	$(\nabla_a\varphi \wedge \nabla_a\psi) \rightarrow \nabla_a(\varphi \wedge \psi)$
(N)	$\nabla_a\top$
(No)	$\neg\nabla_a\top$
(T)	$\nabla_a\varphi \rightarrow \varphi$
(RE)	if $\varphi \leftrightarrow \psi$ then $\nabla_a\varphi \leftrightarrow \nabla_a\psi$

**Figure 1.1:** Some principles of agency operators.  $\nabla_a\varphi$  is a general notation and reads “the agent  $a$  brings it about that  $\varphi$ ”.

Assuming one or another principle is committing the theory of agency to some interpretation. (N) and (No) are inconsistent and express contradictory properties of the operator. Preferring (N) over (No) is accepting that an agent can be agentive for something that is settled. Almost every logic of agency, if not all, take one or the other as an axiom. (M) forces that if an agent ensures a state of affairs, it also ensures its parts. (C) is the inverse implication: an agent ensures all the parts of a state of affairs only if it ensures the whole. (T) confers to agency the characteristic of being *successful*: if an agent is agentive for  $\varphi$  then  $\varphi$  holds now. “ $a$  ensures  $\varphi$ ” hence has to be understood as “ $a$  has performed an action that caused  $\varphi$  to be true now”. A logic of agency accepting this axiom makes its operator mark the result of an action. (RE) is a rule of inference stating that by

bringing about a state of affairs, an agent brings about every equivalent state of affairs.

## 1.2 Agency in philosophy of action

**Chellas.** In [Che69], Brian Chellas proposes what constitutes the first semantics of a logic of action and follows the paradigm of St Anselm. Chellas operator  $\Delta_a\varphi$  reads “agent  $a$  sees to it that  $\varphi$ ”. It is interpreted in terms of agents, times, histories and for each agent, relations for ‘actional alternatives’. These latter permit to collect together histories that an agent cannot single out at a given time. An agent is agentive or responsible for  $\varphi$  if at the previous time it triggered an action that made  $\varphi$  certain at the current time on every history in actional alternatives. The logic of  $\Delta_a\varphi$  validates every principle listed above except (No) and is then a normal logic.

We present Chellas’s semantics in more details in Section 5.3.3.

**Pörn.** Following Stig Kanger [Kan72], Ingmar Pörn also designed a language of agency along Anselmian lines [Pör70, Pör77]. It has particularly gained notoriety among authors interested by institutional power [JS96, Roy00, CP01]. The idea is to combine two normal modal operators to create a non-normal one. Semantics is in terms of Kripke models providing two relations  $R_D$  and  $R_{D'}$  over possible worlds.  $R_D$  is assumed reflexive and transitive and  $R_{D'}$  is given to be irreflexive and serial. The standard modal operators of necessity corresponding to the relations are  $D_a\varphi$  reading “it is necessary for something which  $a$  does that  $\varphi$ ” and  $D'_a\varphi$  reading “but for  $a$ ’s action it would not be the case that  $\varphi$ ”. Pörn defines  $D_a\varphi$  to be true at a world  $w$  if  $\varphi$  is true at every *hypothetical situation* (alias possible world) where agent  $a$  “does at least as much as he does in  $w$ ” [Pör77, p. 5].  $D'_a\varphi$  is true in  $w$  if  $\neg\varphi$  is true in every hypothetical situation  $w'$  such that “the opposite of everything  $a$  does in  $w$  is the case in  $w'$ ” [Pör77, p. 6].

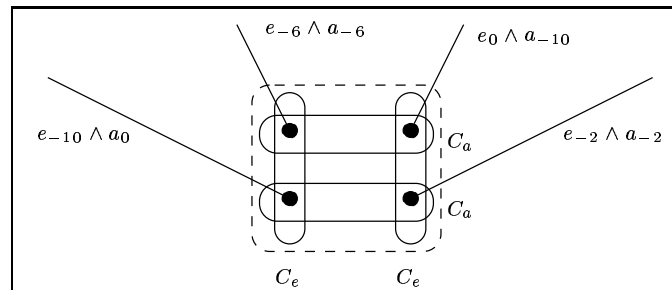
Pörn’s operator of agency is then defined by  $E_a\varphi \triangleq D_a\varphi \wedge \neg D'_a\neg\varphi$  and reads “the agent  $a$  brings it about that  $\varphi$ ”. Non-normality of the operator  $E$  makes it likely that it could have interesting properties of agency. In particular, it obeys the principles (M), (No), (T) and (RE).

We do not go further in the presentation of this framework that does not assume specific interactions between agents. It is also considered complicated and subject to criticism. Segerberg says “the intuitive significance of this semantics is not altogether clear” [Seg92, p. 368]. Horgan reviewed

Pörn's book and claimed that "one problem with the proposed semantics is that 'doing at least as much as' he does in [a world], and the notion of an agent doing 'the opposite' of everything he does in [a world], are of dubious intelligibility without substantial further elucidation, and Pörn offers none." ([Hor79, p. 310]).

Belnap on the contrary writes that Pörn's explanations are "the most detailed working out of the modal logic of agency as based on binary relational semantics" [Bel91, p. 784]. He rather concludes that the semantics proper is problematic.

**Belnap and others.** In [BP88], Nuel Belnap and Michael Perloff introduce the logic of the *achievement stit* operator [ $a \text{ stit} : \varphi$ ]. This modal formula is true if there is a moment in the past where  $a$  made a choice that ensured that  $\varphi$  would be true now, and could have chosen otherwise. Its motivation is primarily linguistic and given to test agentiveness of a sentence. But what distinguishes STIT from the other proposals is its flourishing semantics that shares features with Chellas's one, although more grounded in philosophy of action. In particular, while 'actional alternatives' in Chellas's account are shallowly specified, the choices of agents are strictly characterized in STIT.



**Figure 1.2:** A STIT model.  $C_a$  are classes of choice of Abelard, and  $C_e$  those of Eloise. The four 'dots' correspond to four histories passing through a same moment. The dotted line represents the extension of the moment and limits the possible outcomes at that moment. Straight lines represent the histories and suggest that time continues.

We can represent the information contained in a moment as in Figure 1.2. Abelard, the agent  $a$ , has two alternative choices. He can choose the outcome to be either  $e_{-6} \wedge a_{-6}$  or  $e_0 \wedge a_{-10}$  (if he does so, he ensures one of those outcomes) or he can choose it to be  $e_{-10} \wedge a_0$  or  $e_{-2} \wedge a_{-2}$ . Here, indeterminism depends on the choice Eloise will adopt. Eloise, agent  $e$ , can also choose twofold. She has the choice of forcing  $e_{-6} \wedge a_{-6}$  or  $e_{-10} \wedge a_0$

and the choice of forcing  $e_0 \wedge a_{10}$  or  $e_{-2} \wedge a_{-2}$ . Here, every choice of the two individuals has always two possible outcomes, and as a consequence, none of them can determine a unique outcome at that very moment. The eventual outcome is the one that is in the intersection of Eloise's actual choice with Abelard's. In this scenario the coalition formed by Eloise and Abelard (the *grand* coalition composed of every agent of the system) can force every single outcome. (This is not always the case.)

Segerberg and Chellas show strong support for the theory of the achievement stit.

Belnap and Perloff's "work is probably one of the two most promising avenues of research in current logic of action."<sup>1</sup> [Seg92, p. 374]

Belnap and Perloff's "theories of agency are complex, fascinating, and illuminating – without a doubt the most subtle and sophisticated proposals of their kind to date." [Che92]

However, Chellas is right in saying that the logic of achievement stit is complex. We know some mathematical difficulties about it, and in a sense it does not present an obvious immediate applicability in computer science or artificial intelligence. It was perhaps what John Horty thought too, when he started to investigate operators of agency able to mix adequately with deontic aspects and without yielding an over-engineered framework, that unfortunately often turns a beautiful idea into a 'logician's nightmare'.

Taking inspiration from seminal work in philosophy of action, he proposed the STIT theories the most discussed by 'non-philosopher' authors in logics for artificial intelligence: the *deliberative STIT theories*. They are the logics of two operators of agency. The *deliberative stit* [ $a\ dstit : \varphi$ ] stems from Franz von Kutschera's work [vK86]. It is given to stand for "a deliberately sees to it that  $\varphi$ " and its truth value requires a counterfactual condition for  $a$  being agentive for  $\varphi$ , that is, that  $a$  could act otherwise. Belnap et al. grant this condition a particular importance for agency, and so does Pörn as we have seen. An agent should not be agentive for something inevitable. On the contrary, Chellas argues the opposite in [Che69] and insists in [Che92], giving an interesting criticism of Belnap and Perloff's early account of achievement stit. His main claim is that when an agent

---

<sup>1</sup>To Segerberg's view, the other one is Pratt's Dynamic Logic [Pra76], originally designed to explain program verification and Hoare logic.

sees to it that something holds, it also sees to it that every logical consequence holds and hence also every tautology.<sup>2</sup> The second operator of agency in the deliberative STIT theories is then named *Chellas's stit* [ $a\text{ cstit} : \varphi$ ]. It corresponds to the deliberative stit without the negative condition.

We take Chellas's stit operator as central in a first time in this dissertation, as Horty did in [Hor01]. We then make a move to more fine-grained notions of agency in a second time, putting the light on some issues with this operator. In particular, it would be unfortunate to confuse the reader after this brief introduction to it. Indeed, one must not misconceive it with regard to Chellas's operator  $\Delta_a\varphi$ . Similarly to Chellas's  $\Delta_a\varphi$ , there is no counterfactual requirement in [ $a\text{ cstit} : \varphi$ ]. However, [ $a\text{ cstit} : \varphi$ ] is hardly an operator for causality in agency as  $\Delta_a\varphi$  is, and [ $a\text{ astit} : \varphi$ ] too. There is a slight but still important difference that we will observe in a discrete STIT structure. Roughly, Chellas's stit is evaluated at the moment of choice while  $\Delta_a\varphi$  is evaluated at the very *next* moment. Such a next moment is impossible to apprehend in the original STIT semantics.

An alternative name for Chellas's stit could be "choice stit": it represents what we call *brute choice* or *material choice*. Brute choice has to be understood as the ontological object containing the information of a choice and would just be a component of *rational choice*. A brute choice is simply a set of continuations that an agent (or a group of agents) has chosen or can choose for some reasons that are not part of the description.

### 1.3 A pinch of Game Theory

If there is a misconception of the Chellas stit operator, it does not mean that it does not prove itself to be useful. Horty and Belnap write that it is "simpler and for certain purposes more natural as an analysis of agency". One claim of this dissertation is that it is useful and more natural for an analysis of the strategic behaviour of agents and we support it by providing formal tools for modeling the concept of *uniform strategies*. In game theory [OR94], an agent is said to have a *strategy* for ensuring something when it has *objectively* the *ability* of doing it. A uniform strategy is strengthened by its knowledge of how to achieve it.

---

<sup>2</sup>The question of the closure of agency under logical consequences is a debated question in the philosophical literature we cite in this dissertation. However, we prefer not to enter more in the discussion.

A natural reading of  $[a \text{ cstit} : \varphi]$  should be “ $a$  chooses such that  $\varphi$ ”. Imitating the vocabulary of agency in branching time, we reformulate “ $a$  is able of ensuring that  $\varphi$ ” as “there is a history where  $a$  chooses such that  $\varphi$ ”. “ $a$  has the power of ensuring that  $\varphi$ ” is reformulated “there is a history where  $a$  knows that it chooses such that  $\varphi$ ”.

Social choice is concerned by problems involving complex mechanisms of interaction between agents. Examples of such procedures are fair-division algorithms or voting processes. We argue that Chellas’s stit may be a primitive for social choice modeling because it permits us to reason on models very close to a normal form game representation.

**Strategic games.** In von Neumann and Morgenstern’s Game Theory [MvN44], a *strategic game* or *normal form game* is a way of describing the possible interactions of agents. With two agents, it is a matrix where each entry is a possible outcome. One agent controls the rows, while the other controls the columns, and the outcome that follows is the unique one that is in both selected choices. Figure 1.3 depicts an instance of the most fa-

	defect <sub>e</sub>	silent <sub>e</sub>
defect <sub>a</sub>	(-6, -6)	(0, -10)
silent <sub>a</sub>	(-10, 0)	(-2, -2)

**Figure 1.3:** The prisoners’ dilemma in normal form game. Eloise controls the columns and Abelard controls the rows. defect<sub>a</sub> and defect<sub>e</sub> are the choices in which respectively Abelard and Eloise defect. By silent<sub>a</sub> and silent<sub>e</sub> they remain silent.

amous example of game theory, namely the *prisoners’ dilemma*. Eloise and Abelard have been arrested by the police who do not have enough evidence to be convinced of the guilt of one or the other. The prisoners are left in two different rooms and the police officer makes the same deal with both. If both remain silent (Abelard chooses the bottom row and Eloise the right column), then they will be sentenced to two years of prison each. If only one stays silent, he or she will get the full sentence of ten years while the betrayer is left free. If both denounce the other, they will get a six-year sentence each.

Typical problems of games are those of computations of *solution concepts* exemplified by Pareto optimality or Nash equilibrium. They suggest, and sometimes predict what the outcome of a particular game will be.

**A kernel for logics of agency.** Normal form games deprived of utilities are in fact the shared core of three famous theories: STIT from the philosophical side, Coalition Logic (CL) which originates in research on *social software*<sup>3</sup> and Alternating-time Temporal Logic (ATL), one of the best known logics for multiagent systems [Woo02].

They correspond to a model of agents constrained among other assumptions by the independence of choice. All in all, a normal form game is similar to a moment of STIT theory. We can easily recognize the similarities of the game of Figure 1.3 with the STIT moment represented in Figure 1.2. For the sake of the example, we modeled utilities via propositions, e.g.,  $e_2$  meant that Eloise obtains a two-year sentence: we are not interested here in rationality so we will leave payoff abstract.

Marc Pauly's aim with Coalition Logic (CL) [Pau02] was to model more explicitly games. Still, payoffs are abstracted away. It does not explain why agents have acted or should act but rather what they are able to bring about. Where logics of agency in philosophy are interested in actuality of agency, CL deals with potential agency.<sup>4</sup> Moreover, while researchers in game theory are interested in individual ability of agents, Pauly gives an account of coalitional ability, that is, how individuals can merge their efforts for ensuring tighter outcomes.

This link with normal form games is inherited by ATL which appeared to be an extension of CL, although created independently [Gor01].

## 1.4 Logics of action in computer science

When one thinks about logic of action in computer science, logics of programs as Propositional Dynamic Logic first come to mind. Nonetheless, recently ATL plays an important role in computer science, multiagent systems and artificial intelligence.

In [AHK97, AHK99, AHK02], Alur, Henzinger and Kupferman build a logic of agents on top of a famous logic in computer science: Computational Tree Logic CTL. CTL is a branching time temporal logic with modal operators quantifying (universally (**A**) and existentially (**E**)) over a set of paths.  $A\varphi$  stands for " $\varphi$  is settled", that is,  $\varphi$  is true whatever the future course of time. In ATL, at each state, a coalition of agents  $A$  can *strategi-*

---

<sup>3</sup>Social software is concerned with the issue of constructing and verifying social procedures [Par02].

<sup>4</sup>In philosophy, one interesting example of potential agency is the proposal of Brown that we do not study in this dissertation [Bro88].

*cally* (by a series of choices) force the course of time to be in some subset of paths. We write  $\langle\langle A \rangle\rangle\varphi$  if the coalition  $A$  has a strategy  $\sigma$  in its repertoire such that  $\varphi$  is true at every path compatible with the execution of  $\sigma$ . In other words “ $A$  has the ability to settle  $\varphi$ ”. This setting allows for refinements of the CTL quantification over paths, CTL E corresponding to the ATL  $\langle\langle Agt \rangle\rangle$  and **A** corresponding to  $\langle\langle \emptyset \rangle\rangle$ , where  $\emptyset$  is the empty coalition. Of course ATL inherits the temporal operators of CTL, that provide a capability for linear time reasoning along paths.

Still, we do not want to stick blindly to the Anselmian’s paradigm. We are interested in agency in a broader sense. Andrew Jones and Marek Sergot had those words, talking of the place of a ‘brings it about’ operator (along Pörn’s tradition) in the discipline of computer science:

“The ‘brings it about’ operator abstracts away details of specific actions performed by the agents, changes of states, and the temporal dimension generally; we have indicated that for certain purposes this abstraction is appropriate. But in the context of computer science, a specification employing this operator would be a formal specification at an unusually high level of abstraction. [...] It is clear that some aspects of access control mechanisms and some of the behaviour of distributed computer systems need to be modelled at a finer grain of detail. In these cases, it will be necessary to replace or augment the use of the ‘brings it about’ operator with more standard approaches to action and time in computer science.” [JS93]

Pratt’s Dynamic Logic [Pra76] is among the “more standard approaches to action” referred in the quote. The idea of Dynamic Logic is to represent actions of agents like computer programs. In the last chapters of this dissertation we introduce explicit action labels and investigate some connections between both paradigms. As a consequence, the resulting frameworks not only allow to explain *what* agents bring about or can bring about, but also to specify *how* the agents change the state of the world.

## 1.5 Outline

Chellas’s logic, STIT, CL, ATL, etc, are conceptualizations of very close notions. But we have to commit ourselves to understanding what are the exact meanings of agency operators of their language. Grounding one’s perception of a modal operator after an intuitive reading as we gave in this

introduction is always sloppy. It is not because the various modalities of agency that we consider share the same intuition that it allows us to pick up one of them randomly as soon as one needs an operator of agency for a formal theory. Our aim is here to provide a basis for the understanding of the global picture formed by the collection of logics of agents. We do it by confronting the intrinsic properties of the different approaches. Linking the various fields related in agency is the main objective of this dissertation. Rather than answering old questions or raising new ones we push towards a unification.

Appraising a logic with respect to another gives us an evident knowledge of the structures and mechanisms of every proposal. Transfers of well-known results or interpretations of a formalism to another one are facilitated. Once this work of clarification has been done, it allows us to start the work on a well-founded and tight ontology of agency and action.

STIT theory will be our frame of experimentation and somewhat our Ariadne's thread. In a first time, we give a formal contribution to the STIT theory, working out formalisms for reasoning about individual choice and coalitional choice. We relate it to Coalition Logic. We show that STIT versatile semantics provides complex mechanisms of highest interest for contemporary logic of interaction in artificial intelligence. In a second time, we capitalize on it in an attempt to shed the light on some links between agency and time. We try to reveal some issues in the notions studied in the first chapters that we find out to be too poor.

This dissertation is along the following outline. In Chapter 2, we review elements of the theory of agents and choices in branching time that we think are the minimal requirements for an understanding of STIT. We present models and the operators of the theories of the deliberative stit, a strategic version of Chellas's stit, and the semantics of the achievement stit. Chapter 3 is devoted to the computational aspects of the logics of the deliberative stit and of Chellas's stit. We first simplify the first axiomatics due to Ming Xu. This allows us to discover a link between STIT and product logics, and to simplify the semantics as well. We establish the complexity of reasoning about individual independent choices. In Chapter 4, we capitalize on the significative simplifications of the previous chapter and extend the axiomatization to collective choice and an operator of time. We show that we can embed Coalition Logic in the resulting logic. In fact, it gives a normal semantics to Coalition Logic. We then argue for the relevance of the logic for fine-grained modeling. In particular, we use it to tackle the problem of uniform strategies by a straightforward fusion with standard epistemic logic.

We somewhat stop-and-look in Chapter 5, which is less technical than the rest. We are interested in the relationship between agency and time. We first examine models of the Chellas's stit and show the 'amount of time' that they contain. Next, we propose a discrete framework derived from STIT models with instants. It provides two new collections of operators that we use to show interesting properties of logics of STIT and Chellas's  $\Delta_a\varphi$ . More specifically, we show that we can capture  $\Delta_a\varphi$  more adequately than Chellas's stit does. Then, we try to see how a modal logic can be sufficiently expressive to handle choices of agents, time and actions with duration. Chapter 6 is concerned with the embedding of Alternating-time Temporal Logic in a slightly adapted strategic STIT logic. We propose an embedding, prove its correctness, and discuss what it teaches us about the relationship between logics of agents in computer science and philosophy. Having developed a neat understanding of relevant structures of agency, Chapter 7 is devoted to a fine-grained ontology of action and agency close to the specification in modal logic of Chapter 5. In addition, we show that Chellas's stit suffers from a causal and agentive gap and treat the issue. We discuss perspectives in Chapter 8.

**Notational conventions** For constants of agents, we use  $a, a_0, a_1\dots, b, b_0\dots$  as general notation. We also use constants  $0, 1\dots, i, j, k, l\dots$  when we need a light notation, with a set of agents isomorphic to the set of integers. Groups of agents are named  $A, A_0, A_1\dots$  as in the ATL tradition or  $J, J_0\dots$  following CL tradition.

We use 'STIT' to designate a logic or a particular theory among the theories of agents and choices in branching time, e.g., deliberative STIT theories, deliberative STIT logic, Chellas STIT logic. We use 'stit' to point to an operator, e.g., Chellas's stit, achievement stit.

**Bibliographic notes** This dissertation is based on several published communications. Chapter 3 is based on a joint work with Philippe Balbiani and Andreas Herzig [BHT07a] submitted to a journal in April 2007. Chapter 4 is extracted from a joint work with Jan Broersen and Andreas Herzig [BHT07b]. In Chapter 5, some elements on NSTIT are extracted from [Tro07]. Moreover a preliminary version of the section on a logic of actions with duration (Section 5.4) has been previously published in a joint work with Laure Vieu [TV06]. Chapter 6 is adapted from [BHT06b] written with Jan Broersen and Andreas Herzig. Finally, Chapter 7 is the sequel of a paper with Robert Trypuz and Laure Vieu [TTV06].



# 2

---

## The logic of “Seeing To It That”

### 2.1 Generalities

STIT theory originates in philosophy of action. Probably the first paper to refer to the logic of *seeing to it that* (or *theory of agents and choices in branching time* is [BP88]. It analyzes the needs for a general theory of “an agent making a choice among alternatives that lead to an action”. Already since ancient Greeks, Aristotle in *Nichomachean Ethics* for instance, philosophers have been interested in the notion of agency. It has long been a challenge to make a distinction between sentences which involve agency and those which do not. Belnap and Perloff try to uncover general principles for deciding for example whether “Ishmael sails on board the Pequod” is agentive for Ishmael. It emphasizes a sort of causality and responsibility of an agent for the truth of a state of affairs. For Ishmael being agentive for sailing on the Pequod, there should be a prior choice of Ishmael which permitted it. (E.g. he chose deliberately to engage on the Pequod to break out of his depressive cycle.)

It is then proposed to introduce in a logical language, a binary operator reading roughly “agent  $a$  is agentive for  $\varphi$ ”. After an analysis of several possibilities, it is decided by Belnap and Perloff that “the English verb [...] *sees to it that*, has to [their] ears at least, fewer of the obvious defects of the others, and *sees to it that* has the definite advantage of suggesting alternatives and choices”. It thus suggests a formal operator like  $[a \textit{ stit} : \varphi]$  which reads “agent  $a$  sees to it that  $\varphi$ ”.

**STIT paraphrase thesis** The sentence  $\varphi$  marks the agentiveness of agent  $a$  just in case  $\varphi$  may usefully be paraphrased as  $[a \textit{ stit} : \varphi]$ . Therefore, up to an approximation,  $\varphi$  is agentive for  $a$  whenever  $\varphi \leftrightarrow [a \textit{ stit} : \varphi]$ . This way, deciding whether the sentence  $\varphi$  “Ishmael sails on board the Pequod” is

agentive for Ishmael is deciding whether it is equivalent to “Ishmael sees to it that Ishmael sails on board the Pequod”

[BP88] is a roadmap towards a very rich and justified theory of agency with numerous applications compiled in [BPX01] and [Hor01].

It is worth noting that STIT is influenced by the observation that in a branching time framework, future-tensed statements are ambiguous to evaluate if not impossible. Suppose a moment  $w_0$  and two different moments  $w_1$  and  $w_2$  lying in the future of  $w_0$  on two different courses of time.  $\varphi$  is true at  $w_1$  and false at  $w_2$  and everywhere before and after, on course of time passing through it. (Hence,  $\varphi$  does not hold at  $w_0$ .) What truth value should be assigned to the sentence “ $\varphi$  is true in the future of  $w_0$ ”? Indeed,  $\varphi$  really does lie in the future of  $w_0$ , but what if the course of time happens to go through  $w_2$  instead? There is a truth-value gap: in general, in branching time, a moment alone does not provide enough information to determine the truth value of a sentence about the future.

Arthur Prior [Pri67] and Richmond Thomason [Tho70, Tho84] hence proposed to evaluate future-tensed sentences with respect to a moment *and* a particular course of time running through it. This is why, as we will see, states of the world in STIT models consist of ‘fragmentized’ moments: a moment splits up into as much indexes as there are courses of time running through it.

**Remark 2.1.** *In STIT models, moments may have several valuations, depending on the history they are living in. Thus, at any specific moment, we might have different valuations corresponding to the different possible histories at that moment. This is then naturally (and on purpose) the case for Prior-Thomason tense statements, but for atomic formulas too, for uniformity purpose.*

In this section, we present the elements of the theory that are relevant in this work.

## 2.2 Deliberative STIT theories

### 2.2.1 Models: rudiments

The semantics of STIT is embedded in a branching time structure (BT). It is based on structures of the form  $\langle W, < \rangle$ , in which  $W$  is a *nonempty* set of moments, and  $<$  is a *tree-like ordering* of these moments.

**Assumption 2.1 (Tree order).** *For lighter notation, let  $w_1 \leq w_2$  iff  $w_1 < w_2$  or  $w_1 = w_2$ .*

- $w_1 \leq w_1$ ;
- if  $w_1 \leq w_2$  and  $w_2 \leq w_3$  then  $w_1 \leq w_3$ ;
- if  $w_1 \leq w_2$  and  $w_2 \leq w_1$  then  $w_1 = w_2$ ;
- if  $w_1 \leq w_3$  and  $w_2 \leq w_3$  then  $w_1 \leq w_2$  or  $w_2 \leq w_1$ ;
- for all  $w_1$  and  $w_2$  there is a  $w_0$  such that  $w_0 \leq w_1$  and  $w_0 \leq w_2$ .

A maximal set of linearly ordered moments from  $W$  is a *history*. A history being a set of history,  $w \in h$  denotes that moment  $w$  is on the history  $h$ . We define  $Hist$  as the set of all histories of a STIT structure.  $H_w = \{h | h \in Hist, w \in h\}$  denotes the set of histories passing through  $w$ . An *index* is a pair  $w/h$ , consisting of a moment  $w$  and a history  $h$  from  $H_w$  (i.e., a history and a moment in that history).

To the *BT* structure, one adds choice of agents (AC). Together, they forms the most elementary frame of the STIT theory, called an *agents and choices in branching time* structure, noted *BT + AC*. Throughout this dissertation, we will generally refer to *BT + AC* structures simply as *STIT structures* or *STIT models*. In section 2.4 we talk about a more general structure augmented by a notion of *instants*.

$Agt$  denotes a *nonempty* enumerable set of agents and  $Atm$  denotes a *nonempty* set of atomic propositions.

A *STIT model* is a tuple  $\mathcal{M} = \langle W, <, Choice, v \rangle$ , where:

- $\langle W, < \rangle$  is a branching time structure;
- $Choice : Agt \times W \rightarrow 2^{2^{Hist}}$  is a function mapping each agent and each moment  $w$  into a partition of  $H_w$ ;
- $v$  is valuation function  $v : Atm \rightarrow 2^{W \times Hist}$ .

*Choice* is the most fundamental primitive of *BT + AC* structures. We need to further specify it.

## 2.2.2 Constraints on *Choice*

### 2.2.2.1 Individual choice

The equivalence classes belonging to  $Choice_a^w$  can be thought of as possible choices or actions available to agent  $a$  at  $w$ . Given a history  $h \in H_w$ ,  $Choice_a^w(h)$  represents the particular choice from  $Choice_a^w$  containing  $h$ , or in other words, the particular action performed by  $a$  at the index  $w/h$ .

**Assumption 2.2 (Liveness).**  $Choice_a^w \neq \emptyset$  and  $Q \neq \emptyset$  for every  $Q \in Choice_a^w$ .

We say that two histories  $h_1$  and  $h_2$  are *undivided* at  $w$  iff there is a  $w'$  such that  $w < w'$ , and  $w' \in h_1 \cap h_2$ . An important constraint of  $BT + AC$  structures is the property of *no choice between undivided histories*.

**Assumption 2.3 (No choice between undivided histories).** If two histories  $h_1$  and  $h_2$  are undivided at a moment  $w$ , then  $h_2 \in Choice_a^w(h_1)$  for every agent  $a$ .

### 2.2.2.2 Group choice

In order to deal with group agency, Horty defines in [Hor01, section 2.4], what he calls *group action*. (It is named *joint agency* in [BPX01, Sect. 10C].) Horty first introduces action selection functions  $s_w$  from  $Agt$  into  $2^{H_w}$  satisfying the condition that for each  $w \in W$  and  $a \in Agt$ ,  $s_w(a) \in Choice_a^w$ . So, a selection function  $s_w$  selects a particular action for each agent at  $w$ .

For a given  $w$ ,  $Select_w$  is the set of all selection functions  $s_w$ .

**Assumption 2.4 (Independence of agents).** For every  $s_w \in Select_w$ ,  $\bigcap_{a \in Agt} s_w(a) \neq \emptyset$ .

This constraint corresponds to the assumption that the agents' choices are independent, in the sense that agents can never be deprived of choices due to the choices made by other agents. This property is called *independence of agents* (or *independence of choices*).

Using choice selection functions  $s_w$ , the *Choice* function can be generalized to apply to groups of agents ( $Choice : 2^{Agt} \times W \rightarrow 2^{Hist}$ ). A collective choice for a group of agents  $A \subseteq Agt$  is defined as:

$$Choice_A^w = \left\{ \bigcap_{a \in A} s_w(a) \mid s_w \in Select_w \right\}$$

Again,  $Choice_A^w(h) = \{h' \mid \text{there is } Q \in Choice_A^w \text{ such that } h, h' \in Q\}$ .

### 2.2.3 Truth conditions of operators

A formula is evaluated with respect to a model and an index.

$$\begin{aligned} \mathcal{M}, w/h \models p & \iff w/h \in v(p), p \in Atm. \\ \mathcal{M}, w/h \models \neg\varphi & \iff \mathcal{M}, w/h \not\models \varphi \\ \mathcal{M}, w/h \models \varphi \vee \psi & \iff \mathcal{M}, w/h \models \varphi \text{ or } \mathcal{M}, w/h \models \psi \end{aligned}$$

We have at disposition two weak temporal operators  $P\varphi$  and  $F\varphi$  for reasoning about tense statements along a history.

$$\begin{aligned}\mathcal{M}, w/h \models \mathbf{P}\varphi &\iff \exists w' \in h (w' < w, \mathcal{M}, w'/h \models \varphi) \\ \mathcal{M}, w/h \models \mathbf{F}\varphi &\iff \exists w' \in h (w < w', \mathcal{M}, w'/h \models \varphi)\end{aligned}$$

Historical necessity (or inevitability) at a moment  $w$  in a history is defined as truth in all histories passing through  $w$ :

$$\mathcal{M}, w/h \models \Box\varphi \iff \mathcal{M}, w/h' \models \varphi, \forall h' \in H_w$$

When  $\Box\varphi$  holds at one index of  $w$  then  $\varphi$  is said to be *settled true at  $w$* .  $\Diamond\varphi$  is defined in the usual way as  $\neg\Box\neg\varphi$ , and stands for historical possibility. It reads “it is historically possible that  $\varphi$ ”.

There are several STIT operators; the so-called Chellas’s stit is the most elementary one and the one we will use the most since it purveys the core of *objective choice*, central in the first part of this dissertation. It is named after the author of [Che69, Che92]. The truth condition for Chellas’s stit is as follows:

$$\mathcal{M}, w/h \models [A \text{ cstit} : \varphi] \iff \mathcal{M}, w/h' \models \varphi, \forall h' \in \text{Choice}_A^w(h)$$

Intuitively it means that group  $A$  is *choosing* to ensure  $\varphi$ , whatever other agents outside  $A$  do. (In Section 5.3.3, we study the link between Chellas’s stit and Chellas’s  $\Delta_a\varphi$  operator and point differences.) The more complex *deliberative stit*, inspired by Franz von Kutschera [vK86], can be defined as  $[A \text{ dstit} : \varphi] \triangleq [A \text{ cstit} : \varphi] \wedge \neg\Box\varphi$ . Semantically, its truth value is given by:

$$\begin{aligned}\mathcal{M}, w/h \models [A \text{ dstit} : \varphi] &\iff \mathcal{M}, w/h' \models \varphi, \forall h' \in \text{Choice}_A^w(h) \\ &\text{and } \exists h'' \in H_w, \mathcal{M}, w/h'' \not\models \varphi\end{aligned}$$

**Remark 2.2.** *The notion of group action considered here is a weak one and could be criticized from an ontological point of view. Intentionality left aside,  $[A \text{ cstit} : \varphi]$  even does not capture the fact that every agent of  $A$  were actually necessary to achieve  $\varphi$ . Indeed, for  $A \subseteq B$ ,  $[A \text{ cstit} : \varphi] \rightarrow [B \text{ cstit} : \varphi]$  is valid. This is completely assumed, and we shall see in the following chapters that it is the same for operators of group actions in Coalition Logic and Alternating-time Temporal Logic. Belnap et al. discuss this issue and propose a simple strict joint STIT operator in [BPX01, Sect. 10C] that is true for a group of agents  $A$  only if no subset  $B$  could have ensured the state of affairs.*

We say that a formula  $\varphi$  is valid (noted  $\models \varphi$ ) if  $\mathcal{M}, w/h \models \varphi$  for every STIT model  $\mathcal{M}, h$  in  $\mathcal{M}$  and moment  $w$  in  $h$ .

## 2.3 Strategic ability

[Hor01] and [BPX01] introduce strategies into STIT theory: a *strategy* for an agent  $a$  is a partial function  $\sigma$  on  $W$  such that  $\sigma(w) \in \text{Choice}_a^w$  for each moment  $w$  from  $\text{Dom}(\sigma)$ , the domain of  $\sigma$ . In STIT theory it is assumed that  $\sigma$  may be a partial function. The reason is that there is no need to account for choices at states an agent never arrives at by following  $\sigma$ . In [BPX01, p. 350] it says “A strategy need not tell us what to do at moments that the strategy itself forbids”. We shall see later in Section 6.2 that this contrasts with ATL, where it is implicitly assumed that strategies are total.

As we can see in the definition of the  $[_{cstit} : \_]$  operator, an agent’s choice restricts the set of possible futures. More precisely it restricts the histories to those corresponding with the choice being made. We expect a strategy to be a generalization of this; We want a strategy to restrict the possible histories to those compatible with a series of choices being made at successive moments.

**Definition 2.1 (admitted histories).** *A strategy  $\sigma$  admits a history  $h$  if and only if (i)  $\text{Dom}(\sigma) \cap h \neq \emptyset$  and (ii) for each  $w \in \text{Dom}(\sigma) \cap h$  we have  $h \in \sigma(w)$ . The set of all histories admitted by a strategy  $\sigma$  is denoted  $\text{Adh}(\sigma)$ .*

We will often use the notation  $\sigma_a$ , to name a particular strategy of an agent  $a$ .

Horty [Hor01] proposes strategies with a limited scope for which an agent actually plans. To this end, he introduces the notion of *field at a moment  $w$*  which is a  $<$ -backward closed nonempty subset  $M$  of  $\text{Tree}_w = \{w\} \cup \{w' \mid w < w'\}$ .

Given a corresponding set of *admitted moments*  $\text{Adm}(\sigma) = \{w \mid w \in h, h \in \text{Adh}(\sigma)\}$ , we say a strategy is *perfect* in the field  $M$  if it is *complete* in  $M$  ( $\text{Adm}(\sigma) \cap M \subseteq \text{Dom}(\sigma)$ ) and *irredundant* ( $\text{Dom}(\sigma) \subseteq \text{Adm}(\sigma)$ ). Thus, a strategic operator should be evaluated with respect to a field in addition with the usual index  $w/h$ , and representing the scope of concern of the agent. Unbounded strategies at  $w$  are simply those that are perfect in  $\text{Tree}_w$ .

As discussed in [Hor01], global effectivity by means of a strategy differs from local effectivity induced by a unique choice. Available choices at a moment form a partition of that moment: one history lies in one and only one choice. But, the sets of admitted histories of the strategies available at a given moment do *not* necessarily partition that moment. One history can lie in the sets of admitted histories of two different strategies. Therefore,

since a history alone does not tell us which strategy we have to consider, we cannot evaluate global effectivity as we have done for local effectivity (the  $[_\text{cstit}: \_]$  operator). However, we let aside those semantic difficulties here. We refer the reader to [Hor01, Sect. 7.2.1] and to [BHT06a], where we propose a first solution to this problem in the ATL setting.

Horty points out that we can return to a natural evaluation in the case of an operator for *ability* of agents by using an operator quantifying over strategies. In particular, we can define a *fused* operator for long term strategic ability of groups of agents as follows:

$$\mathcal{M}, w/h/M \models \Diamond_s[a \text{ scstit}: \varphi] \iff \exists \sigma \in \text{Strategy}_a^M \text{ s.t. } \forall h' \in \text{Adh}(\sigma), \mathcal{M}, w/h' \models \varphi$$

where  $M$  is a field at  $w$  and  $\text{Strategy}_a^M = \{\sigma \mid \sigma \text{ perfect in } M\}$ .

Intended readings for  $\Diamond_s[a \text{ scstit}: \varphi]$  are: “it is strategically possible that agent  $a$  sees to it that  $\varphi$ ”, or “ $a$  has the ability to guarantee the truth of  $\varphi$  by carrying out an available strategy”. Horty uses a slightly different syntax and writes this fused operator as  $\Diamond[a \text{ scstit} : \varphi]$ . We use the  $s$ -subscript for the diamond to emphasize that it does not reflect *historical* possibility (written without the  $s$ -subscript as  $\Diamond\varphi$ ) but *strategic* possibility.

The strategic ability operator  $\Diamond_s[a \text{ scstit}: \varphi]$  can be seen to be stronger than the local ability modality  $\Diamond[_\text{cstit}: \_]$ . In particular, it holds that:

$$\models \Diamond[a \text{ cstit}: \varphi] \rightarrow \Diamond_s[a \text{ scstit}: \varphi]$$

In fact, one can define the *Choice* function in terms of admitted histories as follows:

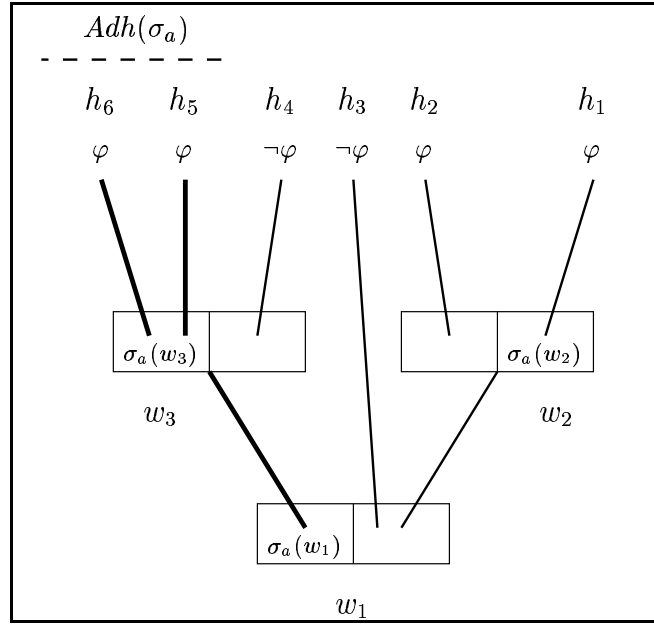
$$\text{Choice}_a^w = \{\text{Adh}(\sigma) \mid \sigma \in \text{Strategy}_a^{\{w\}}\}$$

In the basic deliberative STIT theories, we had at disposition one operator for historical possibility  $\Diamond$  and for agency  $[_\text{cstit}: \_]$ . We can compose them in a formula  $\Diamond[a \text{ cstit}: \varphi]$  which would read “it is possible that  $a$  to guarantees the truth of  $\varphi$ ”.

**Remark 2.3.** *It suggest that at a given moment, every agent’s choice is revocable. A choice can also be seen as a commitment that still could be abandoned.*

So why do we need another (complex) notion of ability? Because the choice of agents is a commitment to a ‘one-step’ strategy. And it makes a difference in the notion of ability that  $\Diamond[a \text{ cstit}: \varphi]$  and  $\Diamond_s[a \text{ scstit}: \varphi]$  capture.

$\Diamond[_\text{cstit} : \_]$  and  $\Diamond_s[_\text{scstit} : \_]$  are not equivalent: in the example of Figure 2.1, we can build a strategy  $\sigma_a$  such that  $\sigma_a(w_1) = \{h_4, h_5, h_6\}$ ,



**Figure 2.1:** Time goes upward. It is strategically possible that agent  $a$  sees to it that some time in the future  $\varphi$ .

$\sigma_a(w_2) = \{h_1\}$  and  $\sigma_a(w_3) = \{h_5, h_6\}$ .  $h_1, h_2$  and  $h_3$  are not admitted because they do not lie in  $\sigma_a(w_1)$ .  $\text{Dom}(\sigma_a) \cap h_4 = \{w_1, w_3\}$ , but  $h_4 \notin \sigma_a(w_3)$ , so  $h_4 \notin \text{Adh}(\sigma_a)$ . However,  $h_5$  and  $h_6$  are in  $\text{Adh}(\sigma_a)$ , and there are no other histories in  $\text{Adh}(\sigma_a)$ . So, there exists a strategy  $\sigma_a$  perfect in  $\{w_1, w_2, w_3\}$  such that for every history in  $\text{Adh}(\sigma_a)$ ,  $\varphi$  is true some time in the future. So, for all  $h \in H_{w_1}$ ,  $\mathcal{M}, w_1/h \models \Diamond_s[a \text{ scstit} : \mathbf{F}\varphi]$ . However, for any  $h \in H_{w_1}$  we also have  $\mathcal{M}, w_1/h \not\models \Diamond[a \text{ cstit} : \mathbf{F}\varphi]$ .

On the contrary, the strategy  $\sigma'_a$  with  $\sigma'_a(w_1) = \{h_1, h_2, h_3\}$ ,  $\sigma'_a(w_2) = \{h_1\}$  and  $\sigma'_a(w_3) = \{h_4\}$  cannot ensure that  $\varphi$  some time in the future, because  $\text{Adh}(\sigma'_a) = \{h_1, h_3\}$ , and  $\mathcal{M}, w_1/h_3 \not\models \mathbf{F}\varphi$ .

$\Diamond_s[a \text{ scstit} : \varphi]$  thus expresses a stronger notion of ability than  $\Diamond[a \text{ cstit} : \varphi]$ . We will see it at work in Chapter 6.

## 2.4 Achievement stit

Historically,  $BT + I + AC$  structures preceded  $BT + AC$  structures, whose purpose was to simplify the framework. We preferred to present them incrementally.  $I$  stands for *instant*. This section introduces into the semantics

a new partial relation among moments allowing to compare moments that may lie on different histories and hence are incomparable by the temporal order relation  $<$ .

A  $BT + I + AC$  model is a tuple  $\mathcal{M} = \langle W, <, Choice, instant, v \rangle$ , where:

- $\langle W, <, Choice, v \rangle$  is a  $BT + AC$  model
- $instant : W \mapsto 2^W$  maps every moment into the set of moments lying in the same instant.

$instant$  can be seen as an equivalence relation, partitioning  $W$  in temporal layers.

**Definition 2.2 (choice equivalence).** *Two moments  $w_1$  and  $w_2$  are  $Choice_A^w$  – equivalent iff (1)  $instant(w_1) = instant(w_2)$  (2)  $w$  is a moment  $w$  prior to both  $w_1$  and  $w_2$  (3)  $w_1$  and  $w_2$  lie on histories belonging to the same  $Choice_A^w$  partition.*

Note that we have the following property:

**Proposition 2.1.** *If  $w'' < w' < w$  then the set of moments  $Choice_A^{w'}$  – equivalent of  $w$  is a subset of the set of the moments  $Choice_A^{w''}$  – equivalent of  $w$ .*

PROOF. Straightforward. ■

We now are able to provide the truth conditions of the *achievement stit* operator:

$$\mathcal{M}, w/h \models [A \text{ stit} : \varphi] \iff \begin{aligned} &(\exists w_0 < w \text{ s.t. } \forall w_1, \forall h' \in H_{w_1} \\ &\text{if } w \text{ is } Choice_A^{w_0} \text{ – equivalent} \\ &\text{then } \mathcal{M}, w_1/h' \models \varphi) \text{ and} \\ &(\exists w_2 \in instant(w), \exists h'' \in H_{w_2} \text{ s.t.} \\ &w < w_2 \text{ and } \mathcal{M}, w_2/h'' \not\models \varphi) \end{aligned}$$

$[A \text{ stit} : \varphi]$  means that agents of  $A$  have ensured that  $\varphi$  holds now by making a choice previously – namely at  $w_0$  that we call the *witness moment* of the operator – and if they had made a different choice,  $\varphi$  could have been false at the present instant.

**Proposition 2.2.** *If  $w' < w$  and  $H_{w'} = H_w$  then  $w'$  is not a witness moment of an achievement stit.*

PROOF. It is sufficient to remark that at such a moment  $w'$  every agent had only one vacuous choice. This is due to the *no choice between undivided histories* constraint on  $BT + AC$  structures. Then it could not be otherwise, as the negative condition of the achievement stit requires. ■

---

In Section 5.3, we propose a discrete  $BT + AC + I$  framework. It makes clearer what are the relationships between the achievement stit and Chellas’s and deliberative stit. Discreteness is a simplification that permits us to shed a syntactical light on the intrinsic links of those operators. Moreover, we are able to give a more precise characterization of Chellas’s  $\Delta_a\varphi$  operator [Che69, Che92].

# 3

---

## Meta-logical aspects of individual choice

### 3.1 Introduction

While STIT has played an important role in philosophical logic since the eighties, it seems to be fair to say that its mathematical aspects have not been developed to the same extent. Most probably the reason is that STIT's models of agency are much more complex than those existing for other modal concepts (such as say necessity, belief, or knowledge): first, the 'seeing-to-it-that' modalities interact (or perhaps better: must be guaranteed not to interact) because the agents' choices are supposed to be independent; second there is another kind of modality involved, viz. the 'master modality' of historic necessity. There are also temporal modalities, but just as most of the other proof-theoretic approaches to STIT, we do not investigate these here.

As a consequence, proof systems for STIT are rather complex, too. To our knowledge the following have been proposed in the literature.

- Xu provides Hilbert-style axiomatizations in terms of the historic necessity operator and Chellas's stit operator [BPX01, Chap. 17], without considering temporal operators. As the deliberative stit operator can be expressed in terms of Chellas's (together with the historic necessity operator), the axiomatization transfers to the deliberative stit. Xu proves their completeness (without considering the temporal dimension), by means of canonical models, and proves decidability by means of filtration. Besides, Xu also gives a complete axiomatization of the one-agent achievement stit [BPX01, Chap. 16].
- Wansing provides a tableau proof system for the deliberative stit [Wan06]. The system is complete, but does not guarantee termi-

nation, and thus “is not tailored for defining tableau algorithms” [Wan06].

- Dégrement gives a dialogical proof procedure for the deliberative stit [Dég06]. Again, the system is complete, but does not guarantee termination, and can therefore only be used to build proofs by hand.

In this chapter, we focus on the so-called Chellas stit named after his proponent [Che69, Che92]. The original operator defined by Chellas is nevertheless notably different since it does not come with the principle of independence of agents that plays a central role here. We use the term **CSTIT** to refer to the logic of that modal operator. We show that Xu’s axiomatics of the logic of the Chellas stit can be greatly simplified. After recalling it (Section 3.2) we propose an alternative one and prove its completeness (Section 3.3). Based on the latter we show that in presence of at least two agents, the modal operator of historic necessity can be defined as an abbreviation (Section 3.4). This leads to a simplified semantics (Section 3.5), and to characterizations of the complexity of satisfiability (Section 3.6).

## 3.2 Xu’s axioms for the individual Chellas STIT (CSTIT)

Some preliminary remarks are due. In [BPX01, Chap. 17], Ming Xu presents *Ldm*, an axiomatization for the basic (that is, without temporal operators) *deliberative* STIT logic. As pointed out, deliberative STIT logic and Chellas STIT logic are interdefinable and just differ in the choice of primitive operators. Following Xu we refer to these two logics as the *deliberative STIT theories*. We here mainly focus on *Ldm* with the Chellas stit operator as primitive.

### 3.2.1 Language

The language of Chellas STIT logic is built from a countably infinite set of atomic propositions *Atm* and a countable set of agents *Agt*. To simplify notation we suppose that *Agt* is an initial subset  $\{0, 1, \dots\}$  of  $\mathbb{N}$  (possibly  $\mathbb{N}$  itself).

Formulas are built by means of the boolean connectives together with modal operators of historic necessity and of agency in the standard way. Usually these modal constructions are noted *Sett* :  $\varphi$  (“ $\varphi$  is settled”) and  $[i \text{ cstit} : \varphi]$  (“ $i$  sees to it that  $\varphi$ ”), where  $i \in \text{Agt}$ . For reasons of conciseness

we here prefer to use  $\Box\varphi$  instead of  $Sett : \varphi$ , and  $[i]\varphi$  instead of  $[i\ cstit : \varphi]$ . The language  $\mathcal{L}_{\text{CSTIT}}$  of the Chellas stit is therefore defined by the following BNF:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid [i]\varphi \mid \Box\varphi$$

where  $p$  ranges over  $\mathcal{A}tm$  and  $i$  ranges over  $\mathcal{A}gt$ . This provides a standard notation for the dual constructions  $\Diamond\varphi$  and  $\langle i \rangle\varphi$ , abbreviating  $\neg\Box\neg\varphi$  and  $\neg[i]\neg\varphi$ , respectively.

The language  $\mathcal{L}_{\text{DSTIT}}$  of the deliberative stit is defined by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid [i\ dstit : \varphi] \mid \Box\varphi$$

Note that neither  $\mathcal{L}_{\text{CSTIT}}$  nor  $\mathcal{L}_{\text{DSTIT}}$  contain temporal operators.

The following function will be useful to compute the number of symbols that are necessary to write down  $\varphi$ .

**Definition 3.1.** We define recursively a mapping  $\|\cdot\|$  from formulas of  $\mathcal{L}_{\text{CSTIT}} \cup \mathcal{L}_{\text{DSTIT}}$  to  $\mathbb{N}$ :  $\|p\| = 1$ ,  $\|\neg\varphi\| = 1 + \|\varphi\|$ ,  $\|(\varphi \wedge \psi)\| = 3 + \|\varphi\| + \|\psi\|$ ,  $\|[i]\varphi\| = 3 + \|\varphi\|$ , and  $\|[i\ dstit : \varphi]\| = 5 + \|\varphi\|$ .

### 3.2.2 Axiomatics

Xu gave the following axiomatics of Chellas's CSTIT:

S5( $\Box$ ) the axiom schemas of S5 for  $\Box$

S5( $i$ ) the axiom schemas of S5 for every  $[i]$

( $\Box \rightarrow i$ )  $\Box\varphi \rightarrow [i]\varphi$

(AIA $_k$ )  $(\Diamond[0]\varphi_0 \wedge \dots \wedge \Diamond[k]\varphi_k) \rightarrow \Diamond([0]\varphi_0 \wedge \dots \wedge [k]\varphi_k)$

The last item is a family of *axiom schemes for independence of agents* that is parameterized by the integer  $k$ .<sup>1</sup>

<sup>1</sup>Xu's original formulation of (AIA $_k$ ) is

$$(\text{diff}(i_0, \dots, i_k) \wedge \Diamond[i_0]\varphi_0 \wedge \dots \wedge \Diamond[i_k]\varphi_k) \rightarrow \Diamond([i_0]\varphi_0 \wedge \dots \wedge [i_k]\varphi_k)$$

for  $1 \leq k$ . The difference predicates  $\text{diff}(i_0, \dots, i_k)$  express that  $i_0, \dots, i_k$  are all distinct. They are defined from an equality predicate  $=$  whose domain is  $\mathcal{A}gt$ . Formally we have to add the axioms:  $\text{diff}(i_0) \leftrightarrow \top$ , and

$$\text{diff}(i_0, \dots, i_{k+1}) \leftrightarrow \text{diff}(i_0, \dots, i_k) \wedge i_1 \neq i_{k+1} \wedge \dots \wedge i_k \neq i_{k+1}.$$

In consequence Xu's axiomatics has to contain axioms for equality. We here preferred not to introduce equality in order to stay with the same logical language throughout.

**Remark 3.1.** *As  $(AIA_{k+1})$  implies  $(AIA_k)$ , the family of schemas can be replaced by the single  $(AIA_{|Agt|-1})$  when  $Agt$  is finite.*

Xu's system has the standard inference rules of modus ponens and necessitation for  $\Box$ . From the latter necessitation rules for every  $[i]$  follow by axiom  $(\Box \rightarrow i)$ .

**Theorem 3.1 ([BPX01, Chapter 17]).** *A formula  $\varphi$  of  $\mathcal{L}_{\text{CSTIT}}$  is valid in  $BT+AC$  structures iff  $\varphi$  is provable from the schemas  $S5(\Box)$ ,  $S5(i)$ ,  $(\Box \rightarrow i)$ , and  $(AIA_k)$  by the rules of modus ponens and  $\Box$ -necessitation.*

Xu's decidability proof proceeds by building a canonical model followed by filtration [BPX01, Theorems 17-18]. Although he does not mention complexity issues, when decidability is proved by canonical model construction from which a finite model is obtained by filtration, then "a NEXPTIME algorithm is usually being employed" [BdRV01, Appendix C, p. 515]. Therefore it can be expected that the problem of deciding the satisfiability of a given formula of  $\mathcal{L}_{\text{CSTIT}}$  is in NEXPTIME. We shall characterize complexity precisely in Section 3.6.

### 3.3 An alternative axiomatics

We now prove that  $(AIA_k)$  can be replaced by the family of axiom schemes

$$(AAIA_k) \quad \Diamond\varphi \rightarrow \langle k \rangle \bigwedge_{0 \leq i < k} \langle i \rangle \varphi \quad \text{for } k \geq 1$$

We call  $(AAIA_k)$  the *alternative axiom schema for independence of agents*. Just as Xu's  $(AIA_k)$ ,  $(AAIA_k)$  involves  $k + 1$  agents.

**Lemma 3.1 (validity of  $AAIA_k$ ).** *For each  $k \geq 1$ ,  $\Diamond\varphi \rightarrow \langle k \rangle \bigwedge_{0 \leq i < k} \langle i \rangle \varphi$  is valid in  $BT+AC$  structures.*

PROOF. See annex in Section 3.8. ■

To warm up, we first prove that our  $(AAIA_1)$  implies Xu's  $(AIA_1)$ .

Clearly, each of our  $(AIA_k)$  can be proved from Xu's original  $(AIA_k)$ . The other way round, given  $k$  and pairwise different  $i_0, \dots, i_k$ , suppose w.l.o.g. that  $i_k \geq i_n$  for  $n \leq k$ . Then one can prove Xu's  $(AIA_k)$

$$(\Diamond[i_0]\varphi_{i_0} \wedge \dots \wedge \Diamond[i_k]\varphi_{i_k}) \rightarrow \Diamond([i_0]\varphi_{i_0} \wedge \dots \wedge [i_k]\varphi_{i_k})$$

from our  $(AIA_{i_k})$

$$(\Diamond[0]\varphi_0 \wedge \dots \wedge \Diamond[i_k]\varphi_{i_k}) \rightarrow \Diamond([0]\varphi_0 \wedge \dots \wedge [i_k]\varphi_{i_k})$$

by appropriately choosing  $\varphi_n$  to be  $\top$  for all those  $n < i_k$  that are not among  $i_0, \dots, i_k$ : as  $[n]\varphi_n \leftrightarrow \top$  and  $\Diamond[n]\varphi_n \leftrightarrow \top$  hold, these conjuncts can be dropped from our  $(AIA_{i_k})$ .

**Lemma 3.2.** *The schema (AIA<sub>1</sub>) is provable from S5(□), S5(i), (□→i) and:*

$$(AAIA_1) \ \Diamond\varphi \rightarrow \langle 1 \rangle \langle 0 \rangle \varphi$$

by modus ponens and □-necessitation.

PROOF. We establish the following deduction:

1.  $\Diamond[0]\varphi_0 \rightarrow \langle 1 \rangle \langle 0 \rangle [0]\varphi_0$     from axiom (AAIA<sub>1</sub>), substituting  $[0]\varphi_0$  for  $\varphi$
2.  $\Diamond[0]\varphi_0 \rightarrow \langle 1 \rangle [0]\varphi_0$     from previous line by S5(0)
3.  $\Diamond[0]\varphi_0 \wedge [1]\varphi_1 \rightarrow \langle 1 \rangle [0]\varphi_0 \wedge [1][1]\varphi_1$     from previous line by S5(1)
4.  $\Diamond[0]\varphi_0 \wedge [1]\varphi_1 \rightarrow \langle 1 \rangle ([0]\varphi_0 \wedge [1]\varphi_1)$     from previous line by K(1)
5.  $\Diamond(\Diamond[0]\varphi_0 \wedge [1]\varphi_1) \rightarrow \Diamond\langle 1 \rangle ([0]\varphi_0 \wedge [1]\varphi_1)$   
from previous line by □-necessitation and K(□)
6.  $\Diamond[0]\varphi_0 \wedge \Diamond[1]\varphi_1 \rightarrow \Diamond\langle 1 \rangle ([0]\varphi_0 \wedge [1]\varphi_1)$     from previous line by S5(□)
7.  $\Diamond[0]\varphi_0 \wedge \Diamond[1]\varphi_1 \rightarrow \Diamond([0]\varphi_0 \wedge [1]\varphi_1)$   
from previous line by (□→i) axiom and S5(□)

■

We turn back to an arbitrary number of agents.

**Lemma 3.3.** *Every schema (AIA<sub>k</sub>) is provable from S5(□), S5(i), (□→i) and (AAIA<sub>k</sub>) by the rules of modus ponens and □-necessitation.*

PROOF. We proceed by induction on  $k$ . The base case  $k = 1$  is settled by Lemma 3.2. Now, suppose AIA<sub>k-1</sub> is provable:

$$\Diamond[0]\varphi_0 \wedge \dots \wedge \Diamond[k-1]\varphi_{k-1} \rightarrow \Diamond([0]\varphi_0 \wedge \dots \wedge [k-1]\varphi_{k-1}).$$

We prove AIA<sub>k</sub> with the following steps.

1.  $\bigwedge_{i < k} \Diamond[i]\varphi_i \rightarrow \Diamond \bigwedge_{i < k} [i]\varphi_i$     by induction hypothesis (AIA<sub>k-1</sub>)
2.  $\bigwedge_{i < k} \Diamond[i]\varphi_i \rightarrow \langle k \rangle (\bigwedge_{j < k} \langle j \rangle \bigwedge_{i < k} [i]\varphi_i)$     from previous line by (AAIA<sub>k</sub>)
3.  $\bigwedge_{i < k} \Diamond[i]\varphi_i \rightarrow \langle k \rangle \bigwedge_{j < k} \langle j \rangle [j]\varphi_j$     from previous line by K(j)
4.  $\bigwedge_{i < k} \Diamond[i]\varphi_i \wedge [k]\varphi_k \rightarrow \langle k \rangle (\bigwedge_{j < k} [j]\varphi_j) \wedge [k]\varphi_k$   
from previous line by S5(i)
5.  $\bigwedge_{i < k} \Diamond[i]\varphi_i \wedge [k]\varphi_k \rightarrow \langle k \rangle \bigwedge_{j \leq k} [j]\varphi_j$     from previous line by S5(k)

6.  $\Diamond(\bigwedge_{i < k} \Diamond[i]\varphi_i \wedge [k]\varphi_k) \rightarrow \Diamond\langle k \rangle \bigwedge_{j \leq k} [j]\varphi_j$   
 from previous line by  $\Box$ -necessitation and  $K(\Box)$
7.  $\Diamond(\bigwedge_{i < k} \Diamond[i]\varphi_i \wedge [k]\varphi_k) \rightarrow \Diamond \bigwedge_{j \leq k} [j]\varphi_j$   
 from previous line by  $(\Box \rightarrow i)$  axiom and  $S5(\Box)$
8.  $\bigwedge_{i \leq k} \Diamond[i]\varphi_i \rightarrow \Diamond \bigwedge_{j \leq k} [j]\varphi_j$  from previous line by  $S5(\Box)$

■

**Theorem 3.2.** *A formula of  $\mathcal{L}_{\text{CSTIT}}$  is valid in  $BT+AC$  structures iff it is provable from the axiom schemas  $S5(\Box)$ ,  $S5(i)$ ,  $(\Box \rightarrow i)$  and  $(AAIA_k)$  by the rules modus ponens and  $\Box$ -necessitation.*

PROOF. First, observe that Xu's axiomatics and ours only differ by the schemas  $(AIA_k)$  and  $(AAIA_k)$ .

Soundness follows from:

1. the validity of our schemas  $AAIA_k$  (see Lemma 3.1),
2. the validity of the rest of the axioms, and
3. the fact that modus ponens and  $\Box$ -necessitation preserve validity.

The last two points are warranted by the soundness of Xu's axioms (Theorem 3.1).

Completeness follows from provability of Xu's  $(AIA_k)$  from our  $(AAIA_k)$  (see Lemma 3.3). As observed above, the rest of Xu's axioms is directly present in our axiomatics. ■

An alternative axiomatics of the deliberative stit is obtained viewing  $[i]\varphi$  as an abbreviation of  $[i \text{ dstit} : \varphi] \vee \Box\varphi$ .

### 3.4 Historic necessity is superfluous in presence of two agents or more

In this section, we suppose that  $| \text{Agt} | \geq 2$ , i.e. there are at least agents 0 and 1.

The equivalence  $\Diamond\varphi \leftrightarrow \langle 1 \rangle \langle 0 \rangle \varphi$  is provable from  $(AAIA_1)$ ,  $(\Box \rightarrow i)$  and  $S5(\Box)$ . This suggests that  $\Box\varphi$  can be viewed as an abbreviation of  $[1][0]\varphi$ . Let us take this as an axiom schema.

Def( $\Box$ )  $\Box\varphi \leftrightarrow [1][0]\varphi$

Pushing this further we can prove that under  $\text{Def}(\Box)$ , axiom  $(AAIA_k)$  can be replaced by the family of axiom schemas of general permutation:

$$(G\text{Perm}_k) \langle l \rangle \langle m \rangle \varphi \rightarrow \langle n \rangle \bigwedge_{i \leq k, i \neq n} \langle i \rangle \varphi \quad \text{for } k \geq 0$$

Note that similar to Xu's axiomatization, if  $\mathcal{Agt}$  is finite then the single schema  $(G\text{Perm}_{|\mathcal{Agt}|-1})$  is sufficient.

The next lemma establishes soundness.

**Lemma 3.4.**  *$(G\text{Perm}_k)$  is valid in  $BT+AC$  structures.*

PROOF. See annex in Section 3.8. ■

Now we prove that the principles of the preceding section can be derived.

**Lemma 3.5.** *The axiom schemas of  $S5(\Box)$ , and the schemas  $(\Box \rightarrow i)$  and  $(AAIA_k)$  are provable from  $\text{Def}(\Box)$ ,  $S5(i)$  and  $(G\text{Perm}_k)$  by the rules of modus ponens and  $[i]$ -necessitation, and  $\Box$ -necessitation is derivable.*

PROOF. First let us prove that the logic of  $\Box$  is  $S5$ . Clearly the K-axiom  $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$  is provable using standard modal principles, and the T-axiom  $\Box\varphi \rightarrow \varphi$  follows from  $S5(0)$  and  $S5(1)$ . It remains to prove the 5-axiom  $\Diamond\varphi \rightarrow \Box\Diamond\varphi$ :

1.  $\langle 1 \rangle \langle 0 \rangle \varphi \rightarrow [1] \langle 1 \rangle \langle 0 \rangle \varphi$  by  $S5(1)$ ;
2.  $[1] \langle 1 \rangle \langle 0 \rangle \varphi \rightarrow [1] \langle 0 \rangle \langle 1 \rangle \varphi$  by  $(G\text{Perm}_1)$  and  $K(1)$ ;
3.  $[1] \langle 0 \rangle \langle 1 \rangle \varphi \rightarrow [1][0] \langle 0 \rangle \langle 1 \rangle \varphi$  by  $S5(0)$  and  $K(1)$ ;
4.  $[1][0] \langle 0 \rangle \langle 1 \rangle \varphi \rightarrow [1][0] \langle 1 \rangle \langle 0 \rangle \varphi$  by  $(G\text{Perm}_1)$ ;
5.  $\langle 1 \rangle \langle 0 \rangle \varphi \rightarrow [1][0] \langle 1 \rangle \langle 0 \rangle \varphi$  from lines 1-4.

Finally,  $\Box$ -necessitation is derivable by applying first 0-necessitation and then 1-necessitation.

Concerning  $(AAIA_k)$  it is easy to see that under  $\text{Def}(\Box)$  it is an instance of  $(G\text{Perm}_k)$ , for all  $k \geq 1$ . It remains to prove  $(\Box \rightarrow i)$ . Let us show that  $\langle i \rangle \varphi \rightarrow \langle 1 \rangle \langle 0 \rangle \varphi$ :

1.  $\langle i \rangle \varphi \rightarrow \langle i \rangle \langle j \rangle \varphi$  by  $S5(i)$ ;
2.  $\langle i \rangle \langle j \rangle \varphi \rightarrow \langle 1 \rangle \langle 0 \rangle \varphi$  by  $(G\text{Perm}_1)$ ;
3.  $\langle i \rangle \varphi \rightarrow \langle 1 \rangle \langle 0 \rangle \varphi$  from lines 1-2.



**Theorem 3.3.** *Suppose  $|\mathcal{Agt}| \geq 2$ . Then a formula of  $\mathcal{L}_{\text{CSITIT}}$  is valid in  $\text{BT+AC}$  structures iff it is provable from  $S5(i)$ ,  $\text{Def}(\Box)$ , and  $(\text{GPerm}_k)$  by the rules of modus ponens and  $[i]$ -necessitation.*

**Remark 3.2.** *If  $\mathcal{Agt} = \{0, 1\}$  then the validities of  $\mathcal{L}_{\text{CSITIT}}$  are axiomatized by  $\text{Def}(\Box)$ ,  $S5(1)$ ,  $S5(2)$ , and  $\langle 1 \rangle \langle 0 \rangle \varphi \leftrightarrow \langle 0 \rangle \langle 1 \rangle \varphi$ . Moreover, the Church-Rosser axiom  $\langle 0 \rangle [1] \varphi \rightarrow [1] \langle 0 \rangle \varphi$  can be proved from  $S5(1)$ ,  $S5(2)$  and  $(\text{GPerm}_1)$ . Therefore **STIT** logic with two agents is a so-called product logic, alias a two-dimensional modal logic [Mar99, GKWZ03]. Such product logics are characterized by the permutation axiom  $\langle 0 \rangle \langle 1 \rangle \varphi \leftrightarrow \langle 1 \rangle \langle 0 \rangle \varphi$  together with the Church-Rosser axiom. Hence the logic of the two-agent Chellas **STIT** is nothing but the product  $S5^2 = S5 \otimes S5$ .*

### 3.5 A simpler semantics

All axiom schemes are in the Sahlqvist class [BdRV01], and therefore have a standard possible worlds semantics.

*Kripke models* are of the form  $M = \langle W, R, V \rangle$ , where  $W$  is a nonempty set of possible worlds,  $R$  is a mapping associating to every  $i \in \mathcal{Agt}$  an equivalence relation  $R_i$  on  $W$ , and  $V$  is a mapping from  $\mathcal{Atm}$  to the set of subsets of  $W$ . We impose that  $R$  satisfies the following property:

**Definition 3.2 (general permutation property).** *We say that  $R$  satisfies the general permutation property iff for all  $w, v \in W$  and for all  $l, m, n \in \mathcal{Agt}$ , if  $\langle w, v \rangle \in R_l \circ R_m$  then there is  $u \in W$  such that:  $\langle w, u \rangle \in R_n$  and  $\langle u, v \rangle \in R_i$  for every  $i \in \mathcal{Agt} \setminus \{n\}$ .*

We have the usual truth condition:

$$M, w \models [i]\varphi \text{ iff } M, u \models \varphi \text{ for every } u \text{ such that } \langle w, u \rangle \in R_i$$

and the usual definitions of validity and satisfiability.

**Lemma 3.6.** *For every  $M = \langle W, R, V \rangle$ , and every  $i, j \in \mathcal{Agt}$ ,  $R$  satisfies the following properties:*

1. *If  $i \neq j$  then  $R_i \circ R_j = R_1 \circ R_0$ .*
2.  *$R_i \circ R_j$  is an equivalence relation for every  $i, j \in \mathcal{Agt}$ .*

$$3. (\bigcup_{i \in \mathcal{A}gt} R_i)^* = R_0 \circ R_1 = R_1 \circ R_0.$$

PROOF. (1) follows from the validity of  $\langle i \rangle \langle j \rangle \varphi \rightarrow \langle 1 \rangle \langle 0 \rangle \varphi$  (due to (GPerm<sub>0</sub>)), and the validity of  $\langle 1 \rangle \langle 0 \rangle \varphi \rightarrow \langle i \rangle \langle j \rangle \varphi$  (due to (GPerm<sub>j</sub>), given that  $i \neq j$ ).

(2) follows from (1) and the fact that the S5-axioms are valid for  $\Box$  (see Lemma 3.5).

In (3), the right-to-left inclusion  $R_0 \circ R_1 \subseteq (\bigcup_{i \in \mathcal{A}gt} R_i)^*$  follows from the inclusion  $R_0 \circ R_1 \subseteq (R_0 \cup R_1)^*$ . For the left-to-right inclusion suppose  $\langle w, v \rangle \in (\bigcup_{i \in \mathcal{A}gt} R_i)^*$ . Hence there are  $i_0, \dots, i_k$  such that  $\langle w, v \rangle \in R_{i_0} \circ \dots \circ R_{i_k}$ . As all the  $R_{i_i}$  are equivalence relations we may suppose w.l.o.g. that  $i_l \neq i_{l+1}$ .

- If  $k$  is odd then  $R_{i_0} \circ \dots \circ R_{i_k} = (R_0 \circ R_1)^{k/2}$  by (1). The latter is equal to  $R_0 \circ R_1$  by (2).
- If  $k$  is even then  $R_{i_0} \circ \dots \circ R_{i_k} = (R_0 \circ R_1)^{(k-1)/2} \circ R_{i_k} = (R_0 \circ R_1) \circ R_{i_k}$  by (1) and (2). The latter is equal to  $R_0 \circ R_1 \circ R_0$  again by (1), and to  $R_0 \circ R_0 \circ R_1$  by (2), which is equal to  $R_0 \circ R_1$  because  $R_0$  is an equivalence relation.

It follows that  $(\bigcup_{i \in \mathcal{A}gt} R_i)^* \subseteq R_0 \circ R_1$ . ■

**Theorem 3.4.** *A formula of  $\mathcal{L}_{\text{CSTIT}}$  is valid in Kripke models satisfying the general permutation property iff it is provable from*

S5(*i*)      the axiom schemas of S5 for every [*i*]

Def( $\Box$ )     $\Box \varphi \leftrightarrow [1][0]\varphi$

(GPerm<sub>*k*</sub>)  $\langle l \rangle \langle m \rangle \varphi \rightarrow \langle n \rangle \bigwedge_{i \leq k, i \neq l} \langle i \rangle \varphi$  for  $k \geq 1$

by the rules of modus ponens and [*i*]-necessitation.

PROOF. If  $\mathcal{A}gt$  is finite then Sahlqvist's Theorem warrants that our axiomatics of Section 3.4 is sound and complete w.r.t. Kripke models satisfying the general permutation property. We show in the annex that this can be extended to the infinite case. ■

## 3.6 Complexity

The axiom system of the preceding section allows us to characterize the complexity of satisfiability of STIT formulas. We study separately the cases of Chellas STIT and of the deliberative STIT.

### 3.6.1 Complexity of CSTIT

First, satisfiability of CSTIT-formulas can be decided in nondeterministic exponential time.

**Lemma 3.7.** *The problem of deciding satisfiability of a formula of  $\mathcal{L}_{\text{CSTIT}}$  is in NEXPTIME.*

PROOF. This can be proved by the standard filtration construction, which establishes that in order to know whether a formula  $\varphi$  is satisfiable in the Kripke models of Section 3.5 it suffices to consider models having at most  $2^{\|\varphi\|}$  possible worlds. See the annex for details. ■

In the rest of the section we show that the upper bound is tight if there are at least two agents. As usual we start with the two-agents case.

**Lemma 3.8.** *If  $|\mathcal{Agt}| = 2$  then the problem of deciding satisfiability of a formula of  $\mathcal{L}_{\text{CSTIT}}$  is NEXPTIME-hard.*

PROOF. Remember our observation at the end of Section 3.4: when  $|\mathcal{Agt}| = 2$  then  $\text{CSTIT}_{\mathcal{Agt}}$  is nothing but the product logic  $S5 \otimes S5$ . We can then apply a result of Marx in [Mar99], who proved that the problem of deciding membership of  $\varphi$  in  $S5 \otimes S5$  is NEXPTIME-hard. (Actually Marx also proved membership in NEXPTIME.) ■

Hence two-agent CSTIT logic is NEXPTIME-complete. Now we state NEXPTIME-completeness for any number of agents greater than 2.

**Theorem 3.5.** *If  $|\mathcal{Agt}| \geq 2$  then the problem of deciding satisfiability of a formula of  $\mathcal{L}_{\text{CSTIT}}$  is NEXPTIME-complete.*

PROOF. See annex in Section 3.8. ■

It remains to establish the complexity of single-agent CSTIT. It turns out that it has the same complexity as S5.

**Theorem 3.6.** *If  $|\mathcal{Agt}| = 1$  then the problem of deciding satisfiability of a formula of  $\mathcal{L}_{\text{CSTIT}}$  is NP-complete.*

PROOF. This can be proved by establishing an upper bound on the size of the models that is quadratic in the length of the formula under concern. ■

**Remark 3.3.** *Intriguingly, while one-agent STIT has the same complexity as S5, and two-agent STIT has the same complexity as S5<sup>2</sup>, 3-agent STIT does not have the same complexity as S5<sup>3</sup>: while Xu's proof establishes decidability of  $\mathcal{L}_{\text{CSTIT}}$ -formulas for any number of agents, it was proved by Maddux that S5<sup>3</sup> is undecidable [MM01].*

Thus we have characterized the complexity of satisfiability of CSTIT formulas for all cases.

### 3.6.2 Complexity of the deliberative STIT logic

The complexity results for Chellas STIT do not immediately transfer to DSTIT. Indeed, the definition of the deliberative STIT from the CSTIT through  $[i \text{ dstit} : \varphi] = [i]\varphi \wedge \neg\Box\varphi$  does not directly provide a lower bound for the deliberative STIT because this is not a polynomial transformation. We now establish these results by giving polynomial translations from CSTIT to DSTIT and vice versa.

Let  $\varphi_0$  be any formula of  $\mathcal{L}_{\text{DSTIT}}$ , and let  $sf(\varphi_0)$  be the set of subformulas of  $\varphi_0$ . Let  $\{p_\psi : \psi \in sf(\varphi_0)\}$  be a set of (pairwise distinct) atoms none of which occurs in  $\varphi_0$ . Every  $p_\psi$  abbreviates the subformula  $\psi$  of  $\varphi_0$ . We recursively define equivalences ('bimplications') that capture the logical relation between  $p_\psi$  and  $\psi$ .

**Definition 3.3.** *We define:*

$$\begin{aligned} B_q &= (p_q \leftrightarrow q) \\ B_{\neg\varphi} &= (p_{\neg\varphi} \leftrightarrow \neg p_\varphi) \\ B_{\varphi \wedge \psi} &= (p_{\varphi \wedge \psi} \leftrightarrow p_\varphi \wedge p_\psi) \\ B_{\Box\varphi} &= (p_{\Box\varphi} \leftrightarrow \Box p_\varphi) \\ B_{[i:\text{dstit}\varphi]} &= (p_{[i:\text{dstit}\varphi]} \leftrightarrow [i]p_\varphi \wedge \neg\Box p_\varphi) \end{aligned}$$

**Definition 3.4.** *We define the translation  $tr$  from DSTIT formulas to CSTIT formulas as:  $tr(\varphi_0) = p_{\varphi_0} \wedge \bigwedge_{\psi \in sf(\varphi_0)} \Box B_\psi$ .*

**Theorem 3.7.**  *$tr$  is a polynomial translation from  $\mathcal{L}_{\text{DSTIT}}$  to  $\mathcal{L}_{\text{CSTIT}}$ , and for every formula  $\varphi_0$  of  $\mathcal{L}_{\text{DSTIT}}$ ,  $\varphi_0$  is satisfiable iff  $tr(\varphi_0)$  is satisfiable.*

PROOF. See annex in Section 3.8. ■

It follows that the problem of deciding whether a formula of  $\mathcal{L}_{\text{DSTIT}}$  is satisfiable is in NEXPTIME. We now prove that this bound is tight.

**Definition 3.5.** We define equivalences  $B'_\varphi$  such that

$$B'_{[i]\varphi} = (p_{[i]\varphi} \leftrightarrow [i \text{ dstit} : p_\varphi] \vee \Box p_\varphi)$$

and  $B'_\varphi = B_\varphi$  if  $\varphi$  is an atomic formula or if its main logical connector is boolean.

**Definition 3.6.** We define the translation  $tr'$  from  $\mathcal{L}_{\text{CSTIT}}$  to  $\mathcal{L}_{\text{DSTIT}}$  as:  $tr'(\varphi_0) = p_{\varphi_0} \wedge \bigwedge_{\psi \in sf(\varphi_0)} \Box B'_\psi$ .

**Theorem 3.8.**  $tr'$  is a polynomial translation from  $\mathcal{L}_{\text{CSTIT}}$  to  $\mathcal{L}_{\text{DSTIT}}$ , and for every formula  $\varphi_0$  of  $\mathcal{L}_{\text{CSTIT}}$ ,  $\varphi_0$  is satisfiable iff  $tr(\varphi_0)$  is satisfiable.

PROOF. The proof is analogous to that of Theorem 3.7. ■

Together, Theorems 3.5, 3.6, 3.7 and 3.8 entail:

**Corollary 3.1.** The problem of deciding whether a formula of  $\mathcal{L}_{\text{DSTIT}}$  is satisfiable is NEXPTIME-complete if  $|\mathcal{Agt}| \geq 2$ , and it is NP-complete if  $|\mathcal{Agt}| = 1$ .

## 3.7 Conclusion

In this chapter we have established NEXPTIME-completeness of the satisfiability problem of formulas of Chellas STIT and of the deliberative STIT for the case of two or more agents. All our complexity results appear to be new.

Our new axiom system for STIT of Section 3.3 is an interesting alternative to Xu's. It highlights the central role of the well-known equivalences  $[i][j]\varphi \leftrightarrow \Box\varphi$  and  $[i \text{ dstit} : [j \text{ dstit} : \varphi]] \leftrightarrow \perp$ , for  $i \neq j$  in theories of agency: as we have shown, they allow to capture independence of agents just as Xu's schema (AIA<sub>k</sub>) does.

For the case of more than two agents, Section 3.4 provides a quite simple axiom system that is made up of very basic modal principles, and moreover, does without historic necessity.

As we have pointed out in Section 3.3, an alternative axiomatics for the deliberative STIT follows straightforwardly. We do not know whether the redundancy of historic necessity that we have established for the CSTIT in Section 3.4 transfers to the deliberative STIT.

## 3.8 Annex: Proofs

### A.1: Proof of Lemma 3.1

In order to prove the validity of every schema

$$(AAIA_k) \ \diamond\varphi \rightarrow \langle k \rangle \bigwedge_{0 \leq i < k} \langle i \rangle \varphi \quad \text{for } k \geq 1$$

in BT+AC structures, we show that for every  $w \in W$ ,  $h, h' \in H_w$  and  $k \in \mathcal{Agt}$  there is  $h_k \in \text{Choice}_k^w(h)$  such that  $h' \in \text{Choice}_i^w(h_k)$  for every  $i \in \mathcal{Agt} \setminus \{k\}$ .

Consider the strategy  $s_w$  such that  $s_w(k) = \text{Choice}_k^w(h)$ , and  $s_w(i) = \text{Choice}_i^w(h')$  for every  $i \neq k$ . By the superadditivity constraint there is some  $h_k$  such that  $h_k \in \bigcap_{i \in \mathcal{Agt}} s_w(i)$ . Hence  $h_k \in \text{Choice}_k^w(h)$ , and  $h' \in \text{Choice}_i^w(h_k)$  for  $i \neq k$ .

### A.2: Proof of Lemma 3.4

We have to prove the validity of every schema

$$(GPerm_k) \ \langle l \rangle \langle m \rangle \varphi \rightarrow \langle n \rangle \bigwedge_{i \leq k, i \neq n} \langle i \rangle \varphi \quad \text{for } k \geq 0$$

in BT+AC structures.

A look at the proof of Lemma 3.1 shows that  $\diamond\varphi \rightarrow \langle n \rangle \bigwedge_{i \leq k, i \neq n} \langle i \rangle \varphi$  is valid in BT+AC structures. It therefore suffices to show the validity of  $\langle l \rangle \langle m \rangle \varphi \rightarrow \diamond\varphi$ . The latter is the case because (1)  $\langle l \rangle \langle m \rangle \varphi \rightarrow \diamond\diamond\varphi$  is valid (due to validity of axiom  $(\Box \rightarrow i)$ ), and (2)  $\diamond\diamond\varphi \rightarrow \diamond\varphi$  is valid (due to validity of S5( $\Box$ )).

### A.3: Proof of Theorem 3.4

We prove the theorem for the infinite case, i.e.  $|\mathcal{Agt}| = \mathbb{N}$ . In this case the general permutation property is no longer a first-order property, and Sahlqvist's result does not apply, i.e. the canonical model does not necessarily satisfy the general permutation property.

Let  $\varphi$  be a formula that is consistent w.r.t. the axiomatic system of Section 3.4. Let  $M = \langle W, R, V \rangle$  be the canonical model associated to this system. By arguments following the lines of those in the proof of Lemma 3.6 we have:

- $\forall i \in \mathcal{Agt}, R_i$  is an equivalence relation;
- $\forall i, j \in \mathcal{Agt}$  such that  $i \neq j, R_i \circ R_j = R_1 \circ R_0$ ;

- $(\bigcup_{i \in \mathcal{A}gt} R_i)^* = R_0 \circ R_1 = R_1 \circ R_0$ .

By the truth lemma we may suppose that  $M$  is generated via  $R_1 \circ R_0$  from a possible world  $w \in W$  such that  $M, w \models \varphi$ . Let  $M' = \langle W', R', V' \rangle$  be the filtration of  $M$  w.r.t.  $sf(\varphi)$  (just as done in Annex A.4). Note that  $R'_i = W' \times W'$  for all  $i \in \mathcal{A}gt$  not occurring in  $\varphi$ . This allows us to show that  $M'$  satisfies the general permutation property. From this completeness follows (via the filtration lemma).

#### A.4: Proof of Lemma 3.7

Let  $M = \langle W, R, V \rangle$  be a Kripke model such that every  $R_i$  is an equivalence relation and  $R$  satisfies the general permutation property. Let  $u$  be a world and  $\varphi$  a formula of  $\mathcal{L}_{\text{CSIT}}$  such that  $M, u \models \varphi$ . Suppose that  $M$  is generated from  $w$  through  $R_1 \circ R_0$ . (This can be supposed w.l.o.g. because of Lemma 3.6 of Section 3.5.)  $sf(\varphi)$  being the set of all subformulas of  $\varphi$ , we say  $w$  and  $v$  are  $sf(\varphi)$ -equivalent iff  $\forall \psi \in sf(\varphi)$ ,  $(M, w \models \psi \text{ iff } M, v \models \psi)$ , and note  $w \equiv_{sf(\varphi)} v$ . Let  $|w|_{\equiv_{sf(\varphi)}}$  denote the equivalence class of  $w$  modulo  $\equiv_{sf(\varphi)}$ .

We construct  $M' = \langle W', R', V' \rangle$  such that:

- $W' = W|_{\equiv_{sf(\varphi)}} = \{|w|_{\equiv_{sf(\varphi)}} : w \in W\}$
- $\langle |w|, |v| \rangle \in R'_i$  iff  $\forall [i]\psi \in sf(\varphi)$ ,  $(M, w \models [i]\psi \text{ iff } M, v \models [i]\psi)$
- $V'(p) = \{|w| : w \in V(p)\}$  for all  $p \in sf(\varphi)$

Remark that for all  $i \in \mathcal{A}gt$ , if  $i$  does not occur in  $\varphi$  then  $R'_i = W' \times W'$ .

We must check that every  $R'_i$  is an equivalence relation, that  $M'$  verifies the general permutation property, that for all  $\psi \in sf(\varphi)$  and  $w \in W$ ,  $M, w \models \psi$  iff  $M', |w| \models \psi$ , and that  $|W'|$  is exponential in the length of  $\varphi$ :

1. Every  $R'_i$  is an equivalence relation, and  $M'$  satisfies the general permutation property.

This follows from the definition of  $R'_i$ .

2.  $\forall \psi \in sf(\varphi)$ ,  $\forall w \in W$ ,  $(M, w \models \psi \text{ iff } M', |w| \models \psi)$ .

This follows from the filtration lemma (see [BdRV01] for details).

3.  $|W'| \leq 2^{\|\varphi\|}$

Note that members of  $W'$  are subsets of states of  $W$  satisfying exactly the same formulas of  $sf(\varphi)$ . Thus  $|W'| \leq 2^{|sf(\varphi)|}$  corresponding to the set of subsets of  $sf(\varphi)$ . We can show by induction on  $\psi$  that  $|sf(\psi)| \leq \|\psi\|$  and then conclude.

Hence,  $\forall \varphi \in \mathcal{L}_{\text{CSITIT}}$ , if  $\varphi$  is satisfiable then  $\exists M = \langle W, R, V \rangle$  such that  $|W| \leq 2^{|\varphi|}$  and there is  $w \in W$  such that  $M, w \models \varphi$ . It allows us to propose a decision procedure with input  $\varphi \in \mathcal{L}_{\text{CSITIT}}$ , and which works as follows: guess an integer  $N \leq 2^{|\varphi|}$  and a model  $M = \langle W, R, V \rangle$  such that  $|W| \leq N$ ; then check whether there is a  $w \in W$  such that  $M, w \models \varphi$ .

### A.5: Proof of Theorem 3.5

The upper bound is given by Lemma 3.7.

To establish the lower bound consider the set of formulas where only the agent symbols 0 and 1 occur. We show that deciding satisfiability of any formula of that fragment is NEXPTIME-hard, for any  $\mathcal{Agt}$  such that  $|\mathcal{Agt}| \geq 2$ . If  $\mathcal{Agt}$  is just  $\{0, 1\}$  this holds by Lemma 3.8. Else we prove that if  $\{0, 1\} \subset \mathcal{Agt}$  then the logic of Kripke models for  $\mathcal{Agt}$  is a conservative extension of that for  $\{0, 1\}$ .

Let  $\varphi$  be any formula containing only 0 and 1.

For the left-to-right direction, suppose  $\varphi$  is valid in all Kripke models for the set of agents  $\{0, 1\}$ . By Theorem 3.3,  $\varphi$  can then be proved from axioms (GPerm<sub>1</sub>), (Perm01), S5(0) and S5(1) with the rules of modus ponens, [0]- and [1]-necessitation. Therefore  $\varphi$  is also provable from the ‘bigger’ axiomatics for  $\mathcal{Agt}$ .

For the right-to-left direction, suppose there is a Kripke model  $M = \langle W, R, V \rangle$  for the set of agents  $\{0, 1\}$  and a  $w \in W$  such that  $M, w \models \varphi$ , where  $R : \{0, 1\} \rightarrow \mathcal{P}(W \times W)$  associates to every  $i \in \{0, 1\}$  an equivalence relation  $R_i$  on  $W$ . We are going to build a Kripke model  $M'$  for the bigger set of agents  $\mathcal{Agt}$  such that  $M', w \models \varphi$ . Let  $M' = \langle W, R', V \rangle$  such that  $R' : \mathcal{Agt} \rightarrow \mathcal{P}(W \times W)$  with  $R'_0 = R_0$ ,  $R'_1 = R_1$  and  $R'_i = R_0 \circ R_1$  for  $i \geq 2$ . Clearly  $M', w \models \varphi$ , too. It remains to show that  $M'$  is indeed a Kripke model as required in Section 3.5. By item 2 of Lemma 3.6 every  $R'_i$  is an equivalence relation, so we only have to show that the general permutation property holds in  $M'$ : if  $\langle w, v \rangle \in R'_i \circ R'_m$  then there is  $u_n \in W$  such that:  $\langle w, u_n \rangle \in R'_n$  and  $\langle u_n, v \rangle \in R'_i$  for every  $i \in \mathcal{Agt} \setminus \{n\}$  (cf. Lemma 3.4). First we show that for every  $l$  and  $m$  we have  $R'_l \circ R'_m = R_0 \circ R_1$ .

- If  $i = 0$  and  $j = 1$  then trivially  $R'_i \circ R'_m = R_0 \circ R_1$ .
- If  $l = 1$  and  $m = 0$  then  $R'_l \circ R'_m = R_1 \circ R_0 = R_0 \circ R_1$
- If  $l = 0$  and  $m \geq 2$  then  $R'_l \circ R'_m = R_0 \circ R_0 \circ R_1 = R_0 \circ R_1$
- If  $l = 1$  and  $m \geq 2$  then  $R'_l \circ R'_m = R_1 \circ R_0 \circ R_1 = R_0 \circ R_1 \circ R_1 = R_0 \circ R_1$
- If  $l \geq 2$  and  $m = 0$  then  $R'_l \circ R'_m = R_0 \circ R_1 \circ R_0 = R_0 \circ R_0 \circ R_1 = R_0 \circ R_1$

- If  $l \geq 2$  and  $m = 1$  then  $R'_l \circ R'_m = R_0 \circ R_1 \circ R_1 = R_0 \circ R_1$
- if  $l \geq 2$  and  $m \geq 2$  then  $R'_l \circ R'_m = R_0 \circ R_1 \circ R_0 \circ R_1 = R_0 \circ R_0 \circ R_1 \circ R_1 = R_0 \circ R_1$

(The identities in all these items hold because  $R_0$  and  $R_1$  permute by item 1 of Lemma 3.6, and because  $R_0$  and  $R_1$  are equivalence relations.) Thus  $\langle w, v \rangle \in R'_l \circ R'_m$  implies  $\langle w, v \rangle \in R_0 \circ R_1$ . We have to show that for every  $n \geq 1$  there is  $u_n \in W$  such that:  $\langle w, u_n \rangle \in R'_n$  and  $\langle u_n, v \rangle \in R'_i$ , for every  $i \in \mathcal{Agt}$ .

- For  $n = 1$ ,  $\langle w, v \rangle \in R_0 \circ R_1$  implies that  $\langle w, v \rangle \in R_1 \circ R_0$  by item 1 of Lemma 3.6, and the latter implies that  $\langle w, v \rangle \in R'_1 \circ R'_0$ . Therefore there is a  $u_1$  such that  $\langle w, u_1 \rangle \in R'_1$  and  $\langle u_1, v \rangle \in R'_0$ .
- For  $n \geq 2$ , take  $u_n = v$ :  $\langle w, v \rangle \in R_0 \circ R_1$  implies that  $\langle w, v \rangle \in R'_n$  by definition of  $R'_n$ , and we have  $\langle v, v \rangle \in R'_i$  because every  $R'_i$  is an equivalence relation (for  $i \geq 2$  this is the case by item 2 of Lemma 3.6).

### A.6: Proof of Theorem 3.7

The proof is done via the following lemmata.

**Lemma 3.9.** *For all formulas  $\varphi_0$  in the language of DSTIT, if  $\varphi_0$  is satisfiable then  $tr(\varphi_0)$  is satisfiable.*

PROOF. Suppose there is  $M = \langle W, R_\square, R, V \rangle$  such that  $M, w \models \varphi_0$ . We build a model  $M' = \langle W, R_\square, R, V' \rangle$  such that  $M', w \models tr(\varphi_0)$  by setting  $V'(q) = V(q)$  for all atoms  $q$  appearing in  $\varphi_0$ , and  $V'(p_\psi) = \{w \in W : M, w \models \psi\}$  for all  $\psi \in sf(\varphi_0)$ .

By induction on the structure of  $\psi$  we show that  $M, v \models B_\psi$  for all  $v \in W$  and all  $\psi \in sf(\varphi_0)$ . (Details left to the reader.)

Hence  $M' \models \bigwedge_{\psi \in sf(\varphi_0)} B_\psi$ , and also  $M' \models \bigwedge_{\psi \in sf(\varphi_0)} \Box B_\psi$ . Since  $M, w \models \varphi_0$ , we have  $M', w \models p_{\varphi_0}$  by construction of  $V'$ . Thus  $M', w \models p_{\varphi_0} \wedge \bigwedge_{\psi \in sf(\varphi_0)} \Box B_\psi$ , in other words  $M', w \models tr(\varphi_0)$ . ■

**Lemma 3.10.** *For all formulas  $\varphi_0$  in the language of DSTIT, if  $tr(\varphi_0)$  is satisfiable then  $\varphi_0$  is satisfiable.*

PROOF. Suppose there is  $M = \langle W, R_\square, R, V \rangle$  such that  $M, w \models tr(\varphi_0)$ . Thus  $M, w \models p_{\varphi_0} \wedge \bigwedge_{\psi \in sf(\varphi_0)} \Box B_\psi$ . By induction on the structure of  $\psi$  we show that  $M, v \models p_\psi \leftrightarrow \psi$  for all  $v \in W$  and all  $\psi \in sf(\varphi_0)$ . (Details left to the reader.)

Thus  $M, w \models p_{\varphi_0}$ , and  $M, w \models p_{\varphi_0} \leftrightarrow \varphi_0$ . Hence  $M, w \models \varphi_0$ . ■

**Lemma 3.11.** *tr is a polynomial transformation.*

PROOF. We easily show that  $\|B_\psi\| \leq 12$  and  $\|\bigwedge_{\psi \in sf(\varphi_0)} \Box B_\psi\| \leq \|\varphi_0\| \cdot (2 + \|B_\psi\|)$ . Then,  $\|\bigwedge_{\psi \in sf(\varphi_0)} \Box B_\psi\| \leq 14 \cdot \|\varphi_0\|$ . We conclude that  $\|tr(\varphi_0)\| \leq 1 + 14 \cdot \|\varphi_0\|$ . Remark that  $|sf(\varphi_0)| \leq \|\varphi_0\|$ . Moreover, for every formula  $\varphi$  in the language of CSTIT,  $\|B_\varphi\| = \mathcal{O}(\|\varphi\|)$ . As a result,  $\|tr(\varphi_0)\| = \mathcal{O}(\|\varphi_0\|^2)$ . ■



# 4

---

## Logics of collective choice

### 4.1 Introduction

We have gained insight from the precedent formal study of individual choice. In this chapter our aim is push the axiomatization of individual choice to *coalitional choice*.

We motivate our notion of coalitional choice by investigating its links with Coalition Logic, a famous logic for *coalitional ability*, but also by introducing some mental attitudes. We are particularly interested in the mix of the concept of choice with epistemic aspects. We claim that a language able to model *actual choice* and not only *possible choice* is drastically more relevant when we want to reason about actional-epistemic statements.

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

More recently the issue of ‘knowing how to act’ has come up in the logic ATEL [vdHW02] which is the epistemic extension of the logic of strategic ability Alternating-time Temporal Logic ATL [AHK02], that is the subject of Chapter 6. The problem is often referred to as the problem of *uniform strategies*. In particular, ATEL does not allow to distinguish the situations where:

1. the agent  $a$  knows it has a particular action/choice in its repertoire that ensures  $\varphi$ , possibly without knowing which choice to make to ensure  $\varphi$ .

2. the agent  $a$  ‘knows how to’ / ‘can’ / ‘has the power to’ ensure  $\varphi$ .

In this chapter we do not reason about series of choices, alias strategies. This means that our starting point is not ATL, but its fragment Coalition Logic, CL for short. CL was proposed by Pauly in [Pau01] as a logic for reasoning about social procedures characterized by complex strategic interactions between agents, individuals or groups. Examples of such procedures are fair-division algorithms or voting processes. CL facilitates reasoning about abilities of coalitions in games by extending classical logic with operators  $\langle\!\langle J \rangle\!\rangle\varphi$  for groups of agents  $J$ , reading: “the coalition  $J$  has a joint strategy to ensure that  $\varphi$ ”.<sup>1</sup>

We show how CL is naturally embedded in a variant of so-called theories of agents and choices in branching time for which we have now a more formal acquaintance. We extend CSTIT to coalitions and obtain a logic (determined by its models) that we call  $\mathcal{G}$ STIT for ‘group STIT’. We briefly show some difficulties of simulating Coalition Logic in it. We then extend  $\mathcal{G}$ STIT with a ‘next’ operator, resulting in a logic that we call NCL. We provide a complete axiomatization and prove that CL is embedded. This in itself is an interesting result since it shows that NCL extends CL with capabilities of reasoning about what a coalition is actually doing (as opposed to what it *could* do). We also propose a brief study of abilities of agents in NCL. Finally, we extend NCL with an S5 modal operator for knowledge and show that the resulting complete logic, that we call ‘Conformant NCL’, solves the problem of uniform strategies and discuss it.

## 4.2 Coalition Logic

Let  $\mathcal{A}gt$  be a set of agents and  $\mathcal{A}tm$  a countable set of atomic formulas. The syntax of Coalition Logic is defined as follows:

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

where  $p$  ranges over  $\mathcal{A}tm$  and  $J$  ranges over the subsets of  $\mathcal{A}gt$ . The other boolean connectives are defined as usual. In the sequel, we present two different semantics for that language.

---

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

### 4.2.1 Coalition model semantics

**Definition 4.1 (effectivity function).** Given a set of agents  $\mathcal{Agt}$  and a set of states  $S$ , an effectivity function is a function  $E : 2^{\mathcal{Agt}} \rightarrow 2^{2^S}$ . An effectivity function is said to be:

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

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

**Definition 4.2 (playable effectivity function).** An effectivity function  $E : 2^{\mathcal{Agt}} \rightarrow 2^{2^S}$  is said to be playable iff

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

**Definition 4.3.** A coalition model is a pair  $((S, E), V)$  where:

- $S$  is a nonempty set of states;
- $E : S \rightarrow (2^{\mathcal{Agt}} \rightarrow 2^{2^S})$  is a playable effectivity structure;
- $V : S \rightarrow 2^{Atm}$  is a valuation function.

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

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

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

## 4.2.2 Game semantics

In [Pau02], Marc Pauly investigates the link between coalition models and strategic games.

**Definition 4.4.** A strategic game is a tuple  $G = (S, \{\Sigma_i | i \in \text{Agt}\}, o)$  where  $S$  is a nonempty set,  $\Sigma_i$  is a nonempty set of choices for every agent  $i \in \text{Agt}$ ,  $o : \prod_{i \in \text{Agt}} \Sigma_i \rightarrow S$  is an outcome function which associates an outcome state in  $S$  with every combination of choice of agents (choice profile).

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

**Definition 4.5.** Given a strategic game  $G$ , the effectivity function  $E_G : 2^N \rightarrow 2^{2^S}$  of  $G$  is defined as  $X \in E_G(C)$  iff there is  $\sigma_C \in \prod_{i \in C} \Sigma_i$  such that for every  $\sigma_{\bar{C}} \in \prod_{i \in \bar{C}} \Sigma_i$  we have  $o(\sigma_C \times \sigma_{\bar{C}}) \in X$ .

Pauly then gives the following characterization:

**Theorem 4.1 ([Pau02]).** An effectivity function  $E$  is playable iff it is the effectivity function of some strategic game.

**Definition 4.6.** Let  $E$  be an effectivity function. A set  $Y \subseteq S$  is called a minimal effectivity outcome at  $s$  for  $J$  iff (1)  $Y \in E_s(J)$  and (2) there is no  $Y' \in E_s(J)$  s.t.  $Y' \subset Y$ .

**Definition 4.7.** The non-monotonic core of  $E$  is the mapping  $\mu_E : 2^{\text{Agt}} \times S \rightarrow 2^{2^S}$  such that  $\mu_E(J, s) = \{Y \mid Y \text{ is a minimal effectivity outcome at } s \text{ for } J\}$ .

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

**Proposition 4.1.** If  $E$  is a playable effectivity function then  $\mu_E(\text{Agt}, s)$  is a nonempty set of singletons.

PROOF. This is a corollary of Theorem 4.1. ■

## 4.2.3 Axiomatization

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

- |             |  |
|-------------|--|
| (ProTau)    | all tautologies of classical propositional logic |
| ( $\perp$ ) | $\neg \langle J \rangle \perp$                   |

( $\top$ )	$\langle J \rangle \top$
( $N$ )	$\neg \langle \emptyset \rangle \neg \varphi \rightarrow \langle \mathcal{A}gt \rangle \varphi$
( $M$ )	$\langle J \rangle (\varphi \wedge \psi) \rightarrow \langle J \rangle \psi$
( $S$ )	$\langle J_1 \rangle \varphi \wedge \langle J_2 \rangle \psi \rightarrow \langle J_1 \cup J_2 \rangle (\varphi \wedge \psi)$ if $J_1 \cap J_2 = \emptyset$
( $MP$ )	from $\varphi$ and $\varphi \rightarrow \psi$ infer $\psi$
( $RE$ )	from $\varphi \leftrightarrow \psi$ infer $\langle J \rangle \varphi \leftrightarrow \langle J \rangle \psi$

**Theorem 4.2 ([Pau02]).** *The principles (ProTau), ( $\perp$ ), ( $\top$ ), ( $N$ ), ( $M$ ), ( $S$ ), ( $MP$ ) and ( $RE$ ) are complete with respect to the class of all coalition models.*

Note that the ( $N$ ) axiom follows from the determinism of *choice profiles* (concurrent choices for every agent in the system): when every agent opts for a choice, the next state is fully determined, thus, if a formula is not settled to be true next, the coalition of all agents ( $\mathcal{A}gt$ ) can always coordinate their choices to make its negation true. The axiom ( $S$ ) says that two disjoint coalitions can combine their efforts to ensure a conjunction of properties. Note that from ( $S$ ) and ( $\perp$ ) it follows that  $\langle J_1 \rangle \varphi \wedge \langle J_2 \rangle \neg \varphi$  is not satisfiable for disjoint  $J_1$  and  $J_2$ . So, two disjoint coalitions cannot ensure opposed facts.

Theoremhood and consistency are defined as usual.

### 4.3 Extension of CSTIT to groups of agents ( $\mathcal{G}STIT$ )

For the purpose of multiagent systems, we need to have a system able to reason about coalitions. We present the logic  $\mathcal{G}STIT$  which is an extension of CSTIT to groups of agents.

Let  $\mathcal{A}gt = \{0, \dots, n-1\}$  be a finite set of  $n \geq 1$  agents and  $\mathcal{A}tm$  a countable set of atomic formulas.  $\mathcal{G}STIT$  has the following syntax, where  $p$  ranges over elements of  $\mathcal{A}tm$  and  $J$  ranges over the set of subsets of  $\mathcal{A}gt$ :

$$\varphi ::= p \mid \neg \varphi \mid \varphi \vee \psi \mid [J]\varphi$$

The other boolean connectives are as usual defined by abbreviations, and  $\langle J \rangle \varphi =_{def} \neg [J] \neg \varphi$ .  $\langle J \rangle \varphi$  roughly reads that “ $J$  does not prevent  $\varphi$ ”. In other words, “ $J$  by its current choice does not rule out  $\varphi$  as a possible outcome.”  $\bar{J}$  denotes the complement of  $J$  w.r.t.  $\mathcal{A}gt$ .

We give to  $\mathcal{G}STIT$  the axiomatization of Figure 4.1. We moreover have

(ProTau)	Enough propositional tautologies
S5( $[J]$ )	S5 axioms for every $[J]$
(Mon)	$[J_1]\varphi \rightarrow [J_1 \cup J_2]\varphi$
Elim( $[\emptyset]$ )	$\langle \emptyset \rangle \varphi \rightarrow \langle J \rangle \langle \bar{J} \rangle \varphi$

**Figure 4.1:** Axiomatics of  $\mathcal{G}STIT$ .

the standard inference rules of modus ponens and necessitation for  $[\emptyset]$ . From the latter necessitation for every  $[J]$  follows by the inclusion axiom (Mon). Theoremhood ( $\vdash_{\mathcal{G}STIT}$ ) is defined as usual.

Note that the converse of Elim( $\emptyset$ ) can be proved from (Mon), S5( $\emptyset$ ). Hence, we have  $\vdash_{\mathcal{G}STIT} \langle \emptyset \rangle \varphi \leftrightarrow \langle J \rangle \langle \bar{J} \rangle \varphi$ .

**Definition 4.8.** A  $\mathcal{G}STIT$ -model is a tuple  $\mathcal{M} = (W, R, \pi)$  where:

- $W$  is a set of worlds (alias contexts);
- $R$  is a collection of equivalence relations  $R_J$  (one for every coalition  $J \subseteq \text{Agt}$ ) such that:
  - $R_{J_1 \cup J_2} \subseteq R_{J_1}$
  - $R_\emptyset \subseteq R_J \circ R_{\bar{J}}$
- $\pi : W \rightarrow 2^{\text{Atm}}$  is a valuation function.

The truth conditions are:

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

and as usual for the classical operators. Validity ( $\models_{\mathcal{G}STIT}$ ) is also as usual.

**Theorem 4.3.**  $\mathcal{G}STIT$  is determined by the class of  $\mathcal{G}STIT$ -models.

PROOF. All axioms are Sahlqvist-type and trivially correspond to the constraints on frames (cf. [BdRV01]). Soundness proof is routine. ■

**Embedding CSTIT in  $\mathcal{G}$ STIT**  $\mathcal{G}$ STIT is easily proved to be a conservative extension of CSTIT. For that, we give the following translation from  $\mathcal{L}_{\text{CSTIT}}$  to formulas of  $\mathcal{G}$ STIT:

$$\begin{aligned} tr(p) &= p \\ tr(\neg\varphi) &= \neg tr(\varphi) \\ tr(\varphi \vee \psi) &= tr(\varphi) \vee tr(\psi) \\ tr(\Box\varphi) &= [\emptyset]\varphi \\ tr([i]\varphi) &= [\{i\}]tr(\varphi) \end{aligned}$$

Note that  $[\emptyset]$  will play here the role of STIT's  $\Box$ , and thus denotes historical settledness.

**Lemma 4.1.** *The translation of  $(G\text{Perm}_k)$  by  $tr$  is a theorem of  $\mathcal{G}$ STIT.*

PROOF. By applying (Mon) to  $\langle\{l\}\rangle$  and  $\langle\{m\}\rangle$  in the right part of the tautology  $\langle\{l\}\rangle\langle\{m\}\rangle\varphi \rightarrow \langle\{l\}\rangle\langle\{m\}\rangle\varphi$  and next S5( $[\emptyset]$ ) we have  $\vdash_{\mathcal{G}\text{STIT}} \langle\{l\}\rangle\langle\{m\}\rangle\varphi \rightarrow \langle\emptyset\rangle\varphi$ . Then by Elim( $\emptyset$ ) we obtain  $\vdash_{\mathcal{G}\text{STIT}} \langle\{l\}\rangle\langle\{m\}\rangle\varphi \rightarrow \langle\{n\}\rangle\langle\{n\}\rangle\varphi$ .

Now, by classical principles on instances of (Mon)  $\langle\overline{\{n\}}\rangle\varphi \rightarrow \langle\{i\}\rangle\varphi$  for every  $i \in \text{Agt} \setminus \{n\}$ , we have  $\vdash_{\mathcal{G}\text{STIT}} \langle\overline{\{n\}}\rangle\varphi \rightarrow \bigwedge_{i \in \text{Agt} \setminus \{n\}} \langle\{i\}\rangle\varphi$ . We conclude that  $\vdash_{\mathcal{G}\text{STIT}} \langle\{l\}\rangle\langle\{m\}\rangle\varphi \rightarrow \langle\{n\}\rangle \bigwedge_{i \in \text{Agt} \setminus \{m\}} \langle\{i\}\rangle\varphi$ . ■

We prove that  $\mathcal{G}$ STIT is a conservative extension of CSTIT in presence of at least two agents.

**Theorem 4.4.** *If  $\varphi \in \mathcal{L}_{\text{CSTIT}}$ ,  $\models_{\text{CSTIT}} \varphi$  iff  $\models_{\mathcal{G}\text{STIT}} tr(\varphi)$ .*

PROOF.

( $\Rightarrow$ ) Let a model  $M = \langle W, R, V \rangle$  for CSTIT as defined in Section 3.5 and  $\varphi$  a CSTIT-formula satisfied at  $M, x$  ( $x \in W$ ).

We transform  $M$  in a  $\mathcal{G}$ STIT-model  $\mathcal{M} = (W', R', \pi)$  in a way that:

- $W' = W$ ;
- $R'_J = \bigcap_{j \in J} R_j$ ;
- $R'_\emptyset = R_1 \circ R_0$ ;
- $\pi(w) = V(w)$ , for every  $w \in W'$ .

It is easy to check that the constructed model  $\mathcal{M}$  satisfies every constraint on  $\mathcal{G}$ STIT models and  $\mathcal{M}, x \models_{\mathcal{G}\text{STIT}} tr(\varphi)$ .

( $\Leftarrow$ ) Remind that besides  $(G\text{Perm}_k)$ , the only other principles of CSTIT are S5 axioms for  $[\{i\}]$ . Thus, from S5( $[J]$ ) and Lemma 4.1, we have that every translated axiom of CSTIT is a theorem of  $\mathcal{G}$ STIT. Moreover, translated inference rules preserve validity.



## 4.4 Coalitional choice plus discrete time

### 4.4.1 Motivations

Now we are able to reason about choice of coalitions, it seems that  $\mathcal{G}$ STIT is a suitable logic to simulate Coalition Logic. Indeed, we have seen that the semantics of the CL operator  $\langle J \rangle$  is a  $\exists - \forall$  pattern. Hence, CL formulas of the form  $\langle J \rangle \varphi$  intuitively should correspond to the composition of an existential quantification of outcomes via the  $\langle \emptyset \rangle$  modality (“it is possible that...”), and the modality of agency  $[J]$  (“whatever other agents do,  $J$  sees to it that...”).

It thus suggests a translation  $tr_0$  which maps a CL formula  $\langle J \rangle \varphi$  to a  $\mathcal{G}$ STIT formula  $\langle \emptyset \rangle [J] tr_0(\varphi)$ , and is homomorphic on the propositional fragment of the language of Coalition Logic.

However, such a translation is not correct. To see that, consider the following consistent CL formula

$$\langle \emptyset \rangle p \wedge \langle \emptyset \rangle \langle J \rangle \neg p.$$

It says that  $p$  is a necessary outcome of the current game, and that  $J$  will be able to ensure  $\neg p$  next, at any next one. By  $tr_0$ , it translates to:

$$\langle \emptyset \rangle [\emptyset] p \wedge \langle \emptyset \rangle [\emptyset] \langle \emptyset \rangle [J] \neg p.$$

By S5 modal principles on  $[\emptyset]$ , it collapses to  $[\emptyset] p \wedge \langle \emptyset \rangle [J] \neg p$ , which is not consistent in S5. (It entails  $\langle \emptyset \rangle ([J] p \wedge [J] \neg p)$  by standard S5 modal principles and (Mon), which is not satisfiable because of S5( $[J]$ ).)

The problem here is that existential ( $\langle \emptyset \rangle$ ) and universal ( $[J]$ ) quantifications are done over the same domain. Then those modalities are ‘intricate’, essentially by monotonicity. In Coalition Logic, we first quantify over  $J$ ’s possible choices, then over  $\bar{J}$ ’s.

More conceptually, we can consider a  $\mathcal{G}$ STIT model as a strategic game. A  $\mathcal{G}$ STIT model is a representation of a STIT moment that we already presented in Section 1.3 as a strategic game in which payoffs were abstracted away. Hence, roughly speaking, we need a modality which permits us to ‘jump’ from game to game.

Note that we have not proved that  $\mathcal{G}STIT$  was not able to simulate Coalition Logic; just that a natural translation does not do job. We nevertheless conjecture that there is no such a translation.

#### 4.4.2 Normal Simulation of Coalition Logic (NCL)

Normal Simulation of Coalition Logic (NCL) extends  $\mathcal{G}STIT$  with a discrete  $\mathbf{X}$  operator. Its syntax is given by the following grammar:

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

**Axiomatics.** We give to NCL the principles of  $\mathcal{G}STIT$  plus *ad hoc* axiom schemas as presented on Figure 4.2. As in  $\mathcal{G}STIT$  we moreover have the

$\mathcal{G}STIT$	axioms of $\mathcal{G}STIT$
+	
$\text{Triv}([\mathcal{A}gt])$	$\varphi \rightarrow [\mathcal{A}gt]\varphi$
$\mathbf{K}(\mathbf{X})$	$\mathbf{X}(\varphi \rightarrow \psi) \rightarrow (\mathbf{X}\varphi \rightarrow \mathbf{X}\psi)$
$\mathbf{D}(\mathbf{X})$	$\mathbf{X}\varphi \rightarrow \neg\mathbf{X}\neg\varphi$
$\text{Det}(\mathbf{X})$	$\neg\mathbf{X}\neg\varphi \rightarrow \mathbf{X}\varphi$

Figure 4.2: Axiomatics of NCL.

standard inference rules of modus ponens and necessitation for  $[\emptyset]$ . We also add necessitation for  $\mathbf{X}$ . Note that the converse of  $\text{Triv}(\mathcal{A}gt)$  is obtained by  $S5(\mathcal{A}gt)$ . Hence, we have  $\vdash_{\text{NCL}} \varphi \leftrightarrow [\mathcal{A}gt]\varphi$ .

From  $\mathbf{K}(\mathbf{X})$ ,  $\mathbf{X}$  is a normal modality. It is serial ( $\mathbf{D}(\mathbf{X})$ ) and deterministic ( $\text{Det}(\mathbf{X})$ ). It is naturally a component that captures discreteness. In addition,  $\text{Triv}([\mathcal{A}gt])$  grasps that the outcome is determined by the coalition  $\mathcal{A}gt$ .

Discreteness and determinism are two assumption that will be instrumental in order to embed Coalition Logic.

**Semantics.** We define NCL-models and state the determination of NCL with respect to them.

**Definition 4.9.** An NCL-model is a tuple  $\mathcal{M} = (W, R, F_X, \pi)$  where:

- $(W, R, \pi)$  is a  $\mathcal{G}STIT$  model, further constrained by  $R_{\mathcal{A}gt} = Id$ ;
- $F_X : W \rightarrow W$  is a total function;

An example of NCL-model is given on Figure 4.3.

Truth conditions are:

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

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

**Theorem 4.5.** *NCL is determined by the class of NCL-models.*

PROOF. Soundness is obtained by a routine argument and completeness is immediate from Sahlqvist theorem. ■

**Preliminary facts.** In the rest of the section we prove two lemmas that are useful for the sequel. We identically could have proved them within  $\mathcal{G}\text{STIT}$ , but are simply intended as preliminaries to our embedding of CL into NCL.

**Lemma 4.2.**  $\vdash_{\text{NCL}} \langle \emptyset \rangle \varphi \rightarrow \langle J_1 \rangle \langle J_2 \rangle \varphi$  if  $J_1 \cap J_2 = \emptyset$ .

PROOF. By  $\text{Elim}(\emptyset)$  we have  $\vdash_{\text{NCL}} \langle \emptyset \rangle \varphi \rightarrow \langle J_1 \rangle \langle \overline{J_1} \rangle \varphi$ . Now by hypothesis  $J_1 \cap J_2 = \emptyset$ , or equivalently  $J_2 \subseteq \text{Agt} \setminus J_1$ . Thus by (Mon)  $\vdash_{\text{NCL}} \langle \overline{J_1} \rangle \varphi \rightarrow \langle J_2 \rangle \varphi$ . We obtain  $\vdash_{\text{NCL}} \langle J_1 \rangle \langle \overline{J_1} \rangle \varphi \rightarrow \langle J_1 \rangle \langle J_2 \rangle \varphi$  by standard modal principles for  $[J_1]$ . We conclude that  $\vdash_{\text{NCL}} \langle \emptyset \rangle \varphi \rightarrow \langle J_1 \rangle \langle J_2 \rangle \varphi$ . ■

In [BPX01, Chap. 17] the authors provide an axiomatics of the theories of deliberative STIT in terms of a family of axiom schemas (AIA<sub>k</sub>). (Cf. Section 3.2.2.) It captures the central idea of the STIT theories that agents are independent. We can show a theorem of NCL which generalizes (AIA<sub>1</sub>) from individuals to coalitions, and that will be instrumental later in the proof of superadditivity in Theorem 4.7.

**Lemma 4.3.**  $\vdash_{\text{NCL}} \langle \emptyset \rangle [J_0] \varphi_0 \wedge \langle \emptyset \rangle [J_1] \varphi_1 \rightarrow \langle \emptyset \rangle ([J_0] \varphi_0 \wedge [J_1] \varphi_1)$  for  $J_0 \cap J_1 = \emptyset$ .

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

1.  $\langle \emptyset \rangle [J_0] \varphi_0 \rightarrow \langle J_1 \rangle \langle J_0 \rangle [J_0] \varphi_0$  by Lemma 4.2
2.  $\langle \emptyset \rangle [J_0] \varphi_0 \rightarrow \langle J_1 \rangle [J_0] \varphi_0$  from 1 by S5([J<sub>0</sub>])

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

■

**Theorem 4.6 ([Swa07, BGH<sup>+</sup>07]).** *The problem of deciding the satisfiability of a formula of NCL is PSPACE-complete in the mono-agent case and NEXPTIME-complete with at least two agents.*

## 4.5 STIT embraces Coalition Logic in the realm of normal modal logics

In this section, we prove that Coalition Logic can be embedded in NCL. We give the following translation from Coalition Logic to NCL.

$$\begin{aligned} tr_1(p) &= p \\ tr_1(\langle J \rangle \varphi) &= \langle \emptyset \rangle [J] \mathbf{X} tr(\varphi) \end{aligned}$$

and homomorphic for classical connectives.

**Theorem 4.7.** *If  $\varphi$  is a theorem of CL then  $tr_1(\varphi)$  is a theorem of NCL.*

PROOF. First, the translations of the CL axiom schemas are theorems of NCL.

- $tr_1(\neg \langle J \rangle \perp) = \neg \langle \emptyset \rangle [J] \mathbf{X} \perp$   
By D( $\mathbf{X}$ ),  $\vdash_{\text{NCL}} \mathbf{X} \perp \leftrightarrow \perp$ . By S5( $[J]$ ),  $\vdash_{\text{NCL}} [J] \perp \rightarrow \perp$ . It remains to prove that  $\vdash_{\text{NCL}} \neg \langle \emptyset \rangle \perp$ , which follows from S5( $[\emptyset]$ ).
- $tr_1(\langle J \rangle \top) = \langle \emptyset \rangle [J] \mathbf{X} \top$   
By K( $\mathbf{X}$ ),  $\vdash_{\text{NCL}} \mathbf{X} \top \leftrightarrow \top$ . By S5( $[J]$ ),  $\vdash_{\text{NCL}} [J] \top \leftrightarrow \top$ . Finally, by S5( $[\emptyset]$ ),  $\vdash_{\text{NCL}} \langle \emptyset \rangle \top$ .

- $tr_1(\neg\langle\emptyset\rangle\neg\varphi \rightarrow \langle\mathcal{Agt}\rangle\varphi) = \neg\langle\emptyset\rangle[\emptyset]\mathbf{X}\neg tr_1(\varphi) \rightarrow \langle\emptyset\rangle[\mathcal{Agt}]\mathbf{X}tr_1(\varphi)$ .  
As  $\vdash_{\text{NCL}} [\mathcal{Agt}]\psi \leftrightarrow \psi$  by  $\text{Triv}(\mathcal{Agt})$ , and as  $\vdash_{\text{NCL}} \langle\emptyset\rangle[\emptyset]\psi \leftrightarrow [\emptyset]\psi$  by  $\text{S5}([\emptyset])$ , the translation of  $(N)$  is equivalent to  $\neg[\emptyset]\mathbf{X}\neg tr_1(\varphi) \rightarrow \langle\emptyset\rangle\mathbf{X}tr_1(\varphi)$ . This is again equivalent to  $\langle\emptyset\rangle\neg\mathbf{X}\neg tr_1(\varphi) \rightarrow \langle\emptyset\rangle\mathbf{X}tr_1(\varphi)$  which is proved a theorem from  $\text{Det}(\mathbf{X})$ .
- $tr_1(\langle[J]\rangle(\varphi \wedge \psi) \rightarrow \langle[J]\rangle\psi) = \langle\emptyset\rangle[J]\mathbf{X}(tr_1(\varphi) \wedge tr_1(\psi)) \rightarrow \langle\emptyset\rangle[J]\mathbf{X}tr_1(\psi)$   
 $\mathbf{X}(tr_1(\varphi) \wedge tr_1(\psi)) \rightarrow \mathbf{X}tr_1(\psi)$  by  $\text{K}(\mathbf{X})$ . We have  $\vdash_{\text{NCL}} \langle\emptyset\rangle[J]\mathbf{X}(tr_1(\varphi) \wedge tr_1(\psi)) \rightarrow \langle\emptyset\rangle[J]\mathbf{X}tr_1(\psi)$  by standard modal principles for  $[J]$  and  $[\emptyset]$ .
- $tr_1(\langle[J_1]\rangle\varphi \wedge \langle[J_2]\rangle\psi \rightarrow \langle[J_1 \cup J_2]\rangle(\varphi \wedge \psi)) = \langle\emptyset\rangle[J_1]\mathbf{X}tr_1(\varphi) \wedge \langle\emptyset\rangle[J_2]\mathbf{X}tr_1(\psi)$   
 $\rightarrow \langle\emptyset\rangle[J_1 \cup J_2]\mathbf{X}(tr_1(\varphi) \wedge tr_1(\psi))$ 
  1.  $\langle\emptyset\rangle[J_1]\mathbf{X}tr_1(\varphi) \wedge \langle\emptyset\rangle[J_2]\mathbf{X}tr_1(\psi) \rightarrow \langle\emptyset\rangle([\mathbf{X}tr_1(\varphi) \wedge \mathbf{X}tr_1(\psi)])$ .  
by Lemma 4.3
  2.  $[\mathbf{X}tr_1(\varphi) \wedge \mathbf{X}tr_1(\psi)] \rightarrow [\mathbf{X}(tr_1(\varphi) \wedge tr_1(\psi))]$ .  
by (Mon)
  3.  $\langle\emptyset\rangle([\mathbf{X}tr_1(\varphi) \wedge \mathbf{X}tr_1(\psi)]) \rightarrow \langle\emptyset\rangle([J_1 \cup J_2](\mathbf{X}tr_1(\varphi) \wedge \mathbf{X}tr_1(\psi)))$   
from previous line by standard modal principles
  4.  $\langle\emptyset\rangle[J_1]\mathbf{X}tr_1(\varphi) \wedge \langle\emptyset\rangle[J_2]\mathbf{X}tr_1(\psi) \rightarrow \langle\emptyset\rangle[J_1 \cup J_2]\mathbf{X}(tr_1(\varphi) \wedge tr_1(\psi))$   
from lines 1 and 3 by standard modal principles for  $\mathbf{X}$ .

Second, clearly the translation of modus ponens preserves validity. To prove that the translation of  $\text{CL}$ 's  $(RE)$  preserves validity suppose  $tr_1(\varphi \leftrightarrow \psi) = tr_1(\varphi) \leftrightarrow tr_1(\psi)$  is a theorem of  $\text{NCL}$ . We have to prove that  $tr_1(\langle[J]\rangle\varphi \leftrightarrow \langle[J]\rangle\psi) = \langle\emptyset\rangle[J]tr_1(\varphi) \leftrightarrow \langle\emptyset\rangle[J]tr_1(\psi)$  is a theorem of  $\text{NCL}$ . This follows from the theoremhood of  $tr_1(\varphi) \rightarrow tr_1(\psi)$  by standard modal principles.  $\blacksquare$

**Lemma 4.4.** *Let  $M = ((S, E), V)$  a coalition model and  $selec : S \rightarrow S$  a mapping such that if  $selec(s) = s'$  then  $\{s'\} \in \mu_E(\mathcal{Agt}, s)$ .<sup>2</sup> Let  $\mathcal{M} = (W, R, F_X, \pi)$  be constructed as follows:*

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

<sup>2</sup>Such a function exists by the axiom of choice.

- $\pi(\langle s, s' \rangle) = V(s)$

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

PROOF. The proof consists in checking that the constructed model satisfies every constraint on NCL models. Everything is almost immediate. The main point is that we are permitted to define  $F_X$  this way w.l.o.g. because of Proposition 4.1. ■

**Theorem 4.8.** *If  $\varphi$  is CL-satisfiable then  $tr_1(\varphi)$  is NCL-satisfiable.*

PROOF. Given a coalition model  $M = ((S, E), V)$  we construct an NCL-model  $\mathcal{M}_{NCL} = (W, R, F_X, \pi)$  for some mapping *selec* as in Lemma 4.4. We prove by structural induction that  $M, s \models \varphi$  iff there is a  $\langle s, s' \rangle \in W$  s.t.  $\mathcal{M}_{NCL}, \langle s, s' \rangle \models tr_1(\varphi)$ .

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

1. Suppose,  $M, s \models \langle J \rangle \psi$ . Then, there is  $Z' \in E_s(J)$  such that for all  $t \in Z', M, t \models \psi$ . Then there is a minimal effectivity outcome  $Z \in \mu_E(J, s)$  such that for all  $t \in Z, M, t \models \psi$ . By induction hypothesis, there is a  $\langle s, s' \rangle$  such that  $\mathcal{M}_{NCL}, \langle s, s' \rangle \models tr_1(\psi)$ .
2. By construction,  $F_X(\langle s, y \rangle) = \langle y, selec(y) \rangle$ , for all  $t \in Z$  (the  $Z$  in item 1) and  $\{y\} \in \mu_E(\mathcal{A}gt, s)$  such that  $\{y\} \subseteq Z$ .
3. By (1) and (2) it follows that for all  $\{y\} \in \mu_E(\mathcal{A}gt, s)$  such that  $\{y\} \subseteq Z$ ,  $\mathcal{M}_{NCL}, \langle s, y \rangle \models \mathbf{X}tr_1(\psi)$ , and thus, since  $Z \in \mu_E(J, s)$ , it follows that there is  $\{y\} \subseteq Z$  such that  $\mathcal{M}_{NCL}, \langle s, y \rangle \models [J]\mathbf{X}tr_1(\psi)$ .
4. Finally, there is  $\langle s, y \rangle \in W$  such that  $\mathcal{M}_{NCL}, \langle s, y \rangle \models \langle \emptyset \rangle [J]\mathbf{X}tr_1(\psi)$ .

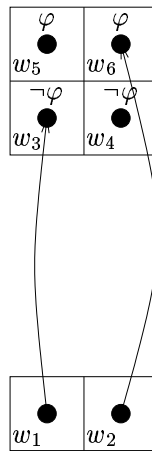
The other direction of the induction hypothesis is verified by reverse arguments. ■

**Corollary 4.1.**  *$\varphi$  is a theorem of CL iff  $tr_1(\varphi)$  is a theorem of NCL.*

PROOF. The right-to-left direction is Theorem 4.7. The left-to-right direction follows from Pauly's completeness result for Coalition Logic and Theorem 4.8. ■

## 4.6 Expressiveness

Coalition Logic is basically a logic of ability, in the sense that its main operator formalizes sentences of the form “agent  $a$  is able to ensure  $\varphi$ ”. As we have seen, NCL embeds CL and is of course suitable for such kind of reasoning about abilities of agents and coalitions. However, the introduction of a STIT-style operator is a move to more expressivity.



**Figure 4.3:** Representation of a NCL-model with two moments and two agents:  $a$  chooses the columns ( $R_{\{a\}}$ ) and  $b$  chooses the rows ( $R_{\{b\}}$ ). The grand coalition can determine a unique outcome:  $R_{\{a,b\}} = Id$ , is represented by the ‘small squares’. Nature ( $\emptyset$ ) cannot distinguish outcomes of a same moment:  $R_{\emptyset} = R_{\{a\}} \circ R_{\{b\}}$  is represented by the ‘big boxes’. Arrows are  $F_X$  transitions.

We attempt to throw some philosophical interest of NCL. Authors in logics of action have often been interested in the notion of ‘making do’. It can be linked to the idea of an agent having the *power over* another agent [Cas03]. On the model of Figure 4.3, it is easy to check that the formula  $\langle \emptyset \rangle [\{a\}] \mathbf{X} [\{b\}] \varphi$  is satisfied at  $w_1$  and  $w_2$ . A direct reading of this formula is “agent  $a$  sees to it that next, agent  $b$  sees to it that  $\varphi$ ”.

For example, in an organizational or normative setting, it in fact reflects adequately the agentive component of a *delegation*. As an illustration of our logic, we see here how NCL can grasp tighter notions of ability than Coalition Logic.

Without  $\mathbf{X}$ , Chellas’s STIT logic has some annoying properties: if we try to model influence of an agent on an other, we are inclined to state it

via the formula  $[a][b]\varphi$ . It is nevertheless equivalent to  $\Box\varphi$ . Hence, in this logic, an agent can force another agent to do something if and only if this something is settled. We must admit this is a poor notion of influence.

In our previous attempts to extend straightforwardly the logic of Chellas's stit with a 'next' operator ([BHT06c]), the formula  $[\{a\}]\mathbf{X}[\{b\}]\varphi \rightarrow \mathbf{X}\Box\varphi$  was valid. It means that if  $a$  forces that next  $b$  ensures  $\varphi$  then next,  $\varphi$  is inevitable. Inserting an  $\mathbf{X}$  operator between the agent's actions gives us a refined notion of influence. Still, it is not completely satisfying, since it suggests that an agent influences another agent  $b$  to do  $\varphi$  by forcing the world to be at a moment where  $\varphi$  is settled. Since an agent at a moment sees to everything being historically necessary (in formula: for every  $a$ ,  $\Box\varphi \rightarrow [a]\varphi$ ), it means that an agent  $a$  influences an agent to do  $\varphi$  if and only if it influences every agent to do  $\varphi$ ,  $a$  included.

On the contrary, the following formula is not a theorem of NCL:

$$[\{a\}]\mathbf{X}[\{b\}]\varphi \rightarrow \mathbf{X}[\emptyset]\varphi.$$

In particular in the model of Figure 4.3, the following formulas are true at  $w_1$  and  $w_2$ :

- $\langle \emptyset \rangle [\{a\}]\mathbf{X}[\{b\}]\varphi$
- $\langle \emptyset \rangle [\{a\}]\mathbf{X}[\{b\}]\neg\varphi$

It somewhat grasps the fact that agent  $a$  controls the truth value of  $\varphi$  by exerting influence on  $b$ . An interesting account of similar concepts but focused on propositional control is given by [vdHW05]. Of course, CL 'fused' operator is not designed for those issues, and Coalition Logic is not suitable for modeling the notion of *power over*.

Even though our quick study does not permit to prove that NCL is indeed a good logic to reason about influence, we think that the consistency of  $\langle \emptyset \rangle [\{a\}]\mathbf{X}([\{b\}]\varphi \wedge \langle \emptyset \rangle [\{b\}]\neg\varphi)$  which is at first sight a drawback, is in fact an interesting property: an agent can force an agent  $b$  to ensure  $\varphi$  even if  $b$  would also be able to ensure  $\neg\varphi$ . It somewhat leaves some place to indeterminism and unsuccessful delegations. What should constrain a delegated agent is not physics but norms. If one wants to rule that property out, one could simply release  $\text{Det}(\mathbf{X})$  and add the axiom schema  $\mathbf{X}\varphi \rightarrow \mathbf{X}[\emptyset]\varphi$ . The nature of time in NCL is simply a very convenient one for embedding CL and is amenable at will. We believe it particularly deserves a work effort in the future. See our conclusive discussion in Section 8.3.

As we will see in the next section for knowledge, the versatility of NCL models allows for smoothness in modeling. The information ‘contained’ in a context, viz. the physical description of the world *and* the actual strategy profile of agents permits to capture fine-grained notions relevant for multiagent systems via Kripke models in the realm of normal modal logics.

## 4.7 Seeing to it under imperfect knowledge

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

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

*“for all epistemically indistinguishable states, there exists a strategy of  $J$  that leads to  $\phi$ ”.*

from

*there exists a strategy  $\sigma$  of the coalition  $J$  such that for all states epistemically indistinguishable for  $J$ ,  $\sigma$  leads to  $\phi$ .”*

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

In [BHT06a] we sketched how the problem can be solved in a STIT-extension of ATL we called ATL-STIT. In the present setting we are only

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

### 4.7.1 Epistemic NCL (ENCL)

**Syntax.** ENCL extends NCL with one epistemic operator  $K_i$  for every agent. Its syntax is given by the following grammar:

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

The logic is obtained by adding to NCL the principles of the standard epistemic logic S5 for every individual agent  $i$ , and pictured in Figure 4.4. As in NCL we moreover have the standard inference rules of modus po-

NCL	axioms of NCL
+	
S5( $K_i$ )	S5-axioms for $K_i$

*Figure 4.4: Axiomatics of ENCL.*

nens, and necessitation for  $[\emptyset]$  and  $\mathbf{X}$ . We also add necessitation for every  $K_i$ .

**Semantics.** ENCL-models simply extend those of NCL with a collection of relations for agents’ uncertainty.

**Definition 4.10.** An ENCL-model is a tuple  $\mathcal{M} = (W, R, F_X, \sim, \pi)$  where:

- $(W, R, F_X, \pi)$  is a model of NCL.
- $\sim$  is a collection of equivalence relations  $\sim_i$  (one for every agent  $i \in \text{Agt}$ ).

**Theorem 4.9.** ENCL is determined by the class of models of ENCL.

PROOF. Again this is a immediate from Salqvist’s theorem. ■

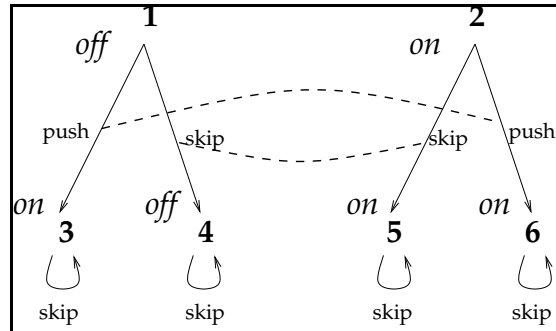
### 4.7.2 Reasoning about uniform strategies

The four basic properties we want to grasp are the following:

$\phi_1 =$	<i>One of Ann's choices ensures the light will be on</i>
$=$	$\langle \emptyset \rangle [\{Ann\}] \mathbf{X} \text{ on}$
$\phi_2 =$	<i>Ann knows one of her choices ensures the light will be on</i>
$=$	$K_{Ann} \langle \emptyset \rangle [\{Ann\}] \mathbf{X} \text{ on}$
$\phi_3 =$	<i>Ann knows she can conformantly see to it that the light is on</i>
$=$	$\langle \emptyset \rangle K_{Ann} [\{Ann\}] \mathbf{X} \text{ on}$
$\phi_4 =$	<i>Ann conformantly sees to it that the light is on</i>
$=$	$K_{Ann} [\{Ann\}] \mathbf{X} \text{ on}$

To explain how ENCL grasps these four properties and then solves the problem of uniform strategies, we present two toy scenarii, and encode them in ENCL. We first present a witness one, where the agent can indeed be said to have a uniform strategy for something. The challenge will be to prove that ENCL can distinguish it from a similar scenario where the agent does not have a uniform strategy. It will be the purpose of the second example.

**Example 4.1.** *Ann is in a room. She is blind and cannot distinguish a world where the light is off from a world where the light is on. The light in the room is controlled by a button that activates a timer. When the button is pushed the light bulb will shine for a determinate time. When the light is on, there is no way to switch it off. Ann can also do nothing (skip). In the actual situation the light is off and Ann is pushing the button.*



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

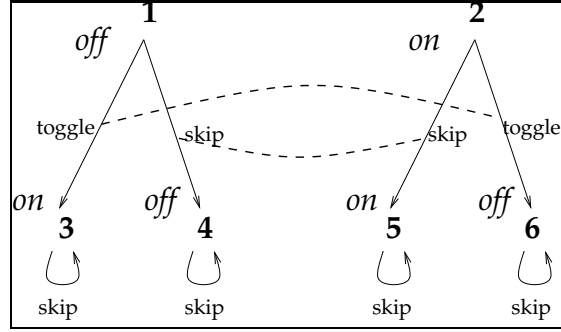
- $W = \{(1, p), (1, s), (2, s), (2, p), (3, s), (4, s), (5, s), (6, s)\}$
- $R_\emptyset = \{\langle(1, p), (1, s)\rangle, \langle(2, s), (2, p)\rangle, \langle(3, s), (3, s)\rangle, \langle(4, s), (4, s)\rangle, \langle(5, s), (5, s)\rangle, \langle(6, s), (6, s)\rangle\}^*$
- $R_{Ann} = \{\langle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

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

It is easy to verify that in  $\mathcal{M}_1$  the first three formulas are true in the first four possible ENCL worlds:  $\mathcal{M}_1, w \models \phi_1 \wedge \phi_2 \wedge \phi_3$  for all  $w \in \{(1, p), (1, s), (2, s), (2, p)\}$ . In particular, in the actual world  $(1, p)$  the third property holds, saying that Ann has a uniform strategy to ensure the light is on. In the actual world also the fourth property holds ( $\mathcal{M}_1, (1, p) \models \phi_4$ ), while in the two worlds where Ann skips, it does not ( $\mathcal{M}_1, (1, s) \not\models \phi_4$  and  $\mathcal{M}_1, (2, s) \not\models \phi_4$ ).

The point now is to show that our framework is able to distinguish the above scenario where Ann had a uniform strategy to put the light on, from another scenario where she cannot.

**Example 4.2.** *Ann is in a room. She is blind and cannot distinguish a world where the light is off from a world where the light is on. The light in the room is controlled by a switch. In her repertoire of actions, Ann can toggle or remain passive (skip), which correspond to switching the state of the light and maintaining the state of the light, respectively. In the actual situation the light is off and Ann toggles.*



This example is encoded by the following ENCL-model  $\mathcal{M}_2 = \langle W, R, F_X, \sim, \pi \rangle$ :

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

Now, in the actual world where the light is off and Ann toggles, the light will actually be on, so the formula  $\mathbf{Xon}$  holds. Yet, Ann does not conformantly see to it that the light is on, since she does not know that the light is off at the present moment. So, the fourth of the above properties does not hold:  $\mathcal{M}_2, (1, t) \not\models \phi_4$ . Also, she does not have a uniform strategy, and indeed the third of the above properties does not hold either:  $\mathcal{M}_2, (1, t) \not\models \phi_3$ . The first and the second property do hold in the actual world, since in each state Ann indeed has an action that ensures the light is on and she knows that: But her problem is that the decision which one to take depends on the state she is in, which is something she does not know:  $\mathcal{M}_2, w \models \phi_1 \wedge \phi_2$  for all  $w \in \{(1, t), (1, s), (2, s), (2, t)\}$ .

### 4.7.3 Discussion

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

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

## 4.8 Concluding remarks

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

A natural investigation concerns the introduction of group knowledge in the picture. In particular the integration of *common knowledge* is a worth challenge: some authors, as Aumann, would say that our system is obsolete without it. It is easy to import to NCL the principles of common knowledge. Nevertheless, axiomatizing common knowledge involve axioms that can not be directly given corresponding first-order semantic conditions. Thus, completeness of the resulting logic does not entail straightforwardly as it did with standard epistemic logic.

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

Last but not least, the clear objective is to extend this work to extensive forms of games. In the remaining of this dissertation, we try to understand the mechanisms of agency over time.

# 5

---

## Agency in branching time

### 5.1 Introduction

We feel the need to take an analysis of STIT temporal structures seriously. In the precedent chapters, we have focused on the CSTIT and its extensions. We showed that the Chellas stit operator was suitable for reasoning about choices of individuals and coalitions. To make a parallele with Game Theory, we have been till now only concerned with strategic games, a STIT moment been analogous to a game in normal form. (See [BPX01, Sect, 10C.2].) In the first two sections of this chapter, we show that the ontological commitment of CSTIT is nevertheless poor with respect the information content in the  $BT + AC$  structures.

We show in Section 5.2 that contrary to what can be expected, the logic of Chellas's stit is able to capture some temporal aspects, even though very poorly. It nevertheless reveals that CSTIT is not an adequate logic for formalizing interesting principles of agency. Then in Section 5.3 we reveal a syntactic relationship between operators of STIT (achievement, deliberative and Chellas's stit). We also see that Chellas's original logic of the operator of agency  $\Delta_a\varphi$  can be now better captured. It intimates that Chellas's stit misconceives the proposition of Chellas. In Section 5.4, we remark that in some circumstances, not being able to explicitly refer to *actions* remains a weakness. It deals with a modal logic allowing to reason about choices of agents and actions with duration controlled by their agent. Although unexpected, this move to the integration of explicit actions in the STIT theory permits us to reveal what we consider as hidden assumptions of original models. It nourishes the further analysis in Chapter 7 of an ontology of action.

## 5.2 Time in CSTIT

In the next chapter we will use the STIT semantics, that is  $BT+AC$  models as a basis of our ontology of action. We have seen that Xu's axiomatization of deliberative STIT theories,  $Ldm$ , determines this class of models. We have considered this theory as a theory of choice in earlier chapters. Because we are interested in a framework in which we can handle complex interaction between time and agency, we preliminary investigate in this section which aspects of time are borrowed by this theory.

### 5.2.1 Some temporal order remains...

Ming Xu did not axiomatize the deliberative STIT theories with tense operators.

“It is surely very natural to combine dstit theory with indeterministic tense logic, especially when we consider *deliberative seeing to something* to be connected with what future will be like. In carrying out some basic technical work in dstit theory, however, we will use a formal language without tense operators, though we will use the historical necessity operator  $Sett$  :, as a primitive.” [BPX01, Chap. 17]

Hence, the language of  $Ldm$  does not furnish tense operators. Nonetheless, it does not rule out the existence of a temporal order in the models characterized by  $Ldm$ . It is indeed true that the language confine us in a unique moment with no possibility to ‘jump’ in time. However, the language is still expressive enough to speak about histories and thus, temporal order. For instance, the formula  $\Diamond\varphi \wedge \Diamond\neg\varphi$  which is a formula of  $Ldm$  can be satisfied in a model consisting of at least three moments  $w_1, w_2$  and  $w_3$  ordered such that  $w_1 < w_2$  and  $w_1 < w_3$ .

Nevertheless, models remain somewhat degenerated. We try here to give the intuition why we cannot describe in  $Ldm$  an elaborated flow of time.

**Definition 5.1.** Let a pointed model  $\mathcal{M} = \langle \langle W, <, Choice, v \rangle, w \rangle$ . We call the minimal model of  $\mathcal{M}$  the pointed model defined as  $\mathcal{M}' = \langle \langle W', <', Choice', v' \rangle, w' \rangle$  such that:

- $W' = \{w'\} \cup \{w_h \mid h \in H_w\}$ ,  $w' <' w_h$  for every  $h \in H_w$ ;
- for every  $a$ ,  $Choice'_a{}^{w'} = Choice_a^w$  and for every  $a$  and  $h \in H_w$ ,  $Choice_a^{w_h} = \{\{w', w_h\}\}$ ; and

- $v'(w'/h) = v(w/h)$  and arbitrarily  $v'(w_h/h) = \emptyset$ .<sup>1</sup>

**Proposition 5.1.** *Let a pointed model  $\mathcal{M} = \langle \langle W, <, Choice, v \rangle, w \rangle$  and its minimal model  $\mathcal{M}' = \langle \langle W', <', Choice', v' \rangle, w' \rangle$ . Then  $\mathcal{M}, w/h \models \varphi$  iff  $\mathcal{M}', w'/\{w', w_h\} \models \varphi$ .*

In plain English, any satisfiable formula of *Ldm* admits a model whose temporal structure consists of a root moment  $w'$  plus one moment  $w_h$  for every  $h$  passing through  $w'$ , such that  $w' < w_h$  for every  $h$ . The basic language of *Ldm*, even though it does not permit tense statements, does not entirely rule out temporal aspects. Nevertheless, the class of minimal models suffices to interpret every formula of *Ldm*, and we can wonder whether some assumptions on time of the theory of agents and choices in branching time still make sense.

## 5.2.2 ... but does not capture fine-grained time

If some features of the temporal order remain, we can wonder which aspects of time *Ldm* does not capture. We show that some postulates on *BT + AC* models (see [BPX01, Appendix 3]) which are part of the philosophical justification of *STIT*, lose their relevance.

**No backward branching.** It is the case of the *no backward branching* principle to be no more relevant in those degenerated models. Indeed, it says that if  $w_1 \leq w_3$  and  $w_2 \leq w_3$  then  $w_1 \leq w_2$  or  $w_2 \leq w_1$  [BPX01].<sup>2</sup> But if we consider minimal models, and if we let trivial cases aside – e.g.  $w_1, w_2$  and  $w_3$  are not pairwise distinct – the root and only the root  $r$  can have a next moment  $x$  such that  $r \leq x$ . Indeed,  $w_h$  and  $w_{h'}$  for  $h \neq h'$  will never be comparable. So, there are three kinds of constraints that this principle is about when considered on minimal models:

1. if  $r \leq w_h$  and  $r \leq w_h$  then  $r \leq r$  or  $r \leq r$
2. if  $r \leq w_h$  and  $w_h \leq w_h$  then  $r \leq w_h$  or  $w_h \leq r$
3. if  $w_h \leq w_h$  and  $w_h \leq w_h$  then  $w_h \leq w_h$  or  $w_h \leq w_h$

which are all commonplace.

<sup>1</sup>We consider valuations  $v$  such that  $v : W \times Hist \mapsto 2^{Atm}$ , characterized from a standard valuation  $v_0 : Atm \mapsto 2^{W \times Hist}$  by: for every  $p \in Atm$ ,  $p \in v(w/h)$  iff  $w/h \in v_0(p)$ .

<sup>2</sup>As usual  $x \leq y$  iff  $x < y$  or  $x = y$ .

**No choice between undivided histories.** Analogously, the *no choice between undivided histories* principle gets irrelevant in those degenerated models where histories are required to contain at most two moments. Our argument follows from basic observations. In [BPX01]  $h$  and  $h'$  are said undivided at  $w$  iff  $w \in h \cap h'$  and there is  $w'$  such that  $w < w'$  and  $w' \in h \cap h'$ .<sup>3</sup> Hence, if we just consider minimal models, since every history is composed of at most two moments, that is, the root  $w'$  plus one single moment which is a leaf of the tree structure, they all are divided at  $w'$ . Thus, as soon as it respects the principle of independence of agents, for such models, everything goes when constructing the *Choice* function.

We regard this property as the most fundamental assumption of  $BT + AC$  models that is not enforced by  $CSTIT$ . However, this assumption is fundamental since it in some sense constrain the temporal structure of causality: an agent or a group of agents can choose between one or another history (and then try to force the time to run through one or the other) only if those histories are divided.

### 5.2.3 Chellas's stit is the brute choice component of agency

Some principles like influence or refraining is captured in an unsatisfying manner in  $CSTIT$ . As we have seen in Chapter 3, given two distinct agents  $a$  and  $b$ , the formula  $[a\text{ cstit} : [b\text{ cstit} : \varphi]]$  reduces to  $\Box\varphi$ . In other words, an agent can influence (or force) another agent to do  $\varphi$  if and only if  $\varphi$  is inevitable. Analogously,  $a$  refrains  $b$  of doing  $\varphi$  if and only if  $b$  cannot achieve  $\varphi$ :  $[a\text{ cstit} : \neg[b\text{ cstit} : \varphi]] \leftrightarrow \neg\Diamond[b\text{ cstit} : \varphi]$ .

This is directly inherited from independence of choice of agents: an agent cannot deprive another agent from its choices. It is a fair assumption whenever it concerns simultaneous choices, that is, without causal order. It is nicely argued by Belnap et al. in [BPX01, p. 218] that takes over a well known argument in games particularly important in the Prisoner's Dilemma and other toy scenarii of game theory:

“If there are agents whose simultaneous choices are not independent, so that the choice of one can “influence” what it is possible for the other to choose even without priority in the

---

<sup>3</sup>In [BPX01, Appendix 4] it is added “unless there is no  $w'$  such that  $w < w'$ ”. So far so good. But, if there is no such a  $w'$ , we have reached a leaf of the tree and just one history is going through it. It reveals that Belnap et al. assume a history and itself to be undivided, which is intuitive but of no use here, since the *Choice* function will impose a history and itself to lie in the same partition.

causal order, then we shall need to treat in the theory of agency a phenomenon just as exotic as those discovered in the land of quantum mechanics by Einstein, Podolsky, and Rosen.”

The only characteristic principle in CSTIT is the independence of agents. CSTIT in some sense is then particularly *ad hoc* for reasoning about a particular moment or, following our analogy of Section 1.3, a particular game with abstract utilities.

Hence, we are inclined to say that more specifically than a logic of agency, CSTIT is a logic of *brute choice*. Brute choice or material choice has to be understood as the ontological object containing the information of a choice and would just be a component of *rational choice*. A brute choice is exactly a set of histories that an agent (or a group of agents) has chosen or can choose for some reasons that are not part of the description.

We suggest that a natural reading of  $[a \text{ cstit} : \varphi]$  is “agent  $a$  chooses such that it sees to it that  $\varphi$ ”. Chellas’s stit describes an *action of choosing*. An alternative name of the operator could be *choice stit*. We see in Section 5.3.3 another argument for preferring it to its original name.

## 5.2.4 Causality in agency

The argument goes, and Belnap et al. add:

“We are in effect postulating that the only way that the choices open to one agent can depend on the choices open to another agent is if the one agent’s choices lie in the causal past of those of another agent.” ([BPX01, p. 218])

Logics of agency in philosophy of action are generally tailored to model causality via agent choices.

If we find ourself in a difficulty this is because the kind of agentive sentences denoted by  $[a \text{ cstit} : \varphi]$  refers to *instantaneous action*. Such actions are not generally accepted and philosophy of action tends to ignore them, if it does not explicitly consider them weird.

Chellas’s stit deals with actions whose result is confounded with their starting point. Some interpretations of time are in terms of actions of agents of the system. An immediate action cannot bring dynamics since somewhat the postconditions are exactly preconditions. Interpreting Chellas’s stit as an action of choosing brings some realism in the picture, as far as we accept that mental actions do not involve movement.

In Chapter 4, we have already used a ruse to overcome the problem. We artificially put some duration to the action denoted by the Chellas stit operator by adding a tense operator  $\mathbf{X}$ .  $[a\ cstit : \mathbf{X}\varphi]$  in some sense simulates the fact that the ‘current’ choice of agent  $a$  underlies an action of  $a$  that ends at the next step with  $\varphi$ . Nevertheless, it does not permit us to distinguish it from a choice underlying a longer action, but whose execution runs along histories for which  $\varphi$  is true at the next moment. It is impossible to measure the length of an action triggered by a choice. The obvious reason is that it is not because we add a tense operator in the scope of the Chellas stit that the Chellas’s stit is more than a simple operator of choice. The more appropriate reading of  $[a\ cstit : \mathbf{X}\varphi]$  to our ears is indeed “agent  $a$  chooses such that  $\varphi$  at the next step”, and does not permit to go further in its interpretation in terms of action than the observation that it underlies an action of choosing, or more concretely of an action of selecting a set of histories.

Achievement stit or even Chellas’s  $\Delta_a\varphi$  operator satisfy (T) axiom because it reflects their nature of success: the agent made a choice *previously* that makes sure that something holds now. Chellas’s stit also obeys the axiom (T) and hence is an operator of success. But not of causality in branching time. In part because CSTIT does not reflect the *no choice between undivided history*, it does not permit us to reason about the causal aspect of agency in branching time. In the next section we exhibit differences between Chellas’s stit and operators of causality. We identify and treat a related problem of *agentive gap* of Chellas’s stit in Chapter 7.

### 5.3 Measuring the length of an action

In this section, we investigate a discrete STIT framework that permits us to distinguish and clarify this different role that choice and causality play in time structure. Fundamental distinctions make that Chellas’s stit is different from an operator of causality.

We propose two new primitive operators that allow to characterize syntactically the Chellas, deliberative and achievement stit operators, but also Chellas’s original operator of agency  $\Delta_a\varphi$ . We show how it highlights their relationship and reveal differences. In particular, we remark that Chellas’s stit is not the more accurate simulation of Chellas’s original proposal  $\Delta_a\varphi$  [Che69, Che92]. A brief preliminary investigation of duration of agents’ activities is given.

### 5.3.1 A discrete STIT framework

What can be now of interest, is to understand the underlying link between the three main versions of the STIT operator, viz. Chellas's stit, deliberative stit and achievement stit. We have already seen that the deliberative stit can be defined from Chellas's plus historical necessity since the following holds:

$$[a \text{dstit} : \varphi] \leftrightarrow [a \text{cstit} : \varphi] \wedge \neg \Box \varphi$$

The other way round, we have  $[a \text{cstit} : \varphi] \leftrightarrow [a \text{dstit} : \varphi] \vee \Box \varphi$ . The link between deliberative and Chellas's stit is then quite obvious. However, a formal link of the achievement stit with them is more involved. We nevertheless claim that, because of the complex semantics of  $[_ \text{astit} : \_ ]$ , such a relationship can provide a neat picture of the fundamental aspects of the theory of choice in time. And in order to stick to the *fundamental* aspects, let us first simplify the framework by some usual assumptions. They at least are usual in a discipline like computer science, and have the merit to rule out some features that were enabled just for a matter of generality, and thus unfortunately hid some other essential features. Belnap and colleagues refrained from taking position on the nature of time.

“[...] no assumption whatsoever is made about the order type that all histories share with each other and with *Instant*. For this reason the present theory of agency is immediately applicable regardless of whether we picture succession as discrete, dense, continuous, well-ordered, some mixture of these, or whatever; and regardless of whether histories are finite or infinite in one direction or the other.” ([BPX01, p. 196].)

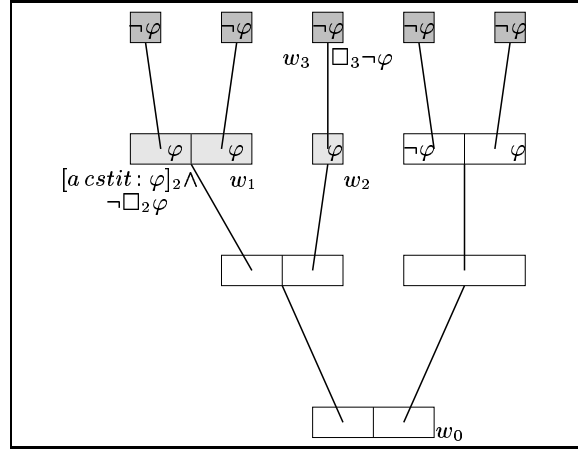
We thus consider the assumption of time isomorphic to the set of natural numbers interesting to study. We would like to investigate how such a simplification can strengthen our understanding of logics of agency. We define discreteness as follows:

**Definition 5.2.** *The total function  $\text{instantof} : W \rightarrow \mathbb{N}$  associates an instant with each moment.*

**Hypothesis 5.1.** *Histories are regular: (1)  $\forall h, h' \in \text{Hist}, \forall w \in h, \exists w' \in h', \text{s.t. } \text{instantof}(w) = \text{instantof}(w')$  (2) for some  $h \in \text{Hist}$  and  $w \in h$ , if  $\text{instantof}(w) = i$  then  $\forall j < i, \exists w' \in h \text{ s.t. } \text{instantof}(w') = j$ .*

Moreover, we assume the existence of a root:

**Hypothesis 5.2.** *There is a moment  $w$  such that there is no  $w'$  such that  $w' < w$ .*



**Figure 5.1:** (Time goes upward.) At  $w_0$ ,  $a$  can make the choice that  $\varphi$  is true in two steps, even though it is not settled it will be true at that instant. At  $w_1$  (or  $w_2$ ) it will be the case  $[a \text{ cstit} : \varphi]$ . Indeed, for some  $h \in w_1$ ,  $w_1/h \models [a \text{ cstit} : \varphi]$  ( $\varphi$  is true at every index – moment/history pair – of  $w_1$  and  $w_2$ ) and  $w_1/h \models \neg \Box_2 \varphi$ . At  $w_0$  it is however already settled that in three steps,  $\varphi$  will be false: for some  $h' \in w_3$ ,  $w_3/h' \models \Box_3 \neg \varphi$ . ( $\varphi$  is true at every (upper) dark grey moment.)

**NSTIT.** In order not to get confused let us call NSTIT the logic interpreted by  $BT + I + AC$  models constrained by the hypothesis previously presented, and syntactically extending the STIT theory presented in Chapter 2 (the *super-language* containing operators Chellas’s, deliberative and achievement stit) with the two following collections of operators indexed by a natural number  $k$ :

- $\mathcal{M}, w/h \models \Box_k \varphi \iff \exists w_0 \leq w, \text{instantof}(w_0) = \text{instantof}(w) - k, \forall w' \in \text{Instant}(w) \cap (\bigcup_{h' \in H_{w_0}} h'), \forall h' \in w', \mathcal{M}, w'/h' \models \varphi$   
It reads that “ $k$  instants ago, it was settled that  $\varphi$  would be true now”.
- $\mathcal{M}, w/h \models [a \text{ cstit} : \varphi]_k \iff \exists w_0 \leq w, \text{instantof}(w_0) = \text{instantof}(w) - k, \forall w' \text{Choice}_a^{w_0} - \text{equivalent of } w, \forall h' \in w', \mathcal{M}, w'/h' \models \varphi$   
It reads that “ $k$  instants ago, agent  $a$  ensured that  $\varphi$  would be true now”.

Analogously to the achievement stit, we call  $w_0$  in the previous truth conditions the *witness moment* of  $[a \text{ cstit} : \varphi]_k$  or  $\Box_k \varphi$ .<sup>4</sup>

We offer to NSTIT a mechanism close to what exists in Hybrid Logic [BdRV01, Chap. 7]. We assume the existence of a set  $\{0, 1, \dots\}$  of specific

<sup>4</sup>Recall that the *choice equivalence* is introduced in Definition 2.2.

atomic formulae that we could call *nominals*. We thus constrain the models such that  $\mathcal{M}, w/h \models i$  iff  $instant(w) = i$ . Our account is nevertheless different from Hybrid Logic since genuine nominals should be true at exactly one moment/history pair. See for example [BG01] for a concrete account of hybrid temporal logic.

Now, let us exhibit some interesting validities, candidates to the status of axioms for further developments.

$$(NP) \quad 0 \rightarrow \neg \Box_1 \top$$

$$(P) \quad 0 \vee \Box_1 \top$$

$$(Mon) \quad \Box_{k+1} \top \rightarrow \Box_k \top$$

$$(Sett-1) \quad \Box_k \varphi \rightarrow \Box_k \top$$

$$(Sett-2) \quad k \leftrightarrow \Box_k k$$

(NP) captures that there is no past beyond the instant 0. (P) on the contrary means that whenever we do not stand at instant 0 we can ‘step back’ in the temporal structure. (Mon) means that when we can look back at  $k + 1$  steps, we can look back at  $k$  steps as well. (Sett-1) says that  $k$  steps ago, it was settled that  $\varphi$  would be true now, only if we can look back at  $k$  steps. (Sett-2) means that we are standing at instant  $k$  iff it was already settled  $k$  steps ago that we would stand at instant  $k$  now.

We are now ready to see how the operators of the STIT language relate to our new primitives.

**Proposition 5.2.** *The following formulae are valid in NSTIT.*

- $\Box \varphi \leftrightarrow \Box_0 \varphi$
- $[a \text{ cstit} : \varphi] \leftrightarrow [a \text{ cstit} : \varphi]_0$
- $[a \text{ dstit} : \varphi] \leftrightarrow [a \text{ cstit} : \varphi]_0 \wedge \neg \Box_0 \varphi$
- $i \rightarrow ([a \text{ astit} : \varphi] \leftrightarrow \bigvee_{k=1}^i ([a \text{ cstit} : \varphi]_k \wedge \neg \Box_k \varphi))$

From the last item, we can have a *local* definition of achievement stit for every instant. It is indeed similar to the definition of tense operator ‘until’ and ‘since’. (See [BG01, Sect. 4.1].) Historical necessity, Chellas’s stit and deliberative stit on the other hand, can be completely defined from our new primitives.

Instances of the new primitive operator of agency are intrinsically related and obey the following property:

**Proposition 5.3.**  $[a\ cstit: \varphi]_{k_1} \rightarrow [a\ cstit: \varphi]_{k_2}$ , for every  $k_2 < k_1$ .

PROOF. This a corollary of Proposition 2.1. ■

### 5.3.2 Duration of an activity.

To give some intuitions behind  $[a\ cstit: \varphi]_k$  consider the following example. In an institutional context, it can be useful to reason about the length of an activity. For instance, given an operator for obligation  $\bigcirc$ , we could have a formula like

$$phd(Mary) \rightarrow \bigcirc[Mary\ cstit: Mary\_has\_written\_her\_thesis]_{24}$$

in the domain description, to state that a student can obtain a PhD only if he or she has achieved the writing of the dissertation and has spent *at least* 24 months working on it. (In France a minimum of 2 registrations is imposed.) From Proposition 5.3, it indeed captures the notion of minimum. In such a modeling, it is like Mary chose at least 24 ‘clock ticks’ ago (that happen here to correspond to months) to write a dissertation and it happens to have succeeded *now*.

### 5.3.3 Comments on Chellas’s $\Delta_a\varphi$

In [Che92], Brian Chellas turns back to his operator of agency introduced in [Che69]. As in theories of agents and choices in branching time, truth values of the language are in terms of times (alias instants), histories and agents, plus *certain* relations. Here, we quickly show how we can define  $\Delta_a\varphi$  fairly in NSTIT, and also suggest that Chellas’s stit operator does not match perfectly.

#### 5.3.3.1 Semantics of time and actional alternatives

The set of times is taken to be the set of integers. We write  $t < t'$  to state that  $t$  is earlier than  $t'$  and  $t \leq t'$  to state it is not later. Histories are functions from times to states of affairs (alias moments), and  $h(t)$  represents the state of affairs in history  $h$  at time  $t$ . Two time-indexed relations between histories are then defined.  $h =_t h'$  means that histories  $h$  and  $h'$  have the same past at time  $t$ ;  $h \equiv_t h'$  means that they share the same past *and* the same present. Formally,

- $h =_t h'$  iff  $h(t') = h'(t')$  at every time  $t' < t$

- $h \equiv_t h'$  iff  $h(t') = h'(t')$  at every time  $t' \leq t$

Given a state of affairs  $h_t$ , Chellas uses the term *future cone* for the set of histories emanating from  $h_t$ . Two histories are in the future cone of  $h(t)$  iff  $h \equiv_t h'$ .

**Instigative alternatives.** The relation  $R_t^a(h, h')$  is used to mean that  $h'$  is an *instigative alternative* of  $h$  for agent  $a$  at time  $t$ . The relation is reflexive and  $R_t^a(h, h')$  only if  $h =_t h'$ . Instigative alternatives capture the idea of histories “under the control” or “responsive to the action” of  $a$  at  $t$ .

Truth conditions of the operator of agency is given by:

$$(h, t) \models \Delta_a \varphi \iff (h', t) \models \varphi, \forall h' \text{ s.t. } R_t^a(h, h')$$

### 5.3.3.2 Chellas’s stit is not $\Delta_a \varphi$

In addition to our short overview, it is interesting and helpful to consider Krister Segerberg’s interpretation of the operator in [Seg92]. Segerberg calls  $R_t^a(h, h')$  the cone of ‘actional alternatives’ and observes that in the truth value of  $\Delta_a \varphi$ , “the cone Chellas wishes to consider has its apex at the immediately preceding time”. This is indeed a consequence of the constraint that two histories  $h$  and  $h'$  are instigative alternatives only if  $h =_t h'$ .

Finally, we can define more appropriately the operator in NSTIT as follows:

$$\Delta_a \varphi \triangleq [a \text{ cstit} : \varphi]_1$$

It thus clearly differs from  $[a \text{ cstit} : \varphi]$  which we remind is logically equivalent in NSTIT to  $[a \text{ cstit} : \varphi]_0$ . There is a temporal switch between them. We touch here one of the claims of this dissertation: one must be aware of a possible misconception of the Chellas’s stit, since it does not reflect Chellas’s original operator. If Chellas had in mind something similar to Chellas’s stit when he made up his  $\Delta_a \varphi$  operator, he would have constrained the instigative alternatives (or actional alternatives) such that  $R_t^a(h, h')$  only if  $h \equiv_t h'$ .

Still, it does not mean that  $[a \text{ cstit} : \varphi]_1$  is  $\Delta_a \varphi$  without nuance. Our definition also suffers what Horty and Belnap already pointed out about their Chellas stit: Chellas did not impose any constraint on independence of agents, while we inherit it from  $BT + AC$  structures. Analogously, Chellas did not assume a future branching only, while  $BT + AC$  structures are constrained by *histories connectedness* in the past.

### 5.3.4 Choice vs. causality in branching time

Now, those operators indexed with a natural number  $k$  may seem odd. But this is not odder than an iterated operator ‘next’ permitting to jump from instant to instant along a history. This is actually interesting to see what is going on if we allow such an operator:

$$\mathcal{M}, w/h \models \mathbf{X}\varphi \iff \exists w' w < w', \nexists w'' w < w'' < w', \text{ s.t. } \mathcal{M}, w'/h \models \varphi$$

In order to highlight how our primitive operators behave over time, it is easy to prove that  $[a \text{ cstit} : \mathbf{X}^k \varphi] \leftrightarrow \mathbf{X}^k [a \text{ cstit} : \varphi]_k$ , and  $\Box \mathbf{X}^k \varphi \leftrightarrow \mathbf{X}^k \Box_k \varphi$ . The former formula captures that  $a$  is choosing such that  $\varphi$  in  $k$  steps iff in  $k$  steps  $\varphi$  is settled true by one choice of  $a$ ,  $k$  steps before.

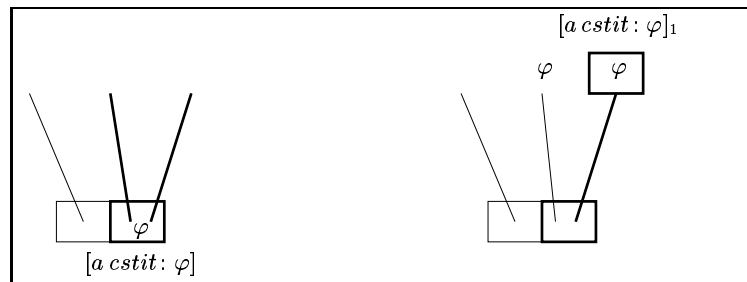


Figure 5.2: Schematic comparison between  $[a \text{ cstit} : \varphi]$  and  $[a \text{ cstit} : \varphi]_1$ .

Let us designate a *chain* as being a set of linearly ordered moments. “In branching time, chains represent certain complex concrete events” [BPX01, p. 181].

While in the original STIT theory the  $[_ \text{cstit} : \_ ]$  permits to express that an agent selects some set of histories (*unbounded* sets of ordered moments), underlying events are loosely characterized: they correspond to every chain we can construct on those histories. With  $[_ \text{cstit} : \_]_k$  we clearly identify the set of events the agent has brought about: events composed of moments between the moment of choice  $w$  and moments that are on the selected histories not farer than  $k$  instants after  $w$ . We see that as a strength of the language. On Figure 5.2, Chellas’s stit on the left side ‘represents’ some bundle of selected histories. On the other hand,  $[_ \text{cstit} : \_]_k$  marks the success of a *previous* choice and regarding the undeterministic nature of time. The former is an operator of choice while the latter marks a causation of an agent.

## 5.4 A modal view of actions with duration

### 5.4.1 Motivation

Many domains, e.g., agent interaction or social law modeling, require a good framework for time, agency and action. Time is the basis to express dynamic properties and indeterminacy of the future, agency deals with what agents can bring about and actions are the various ways to bring about some state of affairs. As far as we know, there is no multiagent system allowing to represent these three domains with sufficient expressivity. In particular, we intend to cover actions that have a duration, and that can be categorized on the basis of properties such as expected effects, temporal or participant structure.

In [Seg92], talking about Belnap and Perloff and the STIT theories, Segerberg writes: “Their work is probably one of the two most promising avenues of research in current logic of action.” According to Segerberg, the other one is Pratt’s Dynamic Logic.

Indeed, these two approaches to action answer to some extent to our needs. Concerning pure action, the well-known Propositional Dynamic Logic (PDL) is a natural candidate. Nevertheless, it is suitable neither for group action nor for individual and group agency. We have already seen that the logic of “Seeing To It That” is a logic of agency embedded in a branching time framework. This is a logic about choices and strategies for individuals and groups. The core idea of logics of agency is that *acting* is best described by what an agent brings about: at some time, an agent *chooses* to constrain that some proposition is true. However, in some circumstances, not being able to explicitly refer to *actions* remains a weakness. One expresses sentences of the form “Mary sees to it that the coyote is dead” but not “Mary shoots at the coyote” or “Mary poisons the coyote”, i.e., the manner of bringing a state of affairs is out of concern. In addition, in STIT, it is generally considered that Mary’s acting does not take time: actions cannot be suspended half-way and one cannot express that an action starts while another is going on. This last point has already been overcome in [Mül05] with the operator *istit*, but this logic still doesn’t involve actions explicitly.

It appears that we need a richer logical system, for reasoning about time, agency and actions with duration. One research avenue *from the point of view of modal logics* is to capitalize on strength of both PDL and STIT. In fact, we just follow the basis of the paradigm of Pratt’s Dynamic Logic, and overcome the problem of what agents “really do” by thinking of the action of an agent like a computer program. Hence, we stick to the

paradigm that tells us to label agents' actions, not to the famous PDL: e.g. we do not introduce complex mechanisms of iterated actions.

The aim of the remaining of this section is to investigate this avenue, offering an expressive logical framework to support time, agency (for individuals and groups) and actions with duration and other properties, for modeling interactions between agents.

## 5.4.2 A modal logic for actions with duration

We give the language of the new logic and describe some elements of its semantics intertwined with ontological justifications. Possible axioms or theorems are proposed in formulas labelled **(n)**.

**Language.**  $Act$  is a finite set of actions, and  $Act_\lambda := \{\alpha_\lambda \mid \alpha \in Act\}$  is the set of continuations of those actions.  $Atm$  is the set of atomic propositions.  $Agnt$  is the set of agents. By notational convention,  $\alpha \in Act$ ,  $\alpha_\lambda \in Act_\lambda$ ,  $\beta \in Act \cup Act_\lambda$ ,  $p \in Atm$ ,  $a \in Agnt$  and  $A \subseteq Agnt$ . A formula can have the following syntactic form:

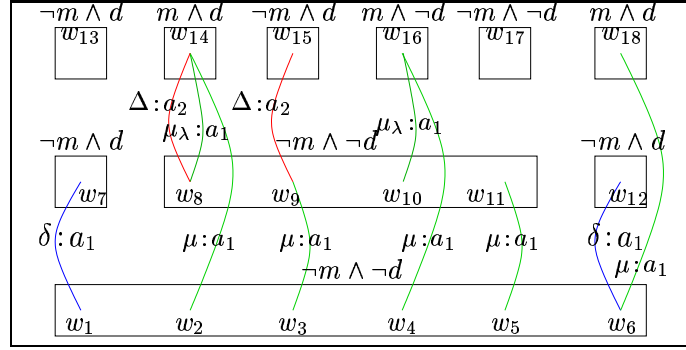
$$\varphi ::= \perp \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \mathcal{S} \varphi \mid \varphi \mathcal{U} \varphi \mid \Box\varphi \mid [\beta:a]\varphi \mid [\overline{\alpha:a}]\varphi$$

$\mathcal{S}$  and  $\mathcal{U}$  are the standard since and until temporal operators. Future and past operators are defined as usual:  $\mathbf{F}\varphi \equiv \top \mathcal{U} \varphi$  and  $\mathbf{P}\varphi \equiv \top \mathcal{S} \varphi$ .  $\Box\varphi$  stands for historical necessity of  $\varphi$ . ( $\varphi$  is true at every index of the moment.)  $\Box$  is an  $S5$  modal necessity constructed over  $R_\Box$  **(1)**. It corresponds to STIT's settledness or historical necessity.

As in PDL,  $[\beta:a]\varphi$  means that  $a$  starts performing action  $\beta$  and that  $\varphi$  holds afterwards. Its truth value is as usual for a necessity modality over  $R_{ACT}(\beta)$ , such that  $a = agent(\beta)$ .  $[\overline{\alpha:a}]\varphi$  means that an action  $\alpha$  performed by  $a$  has just finished and that  $\varphi$  was true before. It is a normal necessity modality over the converse relation  $R_{ACT}(\alpha)$ , such that  $a = agent(\alpha)$ . By tradition,  $\Diamond$  will be used as abbreviation for  $\neg\Box\neg$  and similarly for  $\langle\beta:a\rangle$  and  $[\beta:a]$ .

We will illustrate the logic by the toy example of Figure 5.3. We come back to it throughout the section when we introduce the properties of the logic.

A model is a tuple  $\mathcal{M} = \langle W, <, R_\Box, R_{ACT}, agent, v \rangle$ , where  $W$  is a set of indexes, partially ordered by the strict temporal precedence  $<$ . Non comparable indexes are grouped into *moments* by the equivalence relation  $R_\Box$ . For all  $\beta \in Act \cup Act_\lambda$ ,  $R_{ACT}(\beta)$  is a *partial* function associating an



**Figure 5.3:** The two cooks.  $a_1$  and  $a_2$  have to prepare a meal.  $m$  stands for “the main course is done” and  $d$  for “the dessert is done”.  $\delta$  and  $\Delta$  are actions of (resp.)  $a_1$  and  $a_2$  cooking the dessert, while  $\mu$  is the action of  $a_1$  cooking the main course. Boxes are moments containing indexes.

index  $w$  to the index where the performance at  $w$  of  $\beta$  ends,  $agent$  is a function associating to each action its agent,  $v$  is the valuation function.

We have slightly adapted the semantics of STIT to a  $\mathbb{T} \times \mathbb{W}$  setting where histories are hidden [vK86, vK97]. We can define an *history of an index*  $w$  as the set of indexes in its past or future:  $h(w) = \{w' \mid w' < w \text{ or } w < w' \text{ or } w' = w\}$ . We assume that models satisfy additional properties:

1.  $(w_2 < w_1 \text{ and } w_3 < w_1) \text{ or } (w_1 < w_2 \text{ and } w_1 < w_3)$  implies that  $w_2 \in h(w_3)$ . Time is linear along each history.
2.  $w_1 \in R_\square(w_2)$  implies that nor  $w_1 < w_2$  neither  $w_2 < w_1$ .
3. if  $w_1 < w_2$  and  $w_3 \in R_\square(w_2)$  then there is  $w_4 \in h(w_3)$  s.t.  $R_\square(w_1, w_4)$  and  $w_4 < w_3$ . Time on moments is branching towards the future.
4. for  $\beta \in Act \cup Act_\lambda$ ,  $w < R_{ACT}(\beta)(w)$ . All actions take time: they lead to a future index.
5. for  $\alpha \in Act$ , if  $R_{ACT}(\alpha)(w_1) = w_2$ , there is no  $w_3$  s.t.  $R_{ACT}(\alpha)(w_3)$  is defined and  $w_1 < w_3 < w_2$ . Two occurrences of the same action on the same history cannot overlap in time.
6. if  $R_{ACT}(\alpha_\lambda)(w_1) = w_2$ , it exists  $w_3 < w_1$  s.t.  $R_{ACT}(\alpha)(w_3) = w_2$ . A continuation depends on the current execution of its corresponding action.
7. if  $R_{ACT}(\alpha)(w_1) = w_2$  then for all  $w_3$  s.t.  $w_1 < w_3 < w_2$ ,  $R_{ACT}(\alpha_\lambda)(w_3) = w_2$ . An action is continued at each index between its starting and ending indexes.

We will further constrain the models in Section 5.4.4 when introducing the  $[_{cstit}: \_]$  operator.

### 5.4.3 Ontological justification and some validities

#### 5.4.3.1 Time

As in STIT, our logic assumes a branching time on moments, linear in the past: at each moment, an agent can make different choices, that is, decide to execute different actions bringing about different futures. Maximal linear sets of moments are called histories; indexes can be seen as moment-history pairs. Models do not constrain all moments to be temporally comparable; an extension could be to do so, adding coincidence of moments through the notion of *instant* we have presented in Section 2.4.

#### 5.4.3.2 Actions

All actions take time:

$$\bullet \langle \beta : a \rangle \varphi \rightarrow \mathbf{F}\varphi \quad (2)$$

Actions in the present logic are operators thus not properly speaking “events performed by an agent” since events, when they are acknowledged as citizens of the world, are conceived of as concrete individuals, uniquely situated in time and space [CV96]. These operators correspond to *types*, not tokens, as a given action may occur repeatedly. They are thought of a very restricted sort of types. The agent, as well as all other participants, are fixed. The only remaining parameter is time.

Actions correspond to achievements and accomplishments [Ven67], thus two occurrences of the same action cannot overlap:

$$\bullet [\alpha : a] \mathbf{P}[\overline{\alpha : a}] \varphi \rightarrow \mathbf{P}[\overline{\alpha : a}] \varphi \quad (3)$$

$$\bullet \langle \alpha : a \rangle \top \wedge \mathbf{F}[\alpha : a] \varphi \rightarrow [\alpha : a] \mathbf{F}[\alpha : a] \varphi \quad (4)$$

Each occurrence runs linearly:

$$\bullet \langle \beta : a \rangle \varphi \rightarrow [\beta : a] \varphi \quad (5)$$

$$\bullet \langle \overline{\alpha : a} \rangle \varphi \rightarrow [\overline{\alpha : a}] \varphi \quad (6)$$

$$\bullet [\alpha : a] [\overline{\alpha : a}] \varphi \rightarrow \varphi \quad (7)$$

$$\bullet \varphi \wedge \langle \alpha : a \rangle \top \rightarrow [\alpha : a] [\overline{\alpha : a}] \varphi \quad (8)$$

At a same index, more than one action can be performed, by the same or another agent. In the above example, at  $w_6$ ,  $a_1$  performs  $\mu$  and  $\delta$ :  $w_6 \models \langle \mu : a_1 \rangle \top \wedge \langle \delta : a_1 \rangle \top$ .

An action is simply executed or not at an index, but it can unfold into different courses at different indexes of a same moment: in agreement with the STIT approach, actions are not deterministic. In particular, the duration of an action may be left unspecified. That is, not only different occurrences may have different lengths, but the possibly different courses of the same action occurrence may have different lengths on different histories. Action duration can for instance be influenced by the availability of resources. It is also influenced by the fact that actions may be suspended before completion, for reasons external or internal to the agent [Mül05]. Since actions may abort, starting an action does not imply obtaining some expected result:  $[\alpha : a]\varphi \rightarrow \Box[\alpha : a]\varphi$  is not valid. It grasps that  $\varphi$  may be the outcome of a course of  $\alpha$ , but  $\varphi$  could be not the outcome of another course of  $\alpha$ , even triggered at the same moment. This means that in our approach, actions are not simply characterized by preconditions and results; we rather focus on the decision of the agent to perform an event of some sort.

### 5.4.3.3 Continuation of an action, completed actions

Assuming that actions have a duration and can abort before completion allows to assume that the agent has control over the execution of an action. At each moment during the execution, the agent can decide to keep on performing it or not. On the other hand, in a STIT framework with several agents, agents share the set of indexes, and as a result, whenever an agent makes a choice, all other agents do too. This appears too demanding, as simply continuing what has been initiated before is not really a new choice. For both these reasons, we introduce particular actions representing the *continuation of an action*. Introducing in an explicit manner continuations of an action is actually a good way to formalize the notion of *control* on the action [Sea01]. Here, we follow Searle in holding that actions end when the bodily movement is finished, i.e., all actions are totally under control and thus continued up to their end:

$$\bullet \langle \alpha : a \rangle \top \rightarrow \langle \alpha_\lambda : a \rangle \top \mathcal{U} \langle \overline{\alpha} : \overline{a} \rangle \top \quad (9)$$

Of course, if a continuation is available, it means that the corresponding action has started before:

$$\bullet [\alpha_\lambda : a]\varphi \rightarrow \mathbf{P}[\alpha : a]\varphi \quad (10)$$

Continuations of a given occurrence of an action do not have a unique starting point, as they are repeated till the end of the action, but they all run till the same ending point.

Since actions can abort, when an action ends, it is not necessarily completed. We thus introduce propositions  $comp(\alpha) \in \mathcal{A}tm$ , one for each action  $\alpha$ , that reads “action  $\alpha$  is completed”. An action  $\alpha$  is completed when it has just ended and no continuation is possible; an action aborts if it ends but some continuation is still possible. In our example, action  $\mu$  is aborted in  $w_9$  and  $w_{11}$  because there is an available continuation  $\mu_\lambda$  at  $w_8$ .  $\mu$  is completed in  $w_{14}$ ,  $w_{16}$  and  $w_{18}$ . This notion of completion may be used to express that completed actions do have specific effects; categories of actions could then be introduced on the basis of effects of completed actions.

#### 5.4.3.4 Not doing anything

As observed before, STIT’s requiring that if an agent makes a choice at a moment, all other agents do too, is too demanding, and needs to be fixed not only for action continuations. In fact, agents may remain simply passive when others really choose to act. To express this, we introduce a set of propositions  $\lambda(a) \in \mathcal{A}tm$ , one for each agent  $a$ , that reads “the agent  $a$  remains passive”. An agent  $a$  remains passive when it does not perform an action nor a continuation. In the example, agent  $a_2$  remains passive everywhere but  $w_8$  and  $w_9$ :  $w_8 \models \neg\lambda(a_2)$ .

### 5.4.4 Choices and group agency: a new characterization of independence of agents

We can analyze the combination of choice and action. In multiagent systems, and particularly in STIT, an agent’s choice is understood as choosing to bring about a state of affairs. In the present work, we handle choice as choosing to perform a set of actions. In order to deal with agency we still use an operator equivalent to the Chellas STIT. We use the original notation of the Chellas’ STIT, that is  $[A\ cstit : \varphi]$ , in order not to get confused with PDL-style operators. We define it as follows:

$$[a\ cstit : \varphi] \triangleq \bigvee_{A \subseteq Act \cup Act_\lambda} \left( \left( \bigwedge_{\beta \in A} \langle \beta : a \rangle \top \right) \wedge \square \left( \left( \bigwedge_{\beta \in A} \langle \beta : a \rangle \top \right) \rightarrow \varphi \right) \right).$$

In words, an agent  $a$  sees to  $\varphi$  if there is a set  $A$  of actions of  $a$  such that every action of  $A$  is launched at the current index, and at every index of the moment, if those actions are triggered then  $\varphi$  holds.

Now, if we want to simulate  $[_\_ cstit : \_]$ , we need to ensure that actions of agents are independent. For convenience, we constrain it via a *Choice* function semantically defined as follows.

*Choice* :  $W \times \mathcal{A}gt \mapsto 2^W$  maps each index and agent  $a$  to all indexes of a moment where exactly the same actions or continuations are executed by  $a$ :  $w' \in \text{Choice}(w, a)$  iff  $w' \in R_{\square}(w)$  and  $(\forall \beta \in \mathcal{A}ct \cup \mathcal{A}ct_{\lambda}, R_{ACT}(\beta)(w)$  is defined iff  $R_{ACT}(\beta)(w')$  is defined). Hence, we are able to capture independence of actions by assuming a property similar to the general permutation property of Section 3.5: for all  $w, v \in W$  and for all  $l, m, n \in \mathcal{A}gt$ , if there is a  $w' \in \text{Choice}(w, l)$  and  $v \in \text{Choice}(w', m)$  then there is  $u \in W$  such that:  $u \in \text{Choice}(w, n)$  and  $v \in \text{Choice}(u, i)$  for every  $i \in \mathcal{A}gt \setminus \{n\}$ .

In particular, it forces  $[_\text{cstit} : \_]$  to be an S5 modal operator **(11)**. Moreover, if something is settled, every agent sees to it:  $\square\varphi \rightarrow [a \text{cstit} : \varphi]$  **(12)**.

We have the property that if  $a$  performs  $\alpha$ ,  $a$  sees to it that it performs  $\alpha$ , which can be stated by  $\langle \alpha : a \rangle \top \rightarrow [a \text{cstit} : \langle \alpha : a \rangle \top]$ . The other way round follows from S5( $[a \text{cstit} : \_]$ ). Similarly, it is also true that an agent remains passive if and only if it makes the choice of remaining passive:  $\lambda(a) \leftrightarrow [a \text{cstit} : \lambda(a)]$ . In the example, at the moment of  $w_1$ , agent  $a_1$  has three choices, corresponding to performing action  $\delta$ , performing action  $\mu$  or performing both. For instance, we have:  $w_2 \models [a_1 \text{cstit} : \langle \mu : a_1 \rangle \top] \wedge \diamond[a_1 \text{cstit} : \neg \langle \mu : a_1 \rangle \top]$  and  $w_6 \models [a_1 \text{cstit} : \langle \mu : a_1 \rangle \top \wedge \langle \delta : a_1 \rangle \top]$ .

It remains to ensure that agent's choices are independent. For that, we could impose Xu's axioms for independence of agents, viz. the collection of axiom schemas  $AIA_k$ . But fortunately, we do not need here this machinery that we have tried to simplify in Chapter 3. Actually, actions in the language allow us to state it in a new and intuitive manner via the following collection of formulae:

$$\begin{aligned} & \bullet \diamond \bigwedge_{\beta_0 \in A_0} \langle \beta_0 : a_0 \rangle \top \wedge \dots \wedge \diamond \bigwedge_{\beta_k \in A_k} \langle \beta_k : a_k \rangle \top \\ & \rightarrow \diamond (\bigwedge_{\beta_0 \in A_0} \langle \beta_0 : a_0 \rangle \top \wedge \dots \wedge \bigwedge_{\beta_k \in A_k} \langle \beta_k : a_k \rangle \top) \end{aligned}$$

where  $A_i \in \mathcal{A}ct \cup \mathcal{A}ct_{\lambda}$  are sets of actions or continuations. Then, if an arbitrary number of agents has the possibility to trigger sets of actions at a given moment, then they can launch those actions simultaneously.<sup>5</sup>

For instance,  $AIA_1$  (formulated  $\diamond[a_0 \text{cstit} : \varphi_0] \wedge \diamond[a_1 \text{cstit} : \varphi_1] \rightarrow \diamond([a_0 \text{cstit} : \varphi_0] \wedge [a_1 \text{cstit} : \varphi_1])$ ) can be derived from the definition of Chellas' stit and the two-agent version of our new schema for independence  $\diamond \langle \beta_0 : b_0 \rangle \top \wedge \diamond \langle \beta_1 : b_1 \rangle \top \rightarrow \diamond(\langle \beta_0 : b_0 \rangle \top \wedge \langle \beta_1 : b_1 \rangle \top)$  via simple S5 modal principles. We also have that no agent  $a$  can ensure that another agent  $b \neq a$  performs an action, except if it was settled:  $\models [a \text{cstit} : \langle \beta : b \rangle \top] \rightarrow \square \langle \beta : b \rangle \top$ , if  $a \neq b$ .

<sup>5</sup>Note that, given an arbitrary set of actions  $\{\beta_0, \dots, \beta_k\}$ , the more natural schema  $\diamond \langle \beta_0 : a_0 \rangle \top \wedge \dots \wedge \diamond \langle \beta_k : a_k \rangle \top \rightarrow \diamond(\langle \beta_0 : a_0 \rangle \top \wedge \dots \wedge \langle \beta_k : a_k \rangle \top)$  does not seem sufficient.

Concerning cooperation, at  $w_8$  none of  $a_1$  and  $a_2$  is able to see to it that both the main course and the dessert are cooked ( $\neg\Diamond[a_1 \text{ cstit} : (m \wedge d)] \wedge \neg\Diamond[a_2 \text{ cstit} : (m \wedge d)]$ ) but they can cooperate for that, and actually do at  $w_8$  ( $w_8 \models [\{a_1, a_2\} \text{ cstit} : (m \wedge p)]$ ); it is achieved by means of  $a_1$  continuing to perform  $\mu$  and  $a_2$  executing  $\Delta$ .

## 5.5 Discussion

Neither  $\mathbb{N}$ STIT nor the logic of actions with duration are tailored, or even assumed, as something else than preliminary investigations of the link between choice, time and action. We don't want to give anybody the impression that something has been solved. Nonetheless, even humble, we hope – and think – that this chapter supplies useful clarifications on the intrinsic properties of operators of agency in models of branching time. Dag Elgesem in [Elg93, Sect. 2.IV] defends the thesis that “temporal aspect should not enter into the characterization of agency”. (See also [Elg97].) To argue for it, he tries to interpret his theory of agency into a temporal framework and observes that it does not affect his theory. In this very context, he argues that there seems to be no good reasons for including temporal aspects in his characterization of agency. But the agenda is different: he proposes a wide theory of agency allowing to understand various aspects linked to the more general notion of agency, e.g., ability, unpreventability, independence, opportunity, etc.

In this chapter, the study has been more focused: the scope has essentially been given to some sense of causality and underlying action. We also did a work of clarification on different operators of agency in the literature, that, as we have seen, often just slightly differ by their temporal aspects.

To warm up the more precise proposal of Chapter 7 we have introduced actions with duration, action continuations, and explicit choices of remaining passive in a STIT-like framework. By doing so, we have also cured an annoying feature of STIT which is that when an agent makes a choice, other agents do too. Moreover, choices in STIT are arbitrary partitions of moments; We made the notion of choice clearer by constructing choices over sets of actions. From this preliminary investigation, we think that an interesting problem to pursue is to establish a systematic link between a theory of STIT and Dynamic Logic. This could be a starting point for a solution to a similar problem raised by Johan van Benthem in [vB06]. A Dynamic Logic with a finite number of actions plus an operator of his-

---

toric possibility seem to be an adequate logical composition to simulate deliberative STIT theories. However, after this quick prelude, a complete axiomatization and a result of decidability remain exciting challenges.

We do not go further here. With respect to the scope of this dissertation, we prefer to see it as a preliminary conceptualization in modal logic on which an ontology is constructed in Chapter 7. This is a methodology defended in [TV07].



# 6

---

## Alternating-time Temporal Logic vs. STIT

### 6.1 Introduction

In the philosophical literature operators of STIT have been used in the analysis of agency and in the analysis of deontic concepts [BPX01, Hor01]. We believe that the philosophical intuitions underlying STIT theory are equally relevant for logical models developed to analyze and design multiagent systems. To support this claim, in this chapter we show that there is a close relationship with more recent temporal logics for specification and verification of multiagent systems. In particular, we will study here the relation between Alternating-time Temporal Logic (ATL) proposed by Alur, Henzinger and Kupferman [AHK97, AHK99, AHK02] and the logic of the fused  $\diamond_s[_{scstit} : \_]$  operator, slightly adapted from Horty [Hor01]. ATL was designed as an extension of CTL. CTL is a branching time temporal logic with modal operators quantifying (universally (**A**) and existentially (**E**)) over sets of paths. In ATL, quantification is with respect to strategies, and quantification over paths is implicit as quantification over all paths that are in the *outcome* of a certain strategy. In particular,  $\langle\langle A \rangle\rangle$ , where  $A \subseteq \mathcal{Agt}$  is a group of agents, stands for existential quantification over strategies in the repertoire of  $A$ . In ATL,  $\langle\langle A \rangle\rangle$  is always followed by one of the temporal operators **X** (next), **G** (henceforth) or **U** (until). Evaluation of these temporal operators is with respect to paths that are in the outcome of a strategy. For example,  $\langle\langle A \rangle\rangle \mathbf{X}\varphi$  reads: “group  $A$  has a strategy to ensure that next  $\varphi$ ”. This setting allows for refinements of the CTL quantification over paths, CTL **E** corresponding to the ATL  $\langle\langle \mathcal{Agt} \rangle\rangle$  and **A** corresponding to  $\langle\langle \emptyset \rangle\rangle$ . It was shown by Goranko [Gor01] that ATL is also an extension of Pauly’s Coalition Logic CL [Pau02] that we have reviewed in Section 4.2. The latter is the logic of expressions of the form  $\langle\langle A \rangle\rangle \varphi$ , reading “group  $A$

can ensure that  $\varphi''$ . (As in Chapter 4 the main operator of Coalition Logic is noted  $\langle\langle A \rangle\rangle\varphi$ .) Such expressions correspond to ATL formulas  $\langle\langle A \rangle\rangle\mathbf{X}\varphi$ .

In [BHT06c] we proposed the following translation from CL to STIT.

$$\begin{aligned} tr_{\text{CL}}(p) &= \Box p, \text{ for } p \in \mathcal{Atm} \\ tr_{\text{CL}}(\neg\varphi) &= \neg tr(\varphi) \\ tr_{\text{CL}}(\varphi \vee \psi) &= tr(\varphi) \vee tr(\psi) \\ tr_{\text{CL}}(\langle\langle A \rangle\rangle\varphi) &= \Diamond[A \text{ cstit} : \mathbf{X}tr(\varphi)] \end{aligned}$$

In Chapter 4 we have also given a very similar link between Coalition Logic and STIT theories. This chapter proposes a more general translation from ATL to a discrete version of strategic STIT logic.

In [Wöl04], a close examination of the differences and similarities of the models of STIT theory and ATL is undertaken. It is shown that, under the addition of some specific conditions (e.g., discreteness), the models of the two systems can be seen to obey similar properties, like tree-likeness, uniformity and ‘restrictedness’ (see section 6.3.2). However, these properties are not necessarily expressible in the logics of STIT or ATL. So, although, from a philosophical point of view, it is interesting to look at properties of models as such, here we are essentially interested only in those properties that are expressible in the logics. Where [Wöl04] only compares the models for ATL and STIT, we also compare the logics of both systems.

In Section 6.2 we offer a presentation of Alternating-time Temporal Logic. We overview the complexity of reasoning on ATL and prove a semantic equivalence result on ATL frames in Section 6.3. Section 6.4 consists of a slightly adapted strategic stit operator. Section 6.5 presents the main result of this chapter: we describe a translation from ATL to STIT, and prove that it is correct.<sup>1</sup> We conclude with an informal discussion about the meaning of the results, justifying carefully our assumptions in Section 6.6.

## 6.2 Alternating-time Temporal Logic

The first paper on ATL is [AHK97]. This preliminary work is restricted to turn-based games, i.e., games where each transition is governed by a single agent. [AHK99] comes with general structures called *alternating transition systems* (ATs), where choices are expressed as sets of possible outcomes. In [AHK02] the authors change the models into *concurrent game*

<sup>1</sup>A correct embedding is a sound and complete translation to a fragment of a stronger logic.

structures (CGSs),<sup>2</sup> where choices are identified with explicit labels. ATs and CGSs have been proven equivalent by Goranko and Jamroga [GJ04]. Hence, defining the semantics of ATL in terms of either ATs or CGSs is a matter of convenience.

In what follows,  $Atm$  represents a set of atomic propositions, and  $Agt$  is the finite set of all agents.

**Syntax.** Given that  $p$  ranges over  $Atm$ , and that  $A$  ranges over  $2^{Agt}$ , the language of ATL is defined by the BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle A \rangle\rangle X\varphi \mid \langle\langle A \rangle\rangle G\varphi \mid \langle\langle A \rangle\rangle \varphi U \varphi$$

The intended reading of  $\langle\langle A \rangle\rangle \eta$ , with  $\eta$  a linear temporal formula (branch formula), is that “group  $A$  can ensure  $\eta$  whatever agents in  $Agt \setminus A$  do”.

**Models.** We present models for ATL as in [AHK99], that is, in terms of alternating transition systems which are tuples  $\mathcal{M} = \langle W, \delta, v \rangle$ , where:

- $W$  is a nonempty set of states (alias worlds, alias moments).
- $\delta : W \times Agt \rightarrow 2^{2^W}$  is a transition function mapping each moment and agent to a nonempty family of sets of possible successor moments.
- $v : Atm \rightarrow 2^W$  is a valuation function.

Each  $Q \in \delta(w, a)$  may be seen as the choice by an agent of a particular action in its repertoire.

We use *lock-step synchronous* ATs, which means that in every state, all agents proceed simultaneously (as opposed to the particular case of *turn-based synchronous* ATs). The  $\delta$  function is *non blocking* (agent’s actions are always compatible) and the simultaneous choice of every agent in  $Agt$  determines a *unique next state*: assuming  $Agt = \{a_1, \dots, a_n\}$ , for every state  $w \in W$  and every set  $\{Q_1, \dots, Q_n\}$  of choices  $Q_i \in \delta(w, a_i)$ , the intersection  $Q_1 \cap \dots \cap Q_n$  is a singleton.

A *strategy for an agent  $a$*  is a mapping  $f_a : W^+ \rightarrow 2^W$ , such that it associates to each sequence of states  $w_0 \dots w_k$  an element of  $\delta(w_k, a)$ .<sup>3</sup> A *collective strategy*, for a set of agents  $A \subseteq Agt$  is a tuple  $F_A = \langle f_{a_1}, f_{a_1}, \dots, f_{a_i} \rangle$

<sup>2</sup>An alternative name from the literature is ‘multi-player game model’, abbreviated ‘MGM’.

<sup>3</sup>It actually suffices to use mappings  $f_a : W \rightarrow 2^W$  [GJ04]. However, the current definition is the customary one.

of strategies, one for each agent in  $A$ . The outcome of  $F_A$  from  $w$  is defined as:

$$out(w, F_A) = \{\lambda \mid \lambda = w_0 w_1 w_2 \dots, w_0 = w, \forall i \geq 0 (w_{i+1} \in \bigcap_{a \in A} f_a(w_0 \dots w_i))\}$$

**Definition 6.1 (strategy profile / choice profile).** A strategy profile is a collective strategy  $F_{Agt}$  for all agents of  $Agt$ . Analogously, a tuple  $\langle Q_1, \dots, Q_n \rangle$  (one  $Q_i$  for each  $i \in Agt$ ) is called a choice profile.

**Truth values and axiomatization.**  $\lambda[i]$  is the  $i$ -th position in the path  $\lambda$ . A formula is evaluated with respect to an ATS  $\mathcal{M} = \langle W, \delta, v \rangle$  and a moment  $w \in W$ .

$$\begin{aligned} \mathcal{M}, w \models \langle\langle A \rangle\rangle \mathbf{X}\varphi &\iff \exists F_A, \forall \lambda \in out(w, F_A), \mathcal{M}, \lambda[1] \models \varphi \\ \mathcal{M}, w \models \langle\langle A \rangle\rangle \mathbf{G}\varphi &\iff \exists F_A, \forall \lambda \in out(w, F_A), \mathcal{M}, \lambda[i] \models \varphi, \forall i \geq 0 \\ \mathcal{M}, w \models \langle\langle A \rangle\rangle \varphi \mathbf{U} \psi &\iff \exists F_A, \forall \lambda \in out(w, F_A), \\ &\exists i \geq 0 (\mathcal{M}, \lambda[i] \models \psi, \forall j \in [0, i], \mathcal{M}, \lambda[j] \models \varphi) \end{aligned}$$

Validity is defined as usual. The following complete axiomatization of ATL (as an extension of any axiomatization for propositional logic) is given in [GvD06].  $\mathcal{M}, w \models \langle\langle \emptyset \rangle\rangle \eta$  means that  $\eta$  holds irrespective of the choices made by  $Agt$ .

- ( $\perp$ )  $\neg \langle\langle A \rangle\rangle \mathbf{X}\perp$
- ( $\top$ )  $\langle\langle A \rangle\rangle \mathbf{X}\top$
- ( $N$ )  $\neg \langle\langle \emptyset \rangle\rangle \mathbf{X}\neg\varphi \rightarrow \langle\langle Agt \rangle\rangle \mathbf{X}\varphi$
- ( $S$ )  $\langle\langle A_1 \rangle\rangle \mathbf{X}\varphi \wedge \langle\langle A_2 \rangle\rangle \mathbf{X}\psi \rightarrow \langle\langle A_1 \cup A_2 \rangle\rangle \mathbf{X}(\varphi \wedge \psi)$  if  $A_1 \cap A_2 = \emptyset$
- ( $FP_G$ )  $\langle\langle A \rangle\rangle \mathbf{G}\varphi \equiv \varphi \wedge \langle\langle A \rangle\rangle \mathbf{X}\langle\langle A \rangle\rangle \mathbf{G}\varphi$
- ( $GFP_G$ )  $\langle\langle \emptyset \rangle\rangle \mathbf{G}(\theta \rightarrow (\varphi \wedge \langle\langle A \rangle\rangle \mathbf{X}\theta)) \rightarrow \langle\langle \emptyset \rangle\rangle \mathbf{G}(\theta \rightarrow \langle\langle A \rangle\rangle \mathbf{G}\varphi)$
- ( $FP_U$ )  $\langle\langle A \rangle\rangle \psi \mathbf{U} \varphi \equiv \varphi \vee (\psi \wedge \langle\langle A \rangle\rangle \mathbf{X}\langle\langle A \rangle\rangle \psi \mathbf{U} \varphi)$
- ( $LFP_U$ )  $\langle\langle \emptyset \rangle\rangle \mathbf{G}((\varphi \vee (\psi \wedge \langle\langle A \rangle\rangle \mathbf{X}\theta)) \rightarrow \theta) \rightarrow \langle\langle \emptyset \rangle\rangle \mathbf{G}(\langle\langle A \rangle\rangle \psi \mathbf{U} \varphi \rightarrow \theta)$
- ( $\langle\langle A \rangle\rangle \mathbf{X}$ -Mon) from  $\varphi \rightarrow \psi$  infer  $\langle\langle A \rangle\rangle \mathbf{X}\varphi \rightarrow \langle\langle A \rangle\rangle \mathbf{X}\psi$
- ( $\langle\langle \emptyset \rangle\rangle \mathbf{G}$ -Nec) from  $\varphi$  infer  $\langle\langle \emptyset \rangle\rangle \mathbf{G}\varphi$

Note that the  $(N)$  axiom follows from the determinism of ‘global’ actions (actions constituted by simultaneous choices for every agent in the system): when every agent opts for a choice, the next state is fully determined, thus, if something is not settled, the coalition of all agents ( $\mathcal{Agt}$ ) can always work together to make its negation true. The axiom  $(S)$  says that two coalitions can combine their efforts to ensure a conjunction of properties if they are disjoint. Note that from  $(S)$  it follows that  $\langle\langle A_1 \rangle\rangle\varphi \wedge \langle\langle A_2 \rangle\rangle\neg\varphi$  is not satisfiable for disjoint  $A_1$  and  $A_2$ . So, two disjoint coalitions cannot ensure inconsistent propositions. Axiom  $(FP_G)$  characterizes the global modality as a fixpoint of the next modality, and axiom  $(GFP_G)$  says that this is the greatest fixpoint. Axiom  $(FP_U)$  characterizes the until operator as a (special kind of) fixpoint of the next operator, and axiom  $LFP_U$  expresses that the semantics dictates that we take the least fixpoint.

## 6.3 Formal properties of ATL

### 6.3.1 Complexity

Having provided a complete axiomatics for ATL with Goranko in [GvD06], van Drimmelen was able to prove EXPTIME-completeness of the problem of satisfiability of an ATL formula over a fixed number of agents [vD03]. Hardness follows from the EXPTIME-completeness of CTL, a fragment of ATL. The upper bound is obtained by proving that a satisfiable formula *over a fixed number of agents* is also satisfiable in a tree model of bounded branching degree, which suggests an exponential time satisfiability. However, if the set of agent is not fixed in advance, the construction does not permit to obtain a tight result. As stated in [LWWW06] we can distinguish three cases:

- Given a finite set  $\mathcal{Agt}$  of agents and a formula  $\varphi$  over  $\mathcal{Agt}$ , is  $\varphi$  satisfiable in an ATS over  $\mathcal{Agt}$ ?
- Given a formula  $\varphi$ , is there a finite set  $\mathcal{Agt}$  of agents (containing the agents in  $\varphi$ ) such that  $\varphi$  is satisfiable in an ATS over  $\mathcal{Agt}$ ?
- Given a formula  $\varphi$ , is  $\varphi$  satisfiable in an ATS over exactly the agents which occur in  $\varphi$ ?

Walther et al. hence conclude that each variation of the problem of satisfiability is NEXPTIME-complete.

Concerning model checking, Alur et al. proved that ATL model checking *over ATS* is PTIME-complete. However, recent and very interesting works have emitted critics about the result [vdHLW06]. Their starting point is [AH99] and its *Reactive Modules Language* (RML), which is the language used by the famous MOCHA [AHM<sup>+</sup>98] model checker particularly suitable for modeling multiagent protocols. RML makes model checking ‘practical’ for that one can describe a system of agents by means of small modules, while describing the same system with an alternating transition system makes the size of the specification explode [JD05]. Van der Hoek et al. thus slightly adapt RML to “Simple Reactive Modules Language” (SRML), and prove that model checking an ATL formula over SRML is NEXPTIME-complete. Interestingly, the complexity for the fragment of Coalition Logic comes to be PSPACE-complete. Hence, the complexity of model checking agents properties in a practical manner reveals itself to be as high as the complexity of theorem proving.

### 6.3.2 Semantic equivalence results for ATL

As a first step towards the embedding, we discuss semantic equivalence results for interpreting ATL on ATSs. First we introduce some useful notations:

**Definition 6.2 (successor states / tree-order).** *Given an ATS  $\mathcal{M} = \langle W, \delta, v \rangle$  and an agent  $a \in \text{Agt}$ :*

- $\text{Succ}_a(w) \triangleq \{w' \mid w' \in Q_a, Q_a \in \delta(w, a)\}$
- $\text{Succ}(w) \triangleq \bigcap_{a \in \text{Agt}} \text{Succ}_a(w)$
- $w \prec_\delta w' \triangleq w' \in \text{Succ}(w)$
- $<_\delta$  is the transitive closure of  $\prec_\delta$

Intuitively,  $\text{Succ}_a(w)$  gives the possible successor states from the point of view of agent  $a$ , and  $\text{Succ}(w)$  gives possible successor states for the complete system of agents.

The first steps on the issue of semantical equivalence have already been made by Wölfl [Wöl04], who, among other things, shows how any ATS can be unraveled into  $\langle W, \delta, v \rangle$  in such a way that  $\langle W, <_\delta \rangle$  is a tree. From any ATS we can thus construct a tree-like ATS that is bisimilar. Therefore we may restrict our study to tree-like ATSs.

**Definition 6.3 (tree-like ATSs).** *An ATS  $\mathcal{M} = \langle W, \delta, v \rangle$  where  $\langle W, <_\delta \rangle$  is a tree, is called a tree-like ATS.*

Now, for ATSS it is not necessarily the case that  $Succ_a(w) = Succ(w)$ . The only condition on ATSS is that each intersection of choices by all members of  $Agt$  results in a unique state. This does not guarantee that choices for individual agents do not overlap, and it does not guarantee that there are worlds that seem reachable from the point of view of some agents but are actually not reachable in any simultaneous step by all agents in the system. To be more precise, if  $\delta(w, a) = \{Q_1 \dots Q_n\}$ , then both  $Q_i \cap Q_j \neq \emptyset$  and  $\bigcup_{1 < i < n} Q_i \subsetneq Succ(w)$  for some  $i$  and  $j$  in  $[1, n]$  are allowed. These properties would not hold if, like in STIT, choices for individual agents would partition the set of possible reachable worlds. In this section we will work towards tree-like choice partitioned ATSS and show that they are bisimilar for ATL. For these models we thus have  $Succ_a(w) = Succ(w)$  for every  $a \in Agt$ .

Wölfl explicitly constrains ATSS with the condition that for each agent  $a$  and each state  $w$ ,  $\delta(w, a)$  is a partition of the set of successor states of  $w$ . Here we show that this explicit restriction is not necessary.

**Definition 6.4 (choice partitioned ATSS).** *An ATSS  $\mathcal{M} = \langle W, \delta, v \rangle$  is called a choice partitioned ATSS if for all agent  $a \in Agt$  and for all state  $w \in W$  the choices  $\delta(w, a)$  partition the set  $Succ(w)$ .*

We can now state a result of alternating bisimulation in ATSS. We prove it by roughly following the proof of [BdRV01, Prop. 2.15]. But the reader can refer to [AHKV98] for a particular account of alternating bisimulation.

**Lemma 6.1.** *For any ATSS  $\mathcal{M} = \langle W, \delta, v \rangle$  we can construct an alternating bisimilar tree-like and choice partitioned ATSS  $\mathcal{M}' = \langle W', \delta', v' \rangle$ .*

PROOF.

Elements of  $W'$  are sequences

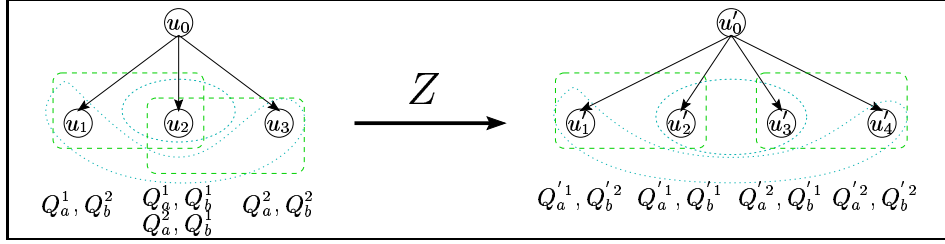
$$(u_0, \langle u_1, \langle Q_1^0 \dots Q_n^0 \rangle \rangle, \dots, \langle u_k, \langle Q_1^{k-1} \dots Q_n^{k-1} \rangle \rangle)$$

satisfying  $k \geq 0$ ,  $Agt = \{a_1, \dots, a_n\}$ ,  $u_i \in W$ ,  $u_{i+1} \in \bigcap_{a \in Agt} Q_a^i$ ,  $Q_a^i \in \delta(a, u_i)$ .  $u_0$  is intended as the root of  $\mathcal{M}$ , and every  $u_i$  is a state reached from  $u_{i-1}$  by agents of  $Agt$ , applying the choice profile  $\langle Q_1^i \dots Q_n^i \rangle$ . Then, for every agent  $a$  and for all  $w' = (u_0, \langle u_1, \langle Q_1^0 \dots Q_n^0 \rangle \rangle, \dots, \langle u_k, \langle Q_1^{k-1} \dots Q_n^{k-1} \rangle \rangle)$  of  $W'$ , we define  $\delta'(a, w') = \{Q'_1 \dots Q'_d\}$  with:

$$Q'_j = \{ (u_0, \langle u_1, \langle Q_1^0 \dots Q_n^0 \rangle \rangle, \dots, \langle u_k, \langle Q_1^{k-1} \dots Q_n^{k-1} \rangle \rangle, \langle u_{k+1}, \langle Q_1^k \dots Q_n^k \rangle \rangle) \mid \delta(a, u_k) = \{Q_1 \dots Q_n\}, u_{k+1} \in Q_a \}$$

For all  $w' = (u_0, \langle u_1, \langle Q_1^0 \dots Q_n^0 \rangle \rangle, \dots, \langle u_k, \langle Q_1^{k-1} \dots Q_n^{k-1} \rangle \rangle)$ , the valuation function  $v'$  is defined by  $v'(w') = v(u_k)$ .

Let  $w' = (u_0, \langle u_1, \langle Q_1^0 \dots Q_n^0 \rangle \rangle, \dots, \langle u_k, \langle Q_1^{k-1} \dots Q_n^{k-1} \rangle \rangle)$  and  $Z : W \rightarrow 2^{W'}$  defined such that  $w' \in Z(u_k)$ . Clearly  $Z$  is a bisimulation between  $\mathcal{M}$  and  $\mathcal{M}'$ . ■



**Figure 6.1:** Construction of a semantically equivalent choice partitioned ATS. Dotted boxes correspond to  $\delta(u_0, a)$  (resp.  $\delta(u'_0, a)$ ) and closed curves correspond to  $\delta(u_0, b)$  (resp.  $\delta(u'_0, b)$ ).

As an illustration, consider a *pre-ATS*<sup>4</sup>  $\mathcal{M}$  over two agents  $a$  and  $b$ . (Left part of Figure 6.1.) From  $u_0$ , agent  $a$  can choose either  $Q_a^1 = \{u_1, u_2\}$  or  $Q_a^2 = \{u_2, u_3\}$ . Agent  $b$  can choose either  $Q_b^1 = \{u_2\}$  or  $Q_b^2 = \{u_1, u_3\}$ . Clearly  $\mathcal{M}$  is not choice partitioned since  $\{Q_a^1, Q_a^2\}$  is not a partition of  $\text{Succ}(u_0)$  ( $Q_a^1 \cap Q_a^2 \neq \emptyset$ ).

We construct the equivalent choice partitioned ATS  $\mathcal{M}' = \langle W', \delta', v' \rangle$  by duplicating  $u_2$  which can be reached by applying two different choice profiles. (Right part of Figure 6.1.) Members of  $W'$  are thus  $u'_0 = (u_0)$ ,  $u'_1 = (u_0, \langle u_1, \langle Q_a^1, Q_b^2 \rangle \rangle)$ ,  $u'_2 = (u_0, \langle u_2, \langle Q_a^1, Q_b^1 \rangle \rangle)$ ,  $u'_3 = (u_0, \langle u_2, \langle Q_a^2, Q_b^1 \rangle \rangle)$  and  $u'_4 = (u_0, \langle u_3, \langle Q_a^2, Q_b^2 \rangle \rangle)$ . The transition function at  $u'_0$  is represented by  $\delta'(u'_0, a) = \{\{u'_1, u'_2\}, \{u'_3, u'_4\}\}$  and  $\delta'(u'_0, b) = \{\{u'_1, u'_4\}, \{u'_2, u'_3\}\}$ .

Lemma 6.1 permits us, without loss of generality, to consider only tree-like choice partitioned ATSS. Wölfl calls these ATSS ‘restricted’. However, as the semantic equivalence shows, this restriction is not a restriction from the viewpoint of modal logic. We come back to the equivalence property in Section 6.6.

## 6.4 A convenient strategic stit operator of ability

We slightly adapt the original logic for strategies defined by Horty. We make some assumptions that we discuss and motivate later in Section 6.6.

<sup>4</sup>Valuation and transition functions from  $u_1, u_2$  and  $u_3$  are irrelevant.

**Syntax.** Given that  $p$  ranges over  $Atm$ , and that  $A$  ranges over  $2^{Agt}$ , a language of strategic stit is defined by:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \phi \mid \Box\varphi \mid \mathbf{X}\varphi \mid \mathbf{G}\varphi \mid \varphi\mathcal{U}\phi \mid \Diamond_s[A \text{ scstit} : \varphi]$$

First we have to explain why we call the logic defined relative to the above syntax a logic of ‘strategic stit *ability*’ in stead of a logic of ‘strategic stit’. The intuitive reading of  $\Diamond_s[A \text{ scstit} : \varphi]$  is “it is strategically possible that agents  $A$  see to it that  $\varphi$ ”. The operator  $\Diamond_s[A \text{ scstit} : \varphi]$ , suggested by Horty [Hor01, p.152], is thus a special (fused) operator that is ‘built’ from an operator for strategic possibility ( $\Diamond_s\varphi$ ) and a strategic version of Chellas’s stit operator ( $[A \text{ cstit} : \varphi]$ ). However, in Horty’s work these separate operators are not given a formal semantics individually; the operators are syntactically forced to occur only in combination (in a recent proposal [BHT06a] we propose a solution to this problem by evaluating with respect to strategy / state pairs). Yet, to understand the semantics of the fused operator, below we discuss the intended semantics of the individual operators.

In this work of mapping ATL, we *do not* need refinement of evaluating the operator of strategic stit ability as in Section 2.3, viz. against  $w/h/M$ , that is the usual index plus a *field*. Indeed, for any strategy at a moment  $w$  we will always consider the field to be the complete set  $Tree_w$ , that is, the sub-tree having  $w$  as root. For evaluation of formulas in the strategic setting we will use the same models and indexes as for the non-strategic setting.<sup>5</sup>

Also, we can trivially extend strategies to groups as follows.

**Definition 6.5 (collective strategy).** A collective strategy for  $A \subseteq Agt$  is a tuple  $\sigma_A = \langle \sigma_a \rangle_{a \in A}$ , and  $Adh(\sigma_A) = \bigcap_{a \in A} Adh(\sigma_a)$ .

It turns out that we can here adapt in consequence the truth value of the  $\Diamond_s[a \text{ scstit} : \varphi]$  operator. We can define a *fused* operator for long term strategic ability of groups of agents as follows:

$$\mathcal{M}, w/h \models \Diamond_s[A \text{ scstit} : \varphi] \iff \exists \sigma \in Strategy_A^w \text{ s.th. } \forall h' \in Adh(\sigma), \mathcal{M}, w/h' \models \varphi$$

<sup>5</sup>It is easy to see that actually histories are not needed to evaluate the strategic ability operator. Horty calls this moment-determinateness of the fused operator. We nevertheless keep the histories for uniformity purposes.

where  $Strategy_A^w = \{\sigma \mid Dom(\sigma) = Tree_w\}$ , is the set of strategies open to  $A$  at moment  $w$ .<sup>6</sup>

In combination with the standard CSTIT theorem  $\Box\varphi \rightarrow \Diamond[A\text{ cstit} : \varphi]$ , for nonempty coalitions  $A \subseteq \mathcal{A}gt$  we arrive at the following property for strategic ability:

$$\models \Box\varphi \rightarrow \Diamond_s[A\text{ scstit} : \varphi]$$

This property ensures that the translation we propose in Section 6.5 embeds the translation we did for CL (cf. the definition of  $tr_{CL}$  in Section 6.1).

For empty coalitions this implication strengthens to an equivalence.<sup>7</sup> We need the next proposition in our proof of Theorem 6.2 below.

**Proposition 6.1.**  $\models \Diamond_s[\emptyset\text{ scstit} : \varphi] \equiv \Box\varphi$

PROOF. Since the empty coalition of agents is just assigned the vacuous choice, at each moment  $w'$ , the empty coalition has no alternative but  $H_{w'}$ . Hence,  $Strategy_\emptyset^w = \{\sigma_\emptyset\}$  with  $\sigma_\emptyset(w') = H_{w'}$  for all  $w' \in Tree_w$ . Therefore, for all  $\sigma$  in  $Strategy_\emptyset^w$ , we have  $Adh(\sigma) = H_w$ .

Thus  $\mathcal{M}, w/h \models \Diamond_s[\emptyset\text{ scstit} : \varphi] \iff \forall h' \in H_w, \mathcal{M}, w/h' \models \varphi$ . Which corresponds to the semantics of the operator of historical necessity. ■

To enable a comparison with ATL we make the following assumptions concerning the nature of time. They are discussed and motivated in Section 6.6.

**Hypothesis 6.1 (countably infiniteness).** *Every history is isomorphic to the set of natural numbers.*

By assuming that histories are countably infinite sets of moments we will be able to reason about temporal properties as in LTL.

As time is discrete in our present setting, we can define the temporal operator **X** (*next*). We also introduce operators **G** (*always*) and **U** (*until*):

$$\begin{aligned} \mathcal{M}, w/h \models \mathbf{X}\varphi &\iff \exists w' \in h (w < w', \mathcal{M}, w'/h \models \varphi, \\ &\quad \nexists w'' \in h (w < w'' < w')). \\ \mathcal{M}, w/h \models \mathbf{G}\varphi &\iff \forall w' \in h (w \leq w', \mathcal{M}, w'/h \models \varphi) \\ \mathcal{M}, w/h \models \varphi\mathbf{U}\psi &\iff \exists w' \in h (w < w', \mathcal{M}, w''/h \models \psi, \\ &\quad \forall w'' (w \leq w'' < w', \mathcal{M}, w''/h \models \varphi) \end{aligned}$$

<sup>6</sup>In the original definition, a set of strategies is denoted  $Strategy_a^M$ , where  $M$  is a field having  $w$  as root. Since we have assumed that  $M$  is always  $Tree_w$ , our notation  $Strategy_A^w$  suffices.

<sup>7</sup>These results of the links between CSTIT and the strategic operator transfer to the general case with fields.

We also make the hypothesis that the intersection of choices of agents in  $\mathcal{Agt}$  must exactly be the set of histories passing through some immediate next moment:

**Hypothesis 6.2 (determinism).**

$$\forall w \in W, \exists w' \in W (w < w' \text{ and } \bigcap_{a \in \mathcal{Agt}} s_w(a) = H_{w'})$$

Note that because STIT frames are discrete trees, the moment  $w'$  is always a next moment.

## 6.5 From ATL to STIT logic

We define the translation  $tr$  from ATL formulae to STIT formulae as:

$$\begin{aligned} tr(p) &= \Box p, \text{ for } p \in \mathcal{Atm} \\ tr(\neg\varphi) &= \neg tr(\varphi) \\ tr(\varphi \vee \psi) &= tr(\varphi) \vee tr(\psi) \\ tr(\langle\langle A \rangle\rangle \mathbf{X}\varphi) &= \Diamond_s[A \text{ scstit} : \mathbf{X}tr(\varphi)] \\ tr(\langle\langle A \rangle\rangle \mathbf{G}\varphi) &= \Diamond_s[A \text{ scstit} : \mathbf{G}tr(\varphi)] \\ tr(\langle\langle A \rangle\rangle \varphi \mathcal{U} \psi) &= \Diamond_s[A \text{ scstit} : tr(\varphi) \mathcal{U} tr(\psi)] \end{aligned}$$

Translating an atom  $p$  into a modal formula  $\Box p$  may seem odd, but is motivated by the remark on page 14. All other clauses of the translation are straightforward, given the intended interpretation of the operators. The remainder of the section is devoted to the proof of the correctness of  $tr$ .

Given a tree-like choice partitioned ATS  $\mathcal{M}_{\text{ATL}} = \langle W_{\text{ATL}}, \delta, v_{\text{ATL}} \rangle$  we associate to it a STIT model  $\mathcal{M}_{\text{STIT}} = \langle W_{\text{STIT}}, \text{Choice}, <, v_{\text{STIT}} \rangle$ , as follows:

- $W_{\text{STIT}} = W_{\text{ATL}}$
- $w < u \iff \exists u_1, \dots, u_n (u_1 = w, u_n = u, \forall i < n (\exists a \in \mathcal{Agt}, Q_a \in \delta(u_i, a), u_{i+1} \in Q_a))$
- $\text{Choice}_a^w = \{\{h \mid Q_a \cap h \neq \emptyset\} \mid Q_a \in \delta(w, a)\}$  for all  $a$  and  $w$
- $\forall h \in H_w, v_{\text{STIT}}(w/h) = v_{\text{ATL}}(w)$

It is clear that the tree property is instrumental for  $\langle W_{\text{STIT}}, < \rangle$  being a tree. We inherit the branching time structure of STIT directly from the tree structure of the ATS. Furthermore, the condition concerning partitions

underlying choice partitioned ATSs prevents that two choices of the same agent have a non-empty intersection, and therefore every  $Choice_a^w$  is a partition of  $H_w$ . If intersections would possibly be non-empty, we could not have constructed the  $Choice$  function as we did: the same history could have been in two different sets of  $Choice_a^w$ .

**Proposition 6.2.**  $\mathcal{M}_{STIT}$  is a discrete STIT model, and  $\mathcal{M}_{STIT}$  is unique.

PROOF. Straightforward. ■

In the following,  $\mathcal{M}_{STIT}$  histories are *maximal* sequences of ATL states respecting  $<$ . Given a history  $h = \{w_0, w_1, \dots\}$  we can construct an infinite sequence of states  $\lambda = q_0q_1\dots$  such that:  $\forall q_i \in \lambda, \exists w_j \in h$  s.th.  $q_i = w_j$ ,  $q_i < q_{i+1}$  and  $\nexists w \in h, q_i < w < q_{i+1}$  (since we have identified  $W_{ATL}$  with  $W_{STIT}$ , we can thus order members of  $W_{ATL}$  with the relation  $<$ ). At such a condition we will say that  $h = \lambda$  (slightly abusing notation). Thus, we will indifferently use a STIT history and the corresponding ATL sequence of states.

**Lemma 6.2.** Let  $u \in W_{ATL}$  be a state in  $\mathcal{M}_{ATL}$ . For every collective ATL strategy  $F_A$  from  $\mathcal{M}_{ATL}$ , there is a collective STIT strategy  $\sigma_A \in Strategy_A^u$  such that  $out(u, F_A) = Adh(\sigma_A)$ .

PROOF. We assume w.l.o.g. that the ATS of  $\mathcal{M}_{ATL}$  is a tree-like choice partitioned structure. Let  $path : W_{STIT} \rightarrow W_{ATL}^+$  map each moment  $w$  into the (unique) *maximal* ordered sequence of states terminated by  $w$ . For all  $f_a$  of the tuple  $F_A$  we construct  $\sigma_a$  s.th.: for all  $u \in W_{STIT}$  and  $w' \in Tree_u$  we have

$$\sigma_a(w') = \{h \mid f_a(path(w')) \cap h \neq \emptyset\}$$

We let  $\sigma_a(w')$  undefined for  $w'$  outside  $Tree_u$ . Let  $\sigma_A = \langle \sigma_a \rangle_{a \in A}$ , we want to show that  $out(u, F_A) = Adh(\sigma_A)$ .

( $\subseteq$ ) Suppose  $\lambda \in out(u, F_A)$ . It means  $\lambda = q_0q_1\dots$  with  $q_0 = u$  and  $\forall i \geq 0, q_{i+1} \in \bigcap_{a \in A} f_a(q_0 \dots q_i)$ . According to the construction of  $\sigma_a$ , we can say that  $\forall i \geq 0, \forall a \in A, \{h \mid q_{i+1} \in h\} \subseteq \sigma_a(q_i)$ , and then  $\{h \mid q_{i+1} \in h\} \subseteq \sigma_A(q_i)$ . Then the concatenation  $path(u)\lambda \in Adh(\sigma_A)$  and thus  $out(u, F_A) \subseteq Adh(\sigma_A)$ .

( $\supseteq$ ) Suppose  $h \in Adh(\sigma_A)$ , and  $\sigma_A \in Strategy_A^u$ . This means that  $h \in Adh(\sigma_a)$  for all  $a \in A$ : therefore we have (i)  $Dom(\sigma_a) \cap h \neq \emptyset$  and (ii)  $\forall w \in Dom(\sigma_a) \cap h, h \in \sigma_a(w)$ . By definition,  $u \in Dom(\sigma_a) \cap h, \forall a \in A$ . According to the construction of  $\sigma_a$  we can say that for

all  $w \in h$  that appear in  $Tree_w$ ,  $f_a(path(w)) \cap h \neq \emptyset$ , and therefore  $(\bigcap_{a \in A} f_a(path(w))) \cap h \neq \emptyset$ . Because  $h$  is a maximal set of linearly ordered moments from  $W$  containing  $u$ , we have that  $h = path(u)q_1q_2 \dots$  with  $q_1 = \bigcap_{a \in A} f_a(path(u))$ , and such that  $q_{i+1} \in \bigcap_{a \in A} f_a(path(q_i))$ . Then  $h \in out(u, F_A)$  and  $Adh(\sigma_A) \subseteq out(u, F_A)$ .

We conclude that  $out(u, F_A) = Adh(\sigma_A)$ . ■

**Theorem 6.1.** *If  $\varphi$  is ATL-satisfiable then  $tr(\varphi)$  is STIT-satisfiable.*

PROOF. Suppose given an ATS  $\mathcal{M}_{ATL} = \langle W_{ATL}, \delta, v_{ATL} \rangle$  and  $w \in W_{ATL}$  s.t.  $\mathcal{M}_{ATL}, w \models \varphi$ . W.l.o.g.  $\mathcal{M}_{ATL}$  is tree-like. We translate it into  $\mathcal{M}_{STIT} = \langle W_{STIT}, Choice, <, v_{STIT} \rangle$ , as described above. Hence by Proposition 6.2,  $\mathcal{M}_{STIT}$  is a STIT model. We prove by structural induction on  $\varphi$  that  $\mathcal{M}_{ATL}, w \models \varphi$  iff  $\mathcal{M}_{STIT}, w/h \models tr(\varphi), \forall h \in H_w$ .

Cases of atomic formulae, negations and disjunctions are trivial, and we here only present the cases of the modal operators.

- Case  $\psi = \langle\langle A \rangle\rangle \mathbf{X}\gamma$ . This means that there is an  $F_A$  s.t. for all  $\lambda \in out(w, F_A)$  we have  $\mathcal{M}_{ATL}, \lambda[1] \models \gamma$ . So by induction hypothesis, for all  $\lambda \in out(w, F_A)$  we have  $\mathcal{M}_{STIT}, \lambda[1]/h \models tr(\gamma)$  for all  $h \in H_{\lambda[1]}$ . By Lemma 6.2, we know that we can construct a collective strategy  $\sigma_A \in Strategy_A^w$  s.th.  $out(w, F_A) = Adh(\sigma_A)$ . So, there is  $\sigma_A$  s.th. for all  $h \in Adh(\sigma_A)$ , we have  $\mathcal{M}_{STIT}, \lambda[1]/h \models tr(\gamma)$ . By construction of  $<$ , and according to the definition of the  $\mathbf{X}$ -operator, this means that  $\mathcal{M}_{STIT}, w/h \models \mathbf{X}tr(\gamma)$ , and we obtain that  $\mathcal{M}_{STIT}, w/h \models \Diamond_s[A \text{ scstit} : \mathbf{X}tr(\gamma)]$ .
- Case  $\psi = \langle\langle A \rangle\rangle \mathbf{G}\gamma$ . This means that there is an  $F_A$  s.th. for all  $\lambda \in out(w, F_A)$  we have  $\mathcal{M}_{ATL}, \lambda[i] \models \gamma, \forall i \geq 0$ . By induction hypothesis, for all  $\lambda \in out(w, F_A)$  we have  $\mathcal{M}_{STIT}, \lambda[i]/h \models \gamma, \forall i \geq 0, \forall h \in H_{\lambda[i]}$ . By Lemma 6.2, there is  $\sigma_A \in Strategy_A^w$  s.th. for all  $h \in Adh(\sigma_A)$ , we have  $\mathcal{M}_{STIT}, \lambda[i]/h \models tr(\gamma), \forall i \geq 0$ . By construction of  $<$ , and according to the definition of the  $\mathbf{G}$ -operator, this means that  $\mathcal{M}_{STIT}, w/h \models \mathbf{G}tr(\gamma), \forall h \in Adh(\sigma_A)$ , and we obtain that  $\mathcal{M}_{STIT}, w/h \models \Diamond_s[A \text{ scstit} : \mathbf{G}tr(\gamma)]$ .
- Case  $\psi = \langle\langle A \rangle\rangle \gamma_1 \mathbf{U} \gamma_2$ . This means that there is an  $F_A$  s.th. for all  $\lambda \in out(w, F_A)$  there exists an  $i \geq 0$  s.th. we have  $\mathcal{M}_{ATL}, \lambda[i] \models \gamma_2$  and  $\forall j, 0 \leq j < i, \mathcal{M}_{ATL}, \lambda[j] \models \gamma_1$ . Using the same arguments as before, we get  $\mathcal{M}_{STIT}, w/h \models \Diamond_s[A \text{ scstit} : tr(\gamma_1) \mathbf{U} tr(\gamma_2)]$  for all  $h$  in  $H_w$ . ■

In addition, for the STIT-fragment corresponding to ATL, it holds that evaluation of formulas does not depend on the history (Horty calls this ‘moment determinedness’). This corresponds with the following property.

**Proposition 6.3.**  $\models_{STIT} tr(\varphi) \equiv \Box tr(\varphi)$

PROOF. The proof is done by induction on the form of  $\varphi$ . It uses the fact that the logic of historical necessity  $\Box$  is S5. ■

We need this proposition in our proof of Theorem 6.2 below.

**Theorem 6.2.** *If  $\models_{ATL} \varphi$  then  $\models_{STIT} tr(\varphi)$ .*

PROOF. We use the ATL axiomatization of [GvD06], and prove that translation of the axioms are valid, and that the translated inference rules preserve validity.

( $\perp$ ), ( $\top$ ), ( $N$ ), ( $S$ ) and ( $\langle\langle A \rangle\rangle X$ -Monotonicity) are axioms of Coalition Logic. Their translation to STIT preserves validity, as we have shown in [BHT06c, Theorem 4.2].<sup>8</sup>

If a formula is STIT-valid, it is true at each index of each STIT model. Then, it is obvious that the translation of ( $\langle\langle \emptyset \rangle\rangle G$ -Necessitation) preserves validity.

- The translation of ( $FP_G$ ) is

$$\Diamond_s[A \text{ scstit} : \mathbf{G}tr(\varphi)] \equiv tr(\varphi) \wedge \Diamond_s[A \text{ scstit} : \mathbf{X}\Diamond_s[A \text{ scstit} : \mathbf{G}tr(\varphi)]].$$

( $\Rightarrow$ ) The left side of the equivalence implies that there is an index where  $tr(\varphi)$  holds. By Proposition 6.3,  $\models_{STIT} tr(\varphi) \rightarrow \Box tr(\varphi)$ , and thus  $tr(\varphi)$  is true at any index of the current moment. If there exists a strategy such that  $tr(\varphi)$  is globally true along admitted histories, then the same strategy also satisfies the right part of the equivalence.

( $\Leftarrow$ ) The right side says there is a strategy  $\sigma_A$  at  $w$ , let us say with  $\sigma_A(w) = Q$ ,  $Q \in \text{Choice}_A^w$ , s.th. at the next step, there is a strategy  $\sigma'_A$  s.th.  $tr(\varphi)$  is globally true. Hence, the strategy  $\sigma''_A$  at  $w$ , defined as  $\sigma''_A(w) = Q$  and  $\forall u \in \text{Dom}(\sigma'_A) \setminus \{w\}, \sigma''_A(u) = \sigma'_A(u)$  satisfies that  $A$  can ensure at  $w$  that  $tr(\varphi)$  is globally true along histories in  $\text{Adh}(\sigma''_A)$ .

- The translation of ( $GFP_G$ ) is

$$\begin{aligned} \Diamond_s[\emptyset \text{ scstit} : \mathbf{G}(tr(\theta) \rightarrow (tr(\varphi) \wedge \Diamond_s[A \text{ scstit} : \mathbf{X}tr(\theta)]))] \rightarrow \\ \Diamond_s[\emptyset \text{ scstit} : \mathbf{G}(tr(\theta) \rightarrow \Diamond_s[A \text{ scstit} : \mathbf{G}tr(\varphi)])]. \end{aligned}$$

<sup>8</sup>The proof of the validity of the translation of the axiom ( $N$ ) involves Hypothesis 6.2 about determinism.

The left member means that it is settled that globally, if we have  $tr(\theta)$  then we also have  $tr(\varphi)$ , and there is strategy s.th.  $tr(\theta)$  is true at the next step. It implies that whenever  $tr(\theta)$  is true, it exists a choice partition that ensures that  $tr(\theta)$  holds at the next step. Thus the strategy  $\sigma_A$  which as soon as  $tr(\theta)$  is true, chooses at each step such a choice partition, ensures that  $tr(\varphi)$  is globally true along histories of  $Adh(\sigma_A)$  (and this, whatever we choose before getting  $tr(\theta)$ ).

- The translation of  $(FP_U)$  is

$$\begin{aligned} \diamond_s[A \text{ scstit} : tr(\psi)\mathcal{U}tr(\varphi)] \equiv \\ tr(\varphi) \vee (tr(\psi) \wedge \diamond_s[A \text{ scstit} : \mathbf{X}\diamond_s[A \text{ scstit} : tr(\psi)\mathcal{U}tr(\varphi)]]). \end{aligned}$$

We prove its validity by using the same arguments as for  $(FP_G)$ .

- The translation of  $(LFP_U)$  is

$$\begin{aligned} \diamond_s[\emptyset \text{ scstit} : \mathbf{G}((tr(\varphi) \vee (tr(\psi) \wedge \diamond_s[A \text{ scstit} : \mathbf{X}tr(\theta)])) \rightarrow tr(\theta))] \rightarrow \\ \diamond_s[\emptyset \text{ scstit} : \mathbf{G}(\diamond_s[A \text{ scstit} : tr(\psi)\mathcal{U}tr(\varphi)] \rightarrow tr(\theta))]. \end{aligned}$$

We use the fact that  $\diamond_s[\emptyset \text{ scstit} : \varphi] \equiv \Box\varphi$  (Proposition 6.1), that  $\Box\mathbf{G}(\varphi \rightarrow \psi) \rightarrow (\Box\mathbf{G}\varphi \rightarrow \Box\mathbf{G}\psi)$  and that  $(\alpha \rightarrow (\beta \rightarrow \gamma)) \equiv (\beta \rightarrow (\alpha \rightarrow \gamma))$ . Thus, we have to prove that  $\beta \rightarrow (\alpha \rightarrow \gamma)$  with  $\beta \equiv \Box\mathbf{G}\diamond_s[A \text{ scstit} : tr(\psi)\mathcal{U}tr(\varphi)]$ ,  $\alpha \equiv \Box\mathbf{G}((tr(\varphi) \vee (tr(\psi) \wedge \diamond_s[A \text{ scstit} : \mathbf{X}tr(\theta)])) \rightarrow tr(\theta))$  and  $\gamma \equiv \Box\mathbf{G}tr(\theta)$ .

Suppose that  $\mathcal{M}, w/h \models \diamond_s[A \text{ scstit} : tr(\psi)\mathcal{U}tr(\varphi)]$ . This means that there is a strategy  $\sigma_A$  s.th.  $\forall h \in Adh(\sigma_A), \exists w_1 \in h, w < w_1$  s.th.  $\mathcal{M}, w_1/h \models tr(\varphi)$  and  $\forall w_2, w \leq w_2 < w_1, \mathcal{M}, w_2/h \models tr(\psi)$ . By  $\alpha$ ,  $tr(\theta)$  is true at  $w_1$ . If  $w_1 = w$  then it is sufficient to conclude. Else, we have  $tr(\psi)$  true at the immediate predecessor of  $w_1$  on  $h$ . So by  $\alpha$ , we also have  $tr(\theta)$ , since  $\diamond_s[A \text{ scstit} : tr(\theta)]$  is true. Still, recursively (this induction is allowed by countably infiniteness of Hypothesis 6.1) as  $tr(\psi)$  is true at each  $w_3 \in h$  s.th.  $w \leq w_3 < w_1$ , we also get  $tr(\theta)$  at  $w_3$  and in particular  $\mathcal{M}, w/h \models tr(\theta)$ . ■

**Corollary 6.1.**  $\varphi$  is satisfiable in ATL iff  $tr(\varphi)$  is satisfiable in STIT.

PROOF. As an immediate corollary of Theorems 6.1 and 6.2 and van Drimelen et al.'s completeness proof for ATL. ■

## 6.6 Discussion

The main contribution of this chapter is, we believe, to build a bridge between two formalisms with a rather different background; The STIT for-

malism originating in philosophy, and ATL originating in computer science (multiagent systems). In this section, we discuss details of our embedding. We address in what sense, and under what assumptions, ATL appears to be a well-identified fragment of a more general and philosophically grounded theory of agency. These assumptions are then insightful and suggestive of a shared core between computer science and the philosophy of agency/action.

It should be noted first that Horty's strategic ability only applies to *individual* agency. Hence, we had to define admitted histories for a *collective* strategy, as the intersection of individual ones. However, this is a straightforward extension of the definition of collective choices; we believe we have neither violated a fundamental aspect of STIT nor forced the embedding by adding too much to the semantics.

We also added some constraints to the original theory of agents and choices in branching time to guarantee that the proposed translation works well. We view these constraints as both relevant and harmless. The constraints are:

1. Histories are isomorphic to the set of natural numbers.
2.  $\forall w \in W, \exists w' \in W (w < w' \text{ and } \bigcap_{a \in \mathcal{A}gt} s_w(a) = H_{w'})$   
Intersection of agents of  $\mathcal{A}gt$ 's choices is *not only* nonempty (which is the only restriction in the original STIT) but must exactly be the set of histories passing through a next moment.

The second condition is the simple counterpart of the ATL constraint stating that when every agent in  $\mathcal{A}gt$  opts for an action then the next state of the world is completely determined. Here we just say that in STIT, the intersection of all agents of  $\mathcal{A}gt$ 's choices must be exactly the set of histories passing through this very completely determined moment.<sup>9</sup> As discussed in [GJ04], the condition of determinism is not a limitation of the modelling capabilities of the language, since we could introduce a neutral agent 'nature', in order to accommodate non-deterministic transitions. Hence, this constraint on ATs should not be considered a fundamental distinction between the two formalisms.

The main difference then concerns the first constraint, that permits us to define the **X** operator, and then to grasp the concept of *next* moments

---

<sup>9</sup>Actually, this condition does not explicitly refer to the *next* moment, but to a future moment. It is nevertheless sufficient, because for all  $h \in H_w$ , and for all  $w' < w$ , we have  $h \in H_{w'}$ . ( $\langle W, < \rangle$  is a tree.)

and outcomes. More generally, it allows us to stick to standard LTL expressivity for temporal properties of paths. This same assumption applies to the temporal component of ATL. This imposes a particular view on time. However, deliberately, Belnap and colleagues do not take a position on the nature of time.

“For this reason the present theory of agency is immediately applicable regardless of whether we picture succession as discrete, dense, continuous, well-ordered, some mixture of these, or whatever; and regardless of whether histories are finite or infinite in one direction or the other.” ([BPX01, p.196].)

Although, from a philosophical point of view, it makes sense wanting to be as general as possible, in computer science it is very common and natural to model the temporal evolution of a system using a transition system. This brings with it a view on time as being discrete. Isomorphism with the natural numbers (and thus non-density) is often assumed in order to keep complexity within acceptable limits, and to avoid discussions about philosophical difficulties reminiscent of problems raised by presocratic philosophers typified by Zeno of Elea: how can time proceed (i.e., how can we interpret a ‘next’ operator) if there is always a moment between two moments? This justifies the assumption concerning isomorphism with the natural numbers.

However the differences in the temporal fragments of both frameworks do not only concern the models, but also the syntax. In particular, note that in STIT we can nest temporal operators without any restriction. In ATL this is syntactically disallowed. In ATL\* we do not have this restriction. However, in some definitions for this stronger logic we cannot unravel ATs into trees under preservation of satisfaction of formulas. It is worth noting that even though STIT operators (e.g.,  $[_\text{cstit}: \_]$ ,  $[_\text{dstit}: \_]$ ,  $\diamond_s[_\text{scstit}: \_]$ ) admit a propositional formula as a complement, they are traditionally intended as referring to a statement about the future.<sup>10</sup> In this respect, the syntactic limitation of ATL against ATL\* should not be considered a philosophical issue.

Obviously, STIT and ATL have some striking resemblances. The concepts of *agent* and *choice* are the same in both theories. In the theory of agents and choices in branching time, agents are “individuals thought of as making choices, or acting, in time” ([BPX01, p.33]). Belnap, as the most

<sup>10</sup>Belnap et col. discuss that in [BPX01, p. 37] for deliberative stit but the same holds for other agentive operators of the logic.

active author of STIT theory, has stressed that STIT agency is not restricted to persons or intentional agents and could equally be applied to processes making random choices. Actions are thus idealized in a way that ignores any mental state. STIT is only interested in the causal structure of choice, regardless of its content. To put it in yet other words, choices are just *objective possibilities* of an agent, selecting some possible courses of time and ruling out some others. All of this equally applies to ATL, where each agent selects a set of next states, and time will go through a state in the intersection of every agent's selection.

Also the notion of *independence of choices* (or equally *independence of agents*) applies to both frameworks. Agent's choices must be non-blocking, i.e., for each possible choice of some agent, the intersection with all possible choices of other agents is non-empty. Belnap et col. admit this to be a fierce constraint. For instance, it follows that two agents  $a$  and  $b$  cannot possibly have identical sets of choices at the same moment  $w$  (in general  $Choice_a^w \neq Choice_b^w$ ) except the vacuous one (in this case,  $Choice_a^w = Choice_b^w = \{H_w\}$ ). It also follows that in STIT, there are not less than  $\prod_{a \in Agt} |Choice_a^w|$  histories passing through a moment  $w$ . Nevertheless the constraint is considered commonplace. In STIT theory it has been argued that if an agent can deprive other agents of some of their choices without any priorities in the causal order, then we would have to deal with a 'quantum' version of agency. We refer the reader to Section 5.2.3 for more details on the consequences of the independence of choices.

ATL structures are not limited to trees. But, as described in [Wöl04], an ATS  $\langle W, \delta, v \rangle$  can easily be unraveled to an ATS where the transition function  $\delta$  in  $\langle W, <_\delta \rangle$  is a tree. ATL, like all other modal formalisms,<sup>11</sup> cannot distinguish the original model from its unraveling into a tree. STIT and ATL thus both embed in branching time structures limited to trees. However, what we show in Section 6.3.2 is stronger. Lemma 6.1 tells us that we can unravel any ATS in a tree satisfying the property that choices of every agent, represented as sets of 'possibly chosen next states', are partitioning the 'possible next states'. Hence, from any ATS, we can construct a bisimilar ATS that meets the constraint STIT imposes to the  $Choice_w^a$  functions. There is no need to enforce this on ATL frames as in [Wöl04]. The property of *empty intersection* of the different simultaneous classes of choice in STIT is not expressible in modal logic.

---

<sup>11</sup>At least this is true according to Van Benthem's definition of 'modal logic' as the bisimulation-invariant subset of first order logic [vB84].

It is worth noting that the present translation is compatible with the one we have proposed in [BHT06c] for Coalition Logic, together with Goranko's translation of  $\langle\langle A \rangle\rangle\varphi$  to  $\langle\langle A \rangle\rangle X\varphi$ .

With the completeness result for ATL in [GvD06], one immediate benefit of our translation is to identify a complete axiomatization of a fragment of STIT. As an interesting perspective, ATL model checking can be applied to a fragment of the STIT languages.

We conclude with the remark [BPX01, p.18] that STIT theory should be understood as a formal characterization of agency, permitting to postpone an ontology. One merit of this work is then to push a significant justification for ATL as an elegant and well-founded framework of agency.



# 7

---

## Ontology of Agency and Action

### 7.1 Introduction

Action and agency are crucial notions for a variety of application domains, e.g., multiagent systems and interaction modelling, planning and robotics, law and social modelling... Accordingly, many different research areas, among which the quite rich discipline of philosophy of action, have proposed theoretical accounts. Unfortunately, these proposals are often unrelated; a correlate is that no well-developed ontology of action and agency is currently available. This chapter is a first attempt at bridging this gap, focusing especially on the relationship between agency and action, mostly studied separately.

As we have seen, STIT logic is a particularly useful logical systems dealing with *agency*, both in terms of expressivity and formal properties. The key idea of agency comes from Anselm around the year 1100, who argued that acting is best described by what an agent brings about. Agency is thus the relationship between an agent (or a group of agents) and the states of affairs it can bring about, without referring to how this is done, i.e., the actions performed. Reducing the ontological commitment is of course positive, but if one wants to reason on actions themselves, considering their preconditions, distinguishing between different ways of reaching a given state of affairs, analysing the internal structure of the action (its participants other than the agent, its way of unfolding in time) and its essential relationship with the agent's mental states, avoiding to introduce actions in the picture becomes impossible.

STIT is a propositional modal logic. Integrating agency and actions in the same framework could be done by extending STIT with some other modal operators dealing more explicitly with actions like those of PDL; this path has begun to be explored in Section 5.4.2. However, with modal operators, the domains of interest and their ontological properties are not made explicit in the language but left hidden in the models. Another di-

rection is to work directly in the more expressive framework of first-order logic, more suitable to easily formulate many properties and explore the variety of possible ontological choices.

The methodology chosen for the work presented here is therefore to first express the ontological assumptions of STIT in a first-order theory, called OntoSTIT; this is the purpose of Section 7.3, after a brief justification of STIT as a basis of an ontology via a summary of its formal properties in Section 7.2. Then, we propose in Section 7.4 to extend this theory by enlarging its language and its domain of interpretation to include actions proper.

## 7.2 Motivations

### 7.2.1 Why going first-order?

Up until now, we have been studying modal logics of agency, and we have seen in Chapter 5 some of limitations of these logics. We think that in order to go further in the analysis, we need to enrich the vocabulary and the structures of agency. This is the purpose of the present chapter, and a natural strategy for gaining expressivity is move up to first-order logic.

We would like to explore the richness of  $BT + AC$  structures on the one hand, and to investigate the gain of possible additions of concepts to the domain. Those additions, as it is often the case in formal ontologies, must be grounded in the literature in philosophy which helps us to identify which concepts are worth expressing in a theory of agency. Starting from STIT models is already motivated by its foundations in philosophy of action, but we think much more should be grasped in the whole picture of agency.

However, our aim is not to consider the resulting ontology as an end in and of itself. We rather see this work as a basis for isolating some promising concepts with respect to a particular application, and that could ideally be captured in a more friendly modal logic. This is not done in this thesis: we just aim at pointing out some paths that of investigation we consider promising.

### 7.2.2 Summary of formal aspects of STIT

STIT is not the only logic of agency. In particular, a family of logics that we could describe as logics of *bringing it about* is reminiscent from [Pör70] and has been largely studied. But what we see as a defect of those logics for

our present purpose is that it lacks temporal aspects that we would wish to rely on for a fine-grained ontology of agency and action.

STIT on its side, enjoys formal properties that make it particularly attractive. One such property is that STIT is more expressive than two well-known logics of *ability* that come from computer science and social software, namely ATL and CL (cf. [BHT06c, BHT06b] and Chapters 4 and 6). Ability is *possible agency*. While agency links an agent to what it *does* bring about, ability links an agent to what it *can* ensure. *Alternating-time Temporal Logic* is a direct extension of *Computational Tree Logic* [CE01] with multiagent systems capabilities. From this famous branching time logic, it adds agents and coalitions of agents who can opt, at every state (or ‘choice point’), for a particular subset of the possible courses of time. *Coalition Logic* [Pau02] has been introduced independently as a logical tool for reasoning about social procedures. Such procedures are exemplified by fair-division algorithms or voting processes, and are characterized by complex strategic interactions between agents. As shown in [Gor01], CL corresponds to a fragment of ATL restricted to some operators. The fact that STIT is more expressive than CL and ATL allows to inherit from the various works in those logics which are very active fields. It is this permissibility of STIT at modeling important properties concerning computer science and game theory that motivates its choice as a starting point of an ontology of action.

The second important property of STIT is the decidability of its core. What we call the *core* of STIT is the logic characterized by the axiomatization presented in Section 3.2. It is the logic of the so-called Chellas stit. We have seen that Chellas STIT brings with it the important notion of *independence of agents*: one agent cannot deprive another from its choices. It can be captured by the observation that an agent  $a$  sees to it that another agent  $a$  sees to it that a state of affairs  $\varphi$  holds only if  $\varphi$  is necessary, and will be formalized by  $[a \text{ cstit} : [b \text{ cstit} : \varphi]] \rightarrow \Box\varphi$ . Its decidability makes the Chellas’ STIT an appropriate tool for reasoning, even though the complexity of the satisfiability problem is now known NEXPTIME-complete (see Chapter 3). For comparison, it is the complexity of the problems of concept satisfiability [Tob01, Lut04] and ABox consistency [Sch94] in OWL-DL.

## 7.3 STIT Ontology of Agency – OntoSTIT

### 7.3.1 A modal or an ontological approach?

As explained in last section, STIT is a quite expressive logic of agency. It has very important formal properties, and accordingly, it knows a growing influence. However, from the ontological point of view, it is not totally clear to what extent STIT captures the intuitions of agency, and how this relates to the notion of action, in particular as it is studied in the philosophy of action.

It is well known that propositional modal logic has expressivity limitations in comparison with first order logic; this is actually why it has better calculability properties. But whereas the latter enables the expression of rich theories capturing almost all intuitions, the former forces us to tie our intuitions into an at times uncomfortable suit. In this sense it is not surprising that inside the ontological community those who deal with the concept of action and agency have little or no interest in the theory of agents and choices in branching time. Belnap et al. complained:

“The modal logic of agency is not popular. Perhaps largely due to the influence of Davidson (see the essays in [Dav91]), but based also on the very different work of such as [Gol70] and [Tho77], the dominant logical template takes an agent as a wart on the skin of an action, and takes an action as a kind of event. This ‘actions as events’ picture is all ontology, not modality, and indeed, in the case of Davidson, is driven by the sort of commitment to first-order logic that counts modalities as Bad.” [BPX01].

On the other hand – the argument goes – STIT is philosophically well motivated and “has the advantage that it permits us to *postpone attempting to fashion an ontological theory*, while still advancing our grasp of some important features of action...” [BPX01].

Although, as said earlier, is it true that the first-order framework is more adequate to ontological studies, we would like to draw a slightly different picture from Belnap’s. As any representation framework, propositional modal logics do carry ontological assumptions, even though these are often hidden in properties of their models rather than explicitly stated in the language. So, even though the focus in STIT work has (deliberately) not been put on ontological questions, STIT is already in some sense an ontology of agency.<sup>1</sup>

---

<sup>1</sup>This is more specifically argued in [GG95] but also [TV07].

In order to clarify what are STIT's ontological assumptions and establish a base ground on which to build a richer ontology of agency and action, we will in the remaining of this chapter, extract those features of action captured by STIT, and make them explicit in a first-order theory that can be proved equivalent to it, that we will call *OntoSTIT*.

### 7.3.2 From STIT to OntoSTIT

We present the new three-sorted first-order language we will be using, and then the axiomatic theory that we call *OntoSTIT*.

We would like to note that there are few changes to the *OntoSTIT* language and axiomatization with respect to its first publication in [TTV06]. We originally designed *OntoSTIT* as an unsorted first-order logic. It was already complete with respect to the class of *BT + AC* models. However, we here prefer to present the theory as a three-sorted FOL. This modification, obviously, does not change expressivity of our theory, but it significantly increases its clarity and transparency.

In the first version of *OntoSTIT*, we have introduced three kinds of individuals, namely agents, moments and histories and we were postulating that our domain contains at least one individual of each of these three types (see axioms *As6*, *As7* and *As9* in [TTV06, p.184]). In the current presentation of *OntoSTIT*, we obtain the same result by introducing tree sorts 0, 1 and 2, containing, respectively, agents, moments and histories. Non-emptiness of each sort is guaranteed by definition of many-sorted first-order logic.

Moreover, the new axiomatization that we present in this section, does not need to contain characterization axioms that restrict the arguments of each relation (see axioms *As1*, *As8* and *As12* in [TTV06, p.184–185]). In the sorted formulation of *OntoSTIT*, the kind of arguments for all relations is specified in the definition of the language, by variable symbols referring to sorts.

#### 7.3.2.1 Language

*OntoSTIT* is a theory of standard three-sorted first-order logic with identity and its language,  $L_{OntoSTIT}$ , is defined in a standard way. We nevertheless assume the following conventions for variables and constants symbols of  $L_{OntoSTIT}$ :

- individual variables of sort 0 ranging over particular agents:  $a, a', \dots, a_1, \dots, a_n$

- individual constants of sort 0 denoting agents:  $\mathbf{a}, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$
- individual variables of sort 1 ranging over particular moments:  $m, m', \dots, m_1, m_2, \dots, m_n$
- individual constants of sort 1 denoting moments:  $\mathbf{m}, \mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n$
- individual variables of sort 2 ranging over particular moments:  $h, h', \dots, h_1, h_2, \dots, h_n$
- individual constants of sort 2 denoting moments:  $\mathbf{h}, \mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n$
- $\Delta = \{In, Pre, PO\}$  is a set of constants denoting (primitive) universals s.t.:
  - “*In*” is of sort  $1 \times 2$ ,
  - “*Pre*” is of sort  $1 \times 1$ ,
  - “*PO*” is of sort  $0 \times 1 \times 2 \times 2$ .
- $\Pi = \{P_1, P_2, \dots\}$  is a finite set of constants denoting universals of the sort  $1 \times 2$ .

The predicate constants of the set  $\Delta$  are understood as, respectively, incidence between a moment and a history, precedence between moments and the relation such that for an agent at a certain moment two histories are both “possible outcomes” of its underlying actions.

The models of OntoSTIT are those of STIT, the class of models  $\mathcal{M}$ . The domain of quantification in which variables of all three sorts are interpreted covers, respectively, agents, moments and histories.

We’d like to note that in the first version of OntoSTIT, we did not express explicitly the fact that three classes of individuals, which existence we were postulating, are disjoint. We get this property in new formulation by the fact that introduced sorts 0, 1 and 2 are declared to be disjoint.

Our language contains also a set of predicate constants  $\Pi$  that corresponds, in 1–1 relation, to the set of atomic propositions of STIT. Each element of  $\Pi$  is interpreted in the equivalent way, in  $\mathcal{M}$ , as the atomic proposition of the set *Atm*. Note, that in [TTV06] we have been using three-place predicate  $HOLDS(m, h, \mathbf{p})$  for saying that the (reified) proposition  $\mathbf{p}$  holds in index  $m/h$ . In present formulation of OntoSTIT, we express the same idea by writing  $P(m, h)$  with  $P \in \Pi$ .

### 7.3.2.2 Characterization of primitive relations and categories

**Order on moments.** The precedence relation  $Pre$  is transitive (Ao1) and irreflexive (Ao2). The linearity in the past is expressed by (Ao3). (Ao4) says that any two moments have a lower bound (historical connection).

$$(Ao1) \quad Pre(m, m') \wedge Pre(m', m'') \rightarrow Pre(m, m'')$$

$$(Ao2) \quad \neg Pre(m, m)$$

$$(Ao3) \quad Pre(m, m'') \wedge Pre(m', m'') \rightarrow m = m' \vee Pre(m, m') \vee Pre(m', m)$$

$$(Ao4) \quad \forall m, m' \exists m'' ((Pre(m'', m) \vee m'' = m) \wedge (Pre(m'', m') \vee m'' = m'))$$

**Moments and histories.** In STIT models, a history is a set of moments and the relationship between a moment and a history is expressed by  $m \in h$ . In OntoSTIT language, a history is denoted by a particular individual and no set theoretical axioms are assumed. We simply express the relation between moments and histories by the relation  $In(m, h)$ : “the moment  $m$  is in the history  $h$ ” or “the history  $h$  passes through the moment  $m$ ”. For any moment, there is some history that passes through it (Ao5). (Ao6) is an axiom schema ensuring that when a predicate  $P_i$  holds at the moment  $m$  and the history  $h$ ,  $m$  is in  $h$ .

$$(Ao5) \quad \forall m \exists h (In(m, h))$$

$$(Ao6) \quad P_i(m, h) \rightarrow In(m, h)$$

**Histories.** That histories denote linearly ordered sets is guaranteed by axiom (Ao7) and the fact that all histories are maximally linearly ordered set is expressed by (Ao8). The predicate  $UD$ , for undivided, can be defined: two histories  $h$  and  $h'$  are undivided at moment  $m$  if and only if for some moment  $m'$  later than  $m$ , it is the case that  $m'$  is in  $h$  and  $h'$  (Do1).

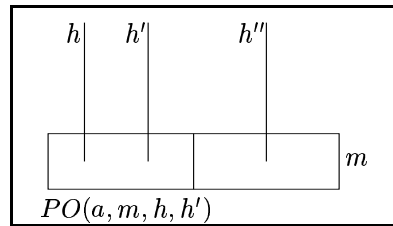
$$(Ao7) \quad (In(m, h) \wedge In(m', h)) \rightarrow (m = m' \vee Pre(m, m') \vee Pre(m', m))$$

$$(Ao8) \quad \forall h \neg \exists h' (h \neq h' \wedge \forall m (In(m, h) \rightarrow In(m, h')))$$

$$(Do1) \quad UD(h, h', m) \triangleq \exists m' (Pre(m, m') \wedge In(m', h) \wedge In(m', h'))$$

**Possible Outcome.** The predicate  $PO(a, m, h, h')$ , for possible outcome, expresses the intuitions that are behind the *Choice* function in STIT: at moment  $m$ , histories  $h$  and  $h'$  – that pass through  $m$  (Ao9)<sup>2</sup> – are the possible outcomes of some action performed by agent  $a$  (see Figure 7.1). We call the histories  $h$  and  $h'$  ‘possible outcomes’ because each of them *might* result from a same action performed by the agent  $a$  at  $m$  although it cannot determine which will be the eventual one. In other words, an agent by his action restricts the possible futures to those histories that are possible outcomes of his action. Note that as STIT, OntoSTIT does not explicitly model action. In other words, actions are not present as individuals in our ontology. That is why we cannot express the intuition, neither in STIT nor in OntoSTIT, that an agent performs a particular action. However this will be possible in OntoSTIT+ (see Section 7.4).

$$(Ao9) PO(a, m, h, h') \rightarrow In(m, h')$$



**Figure 7.1:** At the moment  $m$ , the histories  $h$  and  $h'$  are the possible outcomes of some action performed by the agent  $a$ . At the moment  $m$ , the histories  $h'$  and  $h''$  are not the possible outcomes of some action performed by the agent  $a$ .

When one fixes the first two arguments,  $PO$  is an equivalence relation. It is reflexive (Ao10), transitive (Ao10) and symmetric (Ao12):

$$(Ao10) In(m, h) \rightarrow PO(a, m, h, h)$$

$$(Ao11) PO(a, m, h, h') \wedge PO(a, m, h', h'') \rightarrow PO(a, m, h, h'')$$

$$(Ao12) PO(a, m, h, h') \rightarrow PO(a, m, h', h)$$

Axiom (Ao13) says that histories that are undivided at moment  $y$  are possible outcomes of the same action.

$$(Ao13) PO(a, m, h, h') \wedge UD(h', h'', m) \rightarrow PO(a, m, h, h'')$$

<sup>2</sup>Because of (Ao12), in axiom (Ao9) we do not need to explicitly write that also  $In(m, h)$ .

The axiom schema (Ao14) expresses the independence of choices. It means that at each moment  $y$  there is at least one history  $t$  that is common to possible outcomes of all agents' actual choices.

**(Ao14)**  $PO(a_1, m, h_1, h'_1) \wedge \dots \wedge PO(a_k, m, h_k, h'_k) \rightarrow \exists h(PO(a_1, m, h_1, h) \wedge \dots \wedge PO(a_k, m, h_k, h))$ , for any  $k > 1$

Following the techniques of the Standard Translation [BdRV01] and 'T-encoded semantics' [Ohl98, Man96], it has been shown in [TV07] that CSTIT is equivalent to a special sub-theory of OntoSTIT.

### 7.3.3 Agency in OntoSTIT

The idea of agency is expressed in OntoSTIT by two concepts: possible outcomes ( $PO$ ) and the predicates from the set  $\Pi$  which are intended to be treated as effects of choice/action. This means that actions themselves are not present in our three-sorted first-order theory as they are not present in STIT. We can express in OntoSTIT that an agent saw to it that some state of affairs holds (e.g. *the light is off*), even though we still cannot explicitly say by means of which action the agent has done it (we cannot make sure that *the agent switched off the light* rather than *the agent unscrewed the bulb*).

Consider the instantaneous action of *switching off the light* performed by *Robert*, now. In STIT we can only say that *Robert* sees to it that *the light is off*, leaving the action by means of which the result was achieved unexpressed:

$$[Robert\ cstit: Light-is-off].$$

The same situation can be easily expressed in OntoSTIT by assuring that in all possible outcomes,  $h$ , of this action it is the case that the *light is off* (we assume that the actual moment is named  $\mathbf{n}$  and it is in the actual history  $\mathbf{h}$ ):

**(Es1)**  $\forall h(PO(\mathbf{Robert}, \mathbf{n}, \mathbf{h}, h) \rightarrow Light-is-off(\mathbf{n}, h))$

What is more, we want to say that *Robert switches off the light* is true only if *the light was on* just before the action was performed:

**(Es2)**  $\forall m(Pre(m, \mathbf{n}) \rightarrow \exists m'(Pre(m, m') \wedge Pre(m', \mathbf{n}) \wedge \neg Light-is-off(m', \mathbf{h})) \wedge \forall h(PO(\mathbf{Robert}, \mathbf{n}, \mathbf{h}, h) \rightarrow Light-is-off(\mathbf{n}, h)))$ .

In OntoSTIT (as in STIT) we can also express the idea that an agent brought about some state of affair but could not have done it or simply it could have happened that that state of affair does not hold. For example we say that *Robert switches off the light*, now, but also that *the light might have been still on*.

$$(Es3) \quad \forall h(PO(\mathbf{Robert}, \mathbf{n}, \mathbf{h}, h) \rightarrow Light-is-off(\mathbf{n}, h)) \wedge \exists h'(In(\mathbf{n}, h') \wedge \neg Light-is-off(\mathbf{n}, h'))$$

In STIT this formula can be expressed by  $[Robert\ cstit : Light-is-off] \wedge \neg \Box(Light-is-off)$  which is equivalent to the formula  $[Robert\ dstit : Light-is-off]$ .

Note that in (Es1), (Es2) and (Es3) the moment of choice,  $\mathbf{n}$ , and the moment in which the effect of the action (*the light is off*) comes out, are the same. This expresses the assumption that the action thanks to which the result “*the light is off*” was brought about (e.g. *switching off the light*), is punctual or instantaneous (compare with [BPX01, p.33]). Instantaneity tightly binds the outcome of the action to the choice of performing that action. We may understand a choice here as an *intentional state* which not only triggers an action but is also responsible for controlling and directing the underlying action towards the result, what, indeed, seems to express an agency.

### 7.3.3.1 Agentive and causal gaps

Nevertheless it is possible to separate the moment of choice,  $\mathbf{n}$ , and the moment of appearance of the outcome,  $m$ . In STIT one can do it by extending its language by Prior-Thomason temporal operators. (That is, as we commented in Section 5.2.4.) In such multi-modal logic one can express the fact that *Robert* sees to it that in the future *Light-is-off* ( $\mathbf{F}$  stands for “in the future”):

$$[Robert\ cstit : \mathbf{F}(Light-is-off)].$$

Similarly, we may split a choice and an outcome in OntoSTIT:

$$(Es4) \quad \forall h(PO(\mathbf{Robert}, \mathbf{n}, \mathbf{h}, h) \rightarrow \exists m(Pre(\mathbf{n}, m) \wedge Light-is-off(m, h)))$$

However, splitting a moment of choice and a moment when an outcome appears creates some serious problems, namely an agentive and a causal gap.

**Agentive gap.** Let us consider *swimming* and the specific action *Michele swims from point A to point B*. This action belongs to the group of actions that do not go beyond bodily movements.<sup>3</sup> In STIT, Michele’s action is

<sup>3</sup>As Davidson pointed out in [Dav91, p.59]: “We never do more than move our bodies: the rest is up to nature”. Also [Sea01] shares Davidson’s view that no action goes beyond bodily movement. Here we do not take issue on this.

expressed by the sentence: *(at point A, now) Michele sees to it that he will be in point B*:

$$[Michele\ cstit: \mathbf{F}(Michele\text{-is-in-point-B})].$$

which, if true, means that at the moment of the choice, when Michele is in point A, Michele is guaranteed to reach point B. In a similar way, we formulate the same thought in OntoSTIT:

$$\text{(Es5)} \quad \forall h(PO(\mathbf{Michele}, \mathbf{n}, \mathbf{h}, h) \rightarrow \exists m(Pre(\mathbf{n}, m) \wedge Michele\text{-is-in-point-B}(m, h)))$$

In other words, Michele's choice at point A determines the fact that he is after some time in point B and this very choice cannot be changed anyhow. This is because all actions (or rather, all choices) are successful in STIT. This seems a far too strong assumption, as in real life, agents do change their minds considering for instance some changes in environment and actions can abort.<sup>4</sup> Thus, whenever one expresses an agency by means of STIT with temporal operators or one splits the moment of choice and moment of outcome in OntoSTIT, one has to deal with an agentive gap between the choice and the effects.

In philosophy of action and its mainstream represented *inter alia* by [Ans57], [Bra87], [Dav91, Essay 5], [Mel96] and [Sea01], the problem of agentive gap, for actions with duration that do not go beyond bodily movements, is solved by introducing an intention<sup>5</sup> concurrent with currently performed action. The role of this intention is triggering and sustaining an action. It is also responsible for controlling an action during its lifetime. Thus, the action can be terminated whenever an agent decides to stop it. As we have mentioned earlier, a choice seems to play a role of such intention in the context of a punctual action. However, when we want to take into account actions with duration, for which a choice is present only at their beginning, we face a problem of agentive gap. Thus, we would like to stress the fact that a choice cannot play the role of an intention in actions with duration.

**Causal gap.** A similar problem occurs in the case of actions that do go beyond bodily movement, as for example with *Booth's killing of Lincoln*, (Es6), by shooting him [Pie00]:

<sup>4</sup>Note that *reactivity*, i.e. "being responsive to change in the environment" is one of essential properties of rational agents [Woo00, p.2-5].

<sup>5</sup>Such intention has different names in literature, Searle calls it *intention in action*, Bratman – *present directed intention* and Mele refers to it by "*proximal intention*".

(Es6)  $\forall h(PO(\mathbf{Booth}, \mathbf{n}, \mathbf{h}, h) \rightarrow \exists m(Pre(\mathbf{n}, m) \wedge Lincoln-is-dead(m, h)))$

In STIT with Prior-Thomason's temporal operators, (Es6) would be a counterpart of the formula:

$$[Booth\ cstit : \mathbf{F}(Lincoln\ is\ dead)].$$

Between the moment when *Booth chooses to kill Lincoln* and the moment when *Lincoln is dead*, we have a temporal gap. And we still have the inadequate assumption in STIT that the action consisting of the sequence of events – Booth pulling the trigger, the bullet flying, the bullet entering Lincoln, Lincoln dying – is fully determined by Booth's choice.<sup>6</sup> This means that between the start of the action and the moment when its effect appears, the action cannot be stopped, neither for reasons internal to the agent (which in this case is impossible if we assume the pulling the trigger is instantaneous) nor for any external forces. The temporal gap is here both an agentive and a causal gap.

Solving a causal gap may require introducing a causality into STIT framework. In case of action that do not go beyond bodily movement we may want to talk about a causal relation between an intentional component and bodily movements and in case of actions which go beyond it, we may like to introduce also a causality between a bodily movements and some external events. Causality, even though it's important, can be taken as implicitly contained in the concept of action (similarly as we assumed for intentions).

**Ex post acto.** STIT's assumption that actions are always successful corresponds to the fact that actions are seen *ex post acto*. It is thus in some sense deliberate that only actions that have succeeded are taken into account, which is explicitly expressed by a CSTIT axiom:

$$[a\ cstit : \varphi] \rightarrow \varphi,$$

saying that if an agent *a* sees to it that  $\varphi$ , then  $\varphi$  must occur.

As we have seen, there are nevertheless good reasons to take a different point of view on actions. Indeed, this is why an extension to STIT has been proposed in [Mül05], to include the new operator 'is seeing to it that'. The *ex post acto* view solves the problem of the possible gap between the

---

<sup>6</sup>The irony is that STIT is designed for expressing indeterminacy, while it seems to force the actions with duration – so, the actions which underlies every STIT formula with temporal operators – to be fully determined.

choice and the action's outcome by simply assuming some kind of determinism of choice, and in [Mül05] it is solved by assuming the existence of default 'strategies'.

Obviously, since we just rigorously translated STIT semantics into a first-order theory, those issues on agentive and causal gap are just the drawback inherited from STIT and devised in Chapter 5. It seems obvious that we need to increase the expressivity in order to overcome this issue. We already sketched a solution to this in Section 5.4. The idea has been to extend the realm of agency to actions in order to obtain a logic with both benefits: powerful abstraction of agency and explicit reference to complex sustained actions. In this chapter, our move to FOL provides to us a versatility that was lacking in modal logic to express these complex mechanisms. Next section is devoted to the extension of our ontology of agency to an ontology of action.

## 7.4 Towards an Ontology of Action – OntoSTIT+

In this section we present the new theory OntoSTIT+, obtained by extending OntoSTIT with actions. We show that in OntoSTIT+ some of the problems just described are solved.

The intended models of OntoSTIT+ extend the domain of class  $\mathcal{M}$  in several ways. The expressivity of the language is extended accordingly, in particular by introducing corresponding new sorts. First, we need to refer to actions that possibly have a non-null duration. Since we are in a branching time framework, a single action can develop in different ways over different histories. We therefore need to distinguish between an 'action token', the single action the agent chooses to do at a given moment, and its 'action courses', which are the different possible ways this action token might unfold in time along different histories passing through the starting moment and belonging to the choice. In the light of a foundation ontology such as DOLCE [GGMO03], action courses would be the actual perdurants, whereas action tokens can be seen as more abstract entities, i.e., sets of action courses related by some sort of counterpart relation. Even though action tokens can be seen as sets of perdurants, they are not to be confused with action types. My cooking this egg now is an action token that may unfold differently over different histories if intervening events force the time to branch during the action. These different action courses may vary in their results, duration, etc. On the other hand, "cooking" is a type of action that can be true of many action tokens, possibly occurring at different times, with possibly different agents and different food items

involved. We therefore introduce both a set of action tokens and a set of action courses as individuals in the domain of quantification – and two new sorts in the language –, and action types as relations – and corresponding predicates in the language.

In addition, since actions may have a non-null duration, we introduce a set of intervals in the domain, and yet another sort in the language. Finally, as we want to describe the actions, e.g., saying that this action token is “a cooking by me of this egg”, we will also need to refer to other participants than the agent, in this case, “this egg”. Such entities may be of a wide range of sorts, although not temporal ones like moments, intervals or histories. We then extend the domain to also include a set of arbitrary non-temporal entities and a corresponding last sort in the language. Because many types of actions, e.g., “talking to”, have agents as participants, we need to assume that this last set of non-temporal entities includes the set *Agt* of agents. A richer taxonomy of sorts may be adopted, although we will not pursue this further here.

### 7.4.1 Language.

The language of OntoSTIT+ is that of OntoSTIT first extended with of all new sorts:

- individual variables of sort 3 ranging over intervals:  $i, i', \dots, i_1, \dots, i_n$
- individual constants of sort 3 denoting intervals:  $\mathbf{i}, \mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_n$
- individual variables of sort 4 ranging over action tokens:  $t, t', \dots, t_1, \dots, t_n$
- individual constants of sort 4 denoting action tokens:  $\mathbf{t}, \mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n$
- individual variables of sort 5 ranging over action courses:  $c, c', \dots, c_1, \dots, c_n$
- individual constants of sort 5 denoting action courses:  $\mathbf{c}, \mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$
- individual variables of sort 6 ranging over arbitrary non-temporal entities:  $x, x', \dots, x_1, \dots, x_n$
- individual constants of sort 6 denoting arbitrary non-temporal entities:  $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$

The language is then extended with new primitive universals. Let  $\Delta_+$  be the set of all explicitly introduced universal of OntoSTIT+,  $\Delta_+ = \Delta \cup \{InI, CO, RT, LO_n, AgO, Su\} \cup \Theta \cup \Omega\Theta$ .

The new predicate constants of the first additional set are understood as, respectively, “a moment is in an interval” of sort  $1 \times 3$ , “an action course is a course of an action token” of sort  $5 \times 4$ , “an action course runs through an interval” of sort  $5 \times 3$ , “an action course lies on a history” of sort  $5 \times 2$ , “an agent is the agent of an action token” of sort  $0 \times 4$ , and “an action course is successful” of sort 5.

$\Theta = \{A_1, A_2, \dots, A_k\}$  is a finite set of predicates of basic action types. Each  $A_i$  is of sort  $4 \times 6^{n_i}$ ,  $n_i \geq 0$ .

$\Omega\Theta = \{OA_1, OA_2, \dots, OA_k\}$  is a finite set of predicates describing the expected outcomes associated to an action type, in a one-to-one mapping with  $\Theta$ .  $OA_i$  is associated to  $A_i$  and it is of sort  $1 \times 2 \times 6^{n_i}$ . Assuming there is an expected-outcomes predicate associated to each action type predicate amounts to assuming actions are telic events, i.e., achievements and accomplishments in Vendler’s terminology [Ven67].

Finally, the set  $\Pi$  is extended to include predicate constants of sort  $1 \times 2 \times 6^n$ ,  $n \geq 0$ . So predicates  $P_i$  in  $\Pi$  do not longer correspond to atomic propositions anchored in a moment and history, but to predicates of any arity, increasing thus the expressivity of OntoSTIT+ regarding its ordinary vocabulary from that of a propositional language to that of a first-order language. This extension entails a change in the axioms involving the predicates  $P_i$ , namely (Ao6), which reads now:

$$(Ao6') \quad P_i(m, h, \vec{x}) \rightarrow In(m, h)$$

## 7.4.2 Characterization of new universals

**Intervals.** If this theory were embedded within a top-level ontology covering temporal concepts, some axioms and definitions would of course be inherited from that framework. But for clarity, we wish to specify all the theoretical elements required. We need here a standard notion of intervals. All intervals are linearly ordered (Aop1) and convex (Aop2).

$$(Aop1) \quad InI(m, i) \wedge InI(m', i) \rightarrow m = m' \vee Pre(m, m') \vee Pre(m', m)$$

$$(Aop2) \quad InI(m, i) \wedge InI(m'', i) \wedge Pre(m, m') \wedge Pre(m', m'') \rightarrow InI(m', i)$$

*Ine* is the inclusion between intervals, defined in terms of moments in the intervals (Dop1). Two intervals having the same moments are identical (Aop3).

**(Dop1)**  $Inc(i, i') \triangleq \forall m (InI(m, i) \rightarrow InI(m, i'))$

**(Aop3)**  $Inc(i, i') \wedge Inc(i', i) \rightarrow i = i'$

(Dop2) and (Dop3) define the beginning and end moments of intervals. Any interval has a beginning and an end (Aop4); the unicity of beginning and end for each interval (Top1) is guaranteed by (Dop2), (Dop3) and (Aop1). It is worth noting that nothing prevents a beginning of an interval from being equal to its end, so degenerated intervals are possible.

**(Dop2)**  $Beg(m, i) \triangleq InI(m, i) \wedge \forall m' (Pre(m', m) \rightarrow \neg InI(m', i))$

**(Dop3)**  $End(m, i) \triangleq InI(m, i) \wedge \forall m' (Pre(m, m') \rightarrow \neg InI(m', i))$

**(Aop4)**  $\forall i \exists m, m' (Beg(m, i) \wedge End(m', i))$

**(Top1)**  $\forall i \exists! m \exists! m' (Beg(m, i) \wedge End(m', i))$

(Dop4) defines the relation of temporal part between an interval and a history. For each interval there is a history of which it is temporal part (Aop5). However an interval may belong to more than one history, if those histories are undivided during that interval.

**(Dop4)**  $TP(i, h) \triangleq \forall m (End(m, i) \rightarrow In(m, h))$

**(Aop5)**  $\forall i \exists h TP(i, h)$

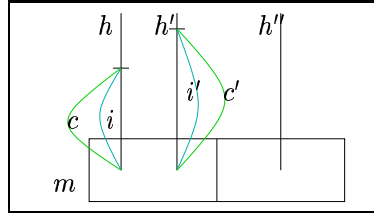
**Actions.** As for all perdurants, the time of each action course is fixed: there is exactly one interval such that it runs through it (Aop6). Action courses depend on action tokens: for each action course there is exactly one action token it is a course of (Aop7). Reciprocally, action tokens have to be realized: for each action token there is at least one action course which is a course of it (Aop8). The agent of each action token exists and is unique (Aop9).

**(Aop6)**  $\forall c \exists! i RT(c, i)$

**(Aop7)**  $\forall c \exists! t CO(c, t)$

**(Aop8)**  $\forall t \exists c CO(c, t)$

**(Aop9)**  $\forall t \exists! a AgO(a, t)$



**Figure 7.2:** Action course  $c$  (resp.  $c'$ ) lies-on history  $h$  ( $h'$ ). Action course  $c$  ( $c'$ ) runs-through interval  $i$  (resp.  $i'$ ).  $c$  and  $c'$  are two courses of a same token.

Figure 7.2 pictures the links between histories, intervals, tokens and courses of actions.

The next important property of action tokens is that each token corresponds to a single choice and a single occurrence. This is obtained by guaranteeing that all of its courses have the same starting point. For expressing this, we first need a few definitions. (Dop5) and (Dop6) define the predicates  $BAct(m, c)$  and  $EAct(m, c)$  which should be understood respectively as “moment  $m$  is the beginning of action course  $c$ ” and “moment  $m$  is the end of action course  $c$ ”. The existence and unicity of the beginning and end of each action course (Top2) is guaranteed by the existence and unicity of the interval of each action course (Aop6) and the existence and unicity of the beginning and end of each interval (Top1).

$$\text{(Dop5)} \quad BAct(m, c) \triangleq \exists i (RT(c, i) \wedge Beg(m, i))$$

$$\text{(Dop6)} \quad EAct(m, c) \triangleq \exists i (RT(c, i) \wedge End(m, i))$$

$$\text{(Top2)} \quad \forall c \exists! m \exists! m' (BAct(m, c) \wedge EAct(m', c))$$

(Aop10) guarantees that all action courses of the same action token have the same starting moment, so the beginning of an action token exists and is unique.

$$\text{(Aop10)} \quad CO(c, t) \wedge CO(c', t) \rightarrow \exists m (BAct(m, c) \wedge BAct(m, c'))$$

Degenerated courses may exist, for instance if the execution of the action is aborted immediately. Truly instantaneous actions are different. An action token is instantaneous if all its courses are, i.e., their beginning is identical to their end:

$$\text{(Dop7)} \quad Ins(t) \triangleq \forall c (CO(c, t) \rightarrow \exists m (BAct(m, c) \wedge EAct(m, c)))$$

An action course is lying on a history only if the interval it runs through is a temporal part of this history (Aop11).<sup>7</sup> The next important constraint is to guarantee that different action courses of the same token correspond to the realization of the token on different histories. In other words, for a given action token and history, only one action course is lying on this history (Aop12).

**(Aop11)**  $LOn(c, h) \wedge RT(c, i) \rightarrow TP(i, h)$

**(Aop12)**  $CO(c, t) \wedge CO(c', t) \wedge LOn(c, h) \wedge LOn(c', h) \rightarrow c = c'$

However, as the same interval can be a temporal part of many different histories it is not clear that all courses lying on different histories need to be distinguished. Two courses of a same token running through the same interval are different only if the histories they lie on are divided at the end (Top3). This avoids the proliferation of action courses without necessity but allows, in particular, an instantaneous action to have various courses, for instance a successful and a failing one. More generally, two courses of a same token running through two intervals, one which is included in the other, are different only if the histories they lie on are divided at the end of the shorter interval (Aop13). With this axiom, (Top3) is a theorem.

**(Aop13)**  $CO(c, t) \wedge CO(c', t) \wedge LOn(c, h) \wedge LOn(c', h') \wedge RT(c, i) \wedge RT(c', i') \wedge Inc(i, i') \wedge End(m, i) \wedge UD(h, h', m) \rightarrow c = c'$

**(Top3)**  $CO(c, t) \wedge CO(c', t) \wedge LOn(c, h) \wedge LOn(c', h') \wedge RT(c, i) \wedge RT(c', i) \wedge End(m, i) \wedge UD(h, h', m) \rightarrow c = c'$

We can define for commodity the predicate  $S(t, m, h)$  whose intended reading is “token  $t$  starts at moment  $m$  and history  $h$ ”:

**(Dop8)**  $S(t, m, h) \triangleq \exists c (CO(c, t) \wedge LOn(c, h) \wedge BAct(m, c))$

We can now characterize successful and unsuccessful actions. Success applies at the level of the course, not the token, as different ways the world turns out to be, i.e., different histories, may affect the realization of an action. An action course is successful if and only if the expected-outcomes predicate associated with the basic action type of its action token holds at the end:

<sup>7</sup> $LOn$  cannot be defined by a formula like  $\exists i (RT(c, i) \wedge TP(i, h))$  because we might need to distinguish courses lying on different histories but running through the same interval. More on this just below.

$$\text{(Aop14)} \quad \forall m, h, \vec{x} \quad (\exists c, t \quad (Su(c) \wedge Lon(c, h) \wedge EAct(m, c) \wedge CO(c, t) \wedge A_i(t, \vec{x})) \\ \leftrightarrow OA_i(m, h, \vec{x}))$$

(Aop14) is an axiom schema – and not a definition – as a new axiom is required for each basic type of action. Specific semantic constraints characterizing the action types through their associated expected outcomes will then be introduced as needed. For instance, one can assume that for the basic action type “*Switching-on*”  $\in \Theta$ , its associated expected-outcomes predicate “*OSwitching-on*”  $\in \Omega\Theta$  entails that the argument (e.g., a light) is “*On*”, *On* being a predicate in  $\Pi$ :

$$\text{(Asem1)} \quad OSwitching-on(m, h, x) \rightarrow On(m, h, x)$$

Such meaning postulates cannot be replaced by definitions of expected-outcomes predicates. We want to allow for actions of different types leading to the same consequences. But more importantly, we want to distinguish states like “being on” from states like “a switching-on has just been achieved”. So the expected outcomes described by the  $OA_i$  predicates hold only at the end moment of a successful action course (Aop14), but the specific effects (in this example, that the light is *On*) may of course persist. (Aop14) states that the effects of a successful action hold at its ending moment. For instantaneous actions, one can wonder whether it makes sense to have their effects holding at the very moment at which the action starts. Actually, one can wonder if there are such things as instantaneous actions at all... In this framework we didn’t want to take issue regarding the existence of instantaneous actions nor regarding the discrete or continuous nature of time. In the continuous case, there is no such thing as a “next moment” at which the effects would hold, therefore no other solution applying to both the discrete and continuous cases is available.

To correctly grasp notions such as continuing and aborting an action, we assume that a course cannot be successful if there is another course of the same token that prolongs it, in other words, an agent cannot continue an action which is already achieved (Aop15). On the other hand, we see here that an unsuccessful action course can be prolonged. In other words, there can be two courses of the same token running through two intervals such that one is included in the other. This is exactly the basis for the notion of aborting an action: an unsuccessful course is *aborted* if the agent can continue it on another history (Dop9), which divides with the history of the aborted course at its ending moment ((Top4) because of (Aop13)). The unsuccessful action course simply fails otherwise.

Note that (Aop15) does not imply that two successful courses of the same token action have the same duration. If the histories on which those

courses lie are divided before any of them end, then, we have no way of comparing the duration of the two courses.

$$\text{(Aop15)} \quad CO(c, t) \wedge RT(c, i) \wedge CO(c', t) \wedge RT(c', i') \wedge Inc(i, i') \wedge \neg Inc(i', i) \rightarrow \neg Su(c)$$

$$\text{(Dop9)} \quad Ab(c) \triangleq \neg Su(c) \wedge \exists t, c', i, i' (CO(c, t) \wedge RT(c, i) \wedge CO(c', t) \wedge RT(c', i') \wedge Inc(i, i') \wedge \neg Inc(i', i))$$

$$\text{(Top4)} \quad Ab(c) \wedge CO(c, t) \wedge RT(c, i) \wedge End(m, i) \wedge LOn(c, h) \wedge CO(c', t) \wedge RT(c', i') \wedge LOn(c', h') \wedge Inc(i, i') \rightarrow \neg UD(h, h', m)$$

All this makes sense if one and only one basic action type, modulo logical equivalence, applies to each action token (Aop16, Aop17).

$$\text{(Aop16)} \quad \forall t \exists \vec{x} (A_1(t, \vec{x}) \vee A_2(t, \vec{x}) \vee \dots \vee A_{|\Theta|}(t, \vec{x}))$$

$$\text{(Aop17)} \quad A_i(t, \vec{x}) \wedge A_j(t, \vec{x}) \rightarrow \forall t', \vec{x}' (A_i(t', \vec{x}') \leftrightarrow A_j(t', \vec{x}')) \wedge \forall m, h, \vec{x}' (OA_i(m, h, \vec{x}') \leftrightarrow OA_j(m, h, \vec{x}'))$$

This assumption corresponds in some sense to a multiplicative view on events, so that, for instance, a given action token cannot be both a “pressing the button” and a “switching on the light”. In fact, the first might be successful while the second is not. With a unifying view on events (motivated, e.g., by considering that bodily motion is essential to actions, see footnote 3), several distinct basic action types can be predicated of the same action token. In this case, one would need to drop (Aop16) and to substitute (Aop14) for an axiom schemata characterizing one “is a successful course of an action token of type  $A_i$ ” predicate – say,  $SuA_i$  – for each basic type  $A_i$ , augmenting the language accordingly. Of course, the multiplicative view assumed here doesn’t prevent the definition of more complex predicates (on the basis of the basic types in  $\Theta$  and possibly other predicates from  $\Pi$ ), that can multiply apply to the same action tokens.

Further characterization of action types can of course be done. We can distinguish achievement from accomplishment types by the fact that all their tokens are instantaneous or not. We can also introduce more specific constraints for some types such that no two tokens with the same participants can take place at intervals that overlap, or even that there cannot be two tokens with the same participants. For instance, no one can eat twice the same apple, and one can travel twice from Rome to Paris, but not at overlapping intervals. The extended literature on aspectuality can of course be valuable for this. Finally, we could associate preconditions, i.e., executability conditions, to actions types in addition to expected outcomes.

### 7.4.3 Agency in OntoSTIT+

We explained that the core of the theory of agents and choices is the independence of the agents. Independence lays at the organizational level of the agents and is captured by the constraints on the function *Choice*. In turn in OntoSTIT, this central aspect of the theory of agents and choice in branching time is brought by the predicate *PO*. Hence, it is important to do away with any misunderstanding of its interpretation. Indeed, we have argued in Section 7.3.3.1, that there is a gap between the notion of choice it borrows and its agentive character. We think that the expressivity of OntoSTIT+ makes it now possible to have a satisfying account of agency even in presence of actions with duration. The remaining of this section aims at binding the intuitions that are behind the notion of *Choice*, or rather *PO*, to the notion of action we just introduced in the OntoSTIT+ framework, and at bridging the agentive gap discussed earlier.

### 7.4.4 Understanding *PO*

Because OntoSTIT is the first-order equivalent of *BT + AC* models, we inherit some of its shallow specifications. In particular, in STIT, the agents share the set of moments and each agent faces a choice at every moment, i.e., in OntoSTIT, *PO* applies to all agents at all moments. If this is not problematic in STIT, where all actions underlying choices are assumed instantaneous, it will suggest an interesting question in OntoSTIT+ where we do have actions with a duration. How can we distinguish the following sentences? At moment  $m$  and history  $h$ , agent  $a$ :

1. starts a new action.
2. explicitly launches the continuation of an action.
3. lets an action go.
4. remains passive.

Some answers are more obvious than others. Starting an action is triggering a new action token independent of any previous token or course on the history  $h$ . The second case corresponds to a scenario where an agent triggers a new action in the meantime of performing another one, and it exists an explicit link between them. The third item refers to the behaviour of an agent that does not trigger any new action token. But one course of an action of its runs through an interval which is a temporal part of  $h$  and containing  $m$ . When an agent has no such activities, we say it remains

passive.<sup>8</sup> Accordingly, there will be some obvious similarities between the formal representations of scenarii 1 and 2, and between those of scenarii 3 and 4. The properties of an agent *passively continuing* an action or of *resting passive* will respectively correspond to the defined predicates  $PC(m, h, t)$  (with  $t$  a token agentive for  $a$ ) and  $RPass(a, m, h)$ .

What we propose here is to consider the particular actions of *continuing an action* as real actions, that is, actions underlying real choices. In [TV06], and because of the limitations of the language of modal logic, it was necessary to introduce explicitly what could be in the present setting a new action type  $A_i^c$ , reading “continuing  $A_i$ ” for every type  $A_i$ . Here however, we are just going to link those continuations via a predicate  $AC(t', t, m)$  which states that “the token  $t'$  is an *active continuation* of  $t$  at moment  $m$ ”. This way, we can refer to a concrete action that continues another happening action without introducing new predicates of action type. Possibly, the type of the continuation could be linked in further research by providing a taxonomy of types, but it is out of the scope of this dissertation. Note the need of a variable of moments in the arguments since a same token can have different continuations depending on the indeterminism or the instant of continuation. On  $AC(t', t, m)$ , we assume:

$$\text{(Aop18)} \quad AC(t', t, m) \rightarrow \forall c, i, m', m'' (CO(c, t) \wedge RT(c, i) \wedge InI(m, i) \wedge BAct(m'', c) \wedge EAct(m', c) \rightarrow \exists c' (CO(c', t') \wedge BAct(m, c') \wedge EAct(m', c') \wedge Pre(m'', m)))$$

$$\text{(Aop18')} \quad AC(t', t, m) \rightarrow \forall c', m' (CO(c', t') \wedge EAct(m', c') \rightarrow BAct(m, c') \wedge \exists c, m_b (CO(c, t) \wedge BAct(m_b, c) \wedge EAct(m', c) \wedge Pre(m_b, m)))$$

$$\text{(Aop19)} \quad AC(t', t, m) \wedge AC(t'', t, m) \rightarrow t' = t''$$

$$\text{(Aop20)} \quad AC(t', t, m) \wedge AgO(a, t) \rightarrow AgO(a, t')$$

(Aop19) ensures the unicity of a continuation of a token  $t$  at a moment  $m$ . (Aop20) constrains every continuation of a token  $t$  to be triggered only by the agent of  $t$ . (Aop18) and (Aop18') give the structure of a continuation whose courses must be each a temporal proper part of the relevant course of the token it continues, running up to the end.

Moreover, we can now make explicit how an action interacts with its continuations. The axiom schema (Aop21) establishes the link between the first two scenarii. It ensures that the expected outcomes of the type of the

---

<sup>8</sup>Remaining passive for human action is debatable, but may be convenient in modelling more abstract systems.

(continuing) token  $t'$  agree with those of the type of  $t$  at *relevant* moment and history.<sup>9</sup>

$$\text{(Aop21)} \quad AC(t', t, m) \wedge CO(c, t) \wedge EAct(m', c) \wedge LOn(c, h) \wedge In(m, h) \wedge A_i(t, \vec{x}_1) \wedge A_j(t', \vec{x}_2) \rightarrow (OA_i(m', h, \vec{x}_1) \leftrightarrow OA_j(m', h, \vec{x}_2))$$

As discussed above, it remains to define two predicates for stating that an agent at a moment  $m$  and history  $h$  “passively continues” a token  $t$  ( $PC$ , which is rather “token  $t$  is passively continued”) or simply rests passive ( $RPass$ ). These are indeed two distinct kinds of agency an agent could face at a same moment and we need to distinguish them for the sake of the meaning of the  $PO$  predicate.

$$\text{(Dop9)} \quad PC(m, h, t) \triangleq \exists c, i (CO(c, t) \wedge LOn(c, h) \wedge RT(c, i) \wedge InI(m, i) \wedge \neg BAct(m, c) \wedge \neg EAct(m, c)) \wedge \neg \exists t' AC(t', t, m)$$

$$\text{(Dop10)} \quad RPass(a, m, h) \triangleq \neg \exists c, t ((BAct(m, c) \wedge CO(c, t) \wedge LOn(c, h) \wedge AgO(a, t)) \wedge \neg \exists t' (AgO(a, t') \wedge PC(m, h, t'))$$

$PC$  is instrumental to the definition of  $RPass$ : an agent  $a$  remains passive at an index  $m/h$  if and only if it does not start any action (including those that are continuations) and does not passively continues any action. Note that the formula  $AC(t', t, m) \wedge PC(m, h, t)$  is consistent. In words, an agent can be both actively continuing a token and passively continuing the same token, at the same moment, although on different histories.

We now can reveal what is the  $PO$  predicate in the light of a language with explicit actions:

$$\text{(Dop11)} \quad PO(a, m, h, h') \triangleq \forall t (AgO(a, t) \rightarrow ((S(t, m, h) \leftrightarrow S(t, m, h')) \wedge (PC(m, h, t) \leftrightarrow PC(m, h', t)))) \wedge (RPass(a, m, h) \leftrightarrow RPass(a, m, h'))$$

With OntoSTIT+ we make the notion of choice explicit by identifying it to an equivalence class of agentive behaviours, be they starting some actions, continuing other ones or simply doing nothing. Formally, two histories are in the same choice partition of an agent at a given moment if and only if exactly the same actions are triggered and exactly the same actions are passively continued, or the agent remains passive at both indexes. This definition allows agents to start actions (which can be an active continuation of some other action) and passively continue others at the same index.

<sup>9</sup>Here we are minimalist by doing the assumption of the equivalence of expected outcomes *only* for relevant moment/history pairs. Alternatively we could have stated  $\forall m, h OA_i(m, h, \vec{x}) \leftrightarrow OA_j(m, h, \vec{x}')$ . But if later types become more specified it could appear too strong. Counter-examples are to be found in contextual purposes Better than freezing water?

Now that  $PO$  is defined, we can remove it from the set of universal primitives  $\Delta$ . It is easy to see that because of the material equivalences in the definition, it follows that  $PO$  will trivially satisfy the properties of an equivalence class. We thus also can do without the axioms (Ao10), (Ao11) and (Ao12).

Interestingly, we have the following theorem:

$$\text{(Top5)} \quad S(t, m, h) \wedge AgO(a, t) \rightarrow \forall h' (PO(a, m, h, h') \rightarrow S(t, m, h'))$$

It states that an action  $t$  of  $a$  is launched only if  $a$  has chosen so. As a consequence of (Top5) and the independence of agents' choices (Ao14), we also have:

$$\text{(Top6)} \quad \forall a, a', m, h, t (\neg(a = a') \wedge AgO(a', t) \wedge \forall h' (PO(a, m, h, h') \rightarrow S(t, m, h'))) \rightarrow \forall h'' (In(m, h'') \rightarrow S(t, m, h''))$$

Thus in  $\text{OntoSTIT}_+$ , an agent  $a$  sees to it that another agent  $a'$  sees to it that an action of  $a'$  is triggered *only if*  $a'$  has no alternative. In other words, an agent cannot force another agent to perform an action  $t$  except if  $t$  is inevitable. Of course, it does not rule out the possibility of an agent to have some kind of influence on another agent. It still can by a *prior* choice force the world to be at a moment where a given action of the other agent is inevitable.

### 7.4.5 Expressivity

We illustrate the expressiveness of these newly defined predicates and the  $\text{OntoSTIT}_+$  theory. We define a notion of an action being under control at a moment as in (Dop13). For this, we first need to define the property of 'happening' of a token at a given index.

$$\text{(Dop12)} \quad Happens(t, m, h) \triangleq \exists c, i (CO(c, t) \wedge RT(c, i) \wedge TP(i, h) \wedge InI(m, i) \wedge (BAct(m, c) \vee \neg EAct(m, c)))$$

$$\text{(Dop13)} \quad IsControlled(t, m) \triangleq \exists h, h' In(m, h) \wedge In(m, h') \wedge Happens(t, m, h) \wedge \neg Happens(t, m, h')$$

We say that a token  $t$  happens at an index  $m/h$  if there is a course  $c$  of  $t$  lying on an interval containing  $m$  and being a temporal part of  $h$ .  $c$  may end at  $m$  only if  $m$  is also its beginning. This way, an action does not happen at its last moment, except if this action is instantaneous.<sup>10</sup>

<sup>10</sup>Without such a constraint, an instantaneous action would happen nowhere!

We say a token is controlled at a moment  $m$  if it happens on a history of  $m$  and does not happen on another history of  $m$ . Note that as for (Top5), a token will happen only if its agent chooses so. Hence, a token is controlled at a moment only if its agent can choose between the action happening and the action ending. In other words, its agent *deliberately* sees to it that it happens.<sup>11</sup> Still in other words, the notion of control on the action is thus simply a control on its eventuality instead of a control on its outcome.

#### 7.4.5.1 Responsibility – filling the causal gap

We criticized in Chapter 5 and Section 7.3.3.1 the fact that the Chellas stit was not an operator of causality but rather of choice of which underlies an instantaneous action. Then the formula  $[a\text{ cstit} : \varphi]$  is not sufficient as a mark of causality of the agent  $a$  for a result  $\varphi$ . However, semantically, the notion of choice (and of alternative choice) is highly relevant. The achievement stit operator (see sections 2.4 and 5.3) is already in our view a very satisfying account of causality and consists in a complex truth condition in  $BT + AC$  models, especially capitalizing on the *Choice* function.

Moreover, in our setting and because of (Top5), an agent launches an action token only if it has chosen so. We were able to define the *PO* predicate via the triggering of a set of actions. It is then no surprise that we can grasp in our ontology of action a satisfying notion an *operator* of causality and responsibility.

**(Dop14)**  $Resp_a^{m,h}\varphi \triangleq \exists t, c, i, m', \vec{x} \text{ AgO}(a, t) \wedge CO(c, t) \wedge RT(c, i) \wedge Su(c) \wedge EAct(m, c) \wedge \bigvee_i (A_i(t, \vec{x}) \wedge (OA_i(m, h, \vec{x}) \rightarrow \varphi)) \wedge Pre(m', m) \wedge InI(m', i) \wedge IsControlled(t, m')$

By (Dop14), we say an agent  $a$  is *responsible* for a state of affairs  $\vec{x}$  iff there is an action token  $t$  whose agent is  $a$  and whose course  $c$  running through an interval  $i$  that successfully ends now with  $\vec{x}$  as an expected outcome, and there is at least at a moment  $m'$  prior to  $m$  on  $i$  where  $t$  has been controlled.

#### 7.4.5.2 Filling the agentive gap

As we explained in Section 7.3.3.1 actions themselves were not present in OntoSTIT, and we were not able to express that *the agent Robert switches off the light* by explicitly referring to a *switching* of a given *light*. In OntoSTIT+ we can now do it. (Esop1) represents again the example (Es1), in which

<sup>11</sup>'Deliberately' in the sense of the deliberative stit.

the action is assumed to be happening now, i.e. at the present moment  $\mathbf{n}$  on the actual history  $\mathbf{h}$ , and being instantaneous and successful:

$$\text{(Esop1)} \quad \exists t \quad (\text{Switching-off}(t, \mathbf{light}) \wedge \text{AgO}(\mathbf{Robert}, t) \wedge \exists c(\text{CO}(c, t) \wedge \text{LOn}(c, \mathbf{h}))) \wedge \text{Ins}(t) \wedge \forall c (\text{CO}(c, t) \rightarrow \text{BAct}(\mathbf{n}, c) \wedge \text{Su}(c))$$

This formula still reflects the underlying assumption of STIT and OntoSTIT that we could implicitly talk of actions through the assertion that their effects were guaranteed by a choice, and the companion assumption that all actions were successful. In OntoSTIT+, we may have successful and unsuccessful action courses of a given action token, even for instantaneous ones. A better rendering of this example might be to state that only the actual course of the action is successful:

$$\text{(Esop1')} \quad \exists t, c (\text{Switching-off}(t, \mathbf{light}) \wedge \text{AgO}(\mathbf{Robert}, t) \wedge \text{Ins}(t) \wedge \text{CO}(c, t) \wedge \text{BAct}(\mathbf{n}, c) \wedge \text{LOn}(c, \mathbf{h}) \wedge \text{Su}(c))$$

To complete the description, we must assume the following postulate that guarantees that a successful switching off the light ensures that the light is off (not on):

$$\text{(Asem2)} \quad \text{OSwitching-off}(m, h, x) \rightarrow \neg \text{On}(m, h, x)$$

Both (Esop1) and (Esop1') then entail with (Asem2) that the light is indeed off when Robert switches off the light:

$$\exists t, c (\text{Switching-off}(t, \mathbf{light}) \wedge \text{AgO}(\mathbf{Robert}, t) \wedge \text{CO}(c, t) \wedge \text{LOn}(c, \mathbf{h}) \wedge \text{EAct}(\mathbf{n}, c) \wedge \neg \text{On}(\mathbf{n}, \mathbf{h}, \mathbf{light}))$$

As for the formula expressing in addition the precondition of the switching off, (Es2) in Section 2.2.4, it is now:

$$\text{(Esop2)} \quad \forall x (\text{Pre}(x, \mathbf{n}) \rightarrow \exists y (\text{Pre}(x, y) \wedge \text{Pre}(y, \mathbf{n}) \wedge \text{On}(\mathbf{n}, \mathbf{h}, \mathbf{light}))) \wedge \exists t, c (\text{Switching-off}(t, \mathbf{light}) \wedge \text{AgO}(\mathbf{Robert}, t) \wedge \text{CO}(c, t) \wedge \text{LOn}(c, \mathbf{h}) \wedge \text{EAct}(\mathbf{n}, c) \wedge \neg \text{On}(\mathbf{n}, \mathbf{h}, \mathbf{light}))$$

Now, let us turn to the cases in which the agentive gap really showed up, namely, non-instantaneous actions. Resolving the causal gap amounts to requiring that the action has been successful only for the actual history  $\mathbf{h}$ , and leaving possible intervening events blocking the action on other histories. So, a faithful representation of example (Es6) is:

$$\text{(Esop6)} \quad \exists t, c (\text{Killing}(t, \mathbf{Lincoln}) \wedge \text{AgO}(\mathbf{Booth}, t) \wedge \text{CO}(c, t) \wedge \text{LOn}(c, \mathbf{h}) \wedge \text{BAct}(\mathbf{n}, c) \wedge \text{Su}(c))$$

**(Asem3)**  $OKilling(m, h, x) \rightarrow Dead(m, h, x)$

Notice that (Esop6) does not share the problems of (Es6) because the outcome of the action is linked to the action of the agent.

Now to solve the agentive gap, we might want to say that the agent may change its mind during the course of the action. This is not very different from the solution to the causal gap, but we can no longer be sure that the action is successful on the actual history.

**(Esop6')**  $\exists t, c, c' (Killing(t, \mathbf{Lincoln}) \wedge AgO(\mathbf{Booth}, t) \wedge CO(c, t) \wedge Lon(c, \mathbf{h}) \wedge BAct(\mathbf{n}, c) \wedge CO(c', t) \wedge Su(c'))$

By extending OntoSTIT on actions and intervals we solved two problems pointed out at the end of section 2.2.4.



---

## Conclusion and perspectives

### 8.1 Summary

We have tried to uncover ‘Devils-in-the-detail’ of logics of agents. The main purpose was to study how the logics relate to each other in order to augment our knowledge of the concepts they model. Our efforts to discover their similarities and differences indeed permitted us to establish some results.

*Chellas’s stit is an operator of brute choice.* If we think about operators of agency as abstractions of an underlying action that has been performed, Chellas’s stit corresponds to an action of choosing. There is no temporal aspect in it; The ending of an action is confounded with its starting. It is particularly striking that Chellas’s stit differs with Chellas’s  $\Delta_a\varphi$  exactly on a slight temporal glide (Section 5.3).

*Xu’s axiomatization of CSTIT can be simplified.* We can in fact replace the axiom schema for independence of agents with a syntactically simpler one. It suggests that (1) in presence of two agents, one can do without the operator of historical necessity (Section 3.4); (2) CSTIT can be given a much simpler semantics than  $BT + AC$  structures (Section 3.5); (3) we can link CSTIT with more standard modal logic and transfer results.

*Reasoning about brute choice of independent agents is complex.* There is a strong link between reasoning with S5 product logics and individual choice of independent agents. Taking up van Benthem and Pacuit’s slogan: “Grids are Dangerous” [vBP06]. However, and intriguingly, while one-agent CSTIT has the same complexity as S5, and two-agent CSTIT has the same complexity as  $S5^2$ , 3-agent CSTIT remains decidable and  $S5^3$  is not. But it is in any case sufficient to make it a complex logic, since in presence of at least two agents, CSTIT is NEXPTIME-complete (Theorem

3.5).

*Reasoning about coalitional choice is not more complex than reasoning about individual choice.* Chapter 4 extends the previous results on individual choice to coalitions. Finally it does not present any increase of complexity and remains NEXPTIME-complete with at least two agents. (See Theorem 4.6 and [BGH<sup>+</sup>07].)

*Coalition Logic can be evaluated w.r.t. Kripke semantics.* The non-normal operator of coalitional ability of CL can be simulated by the composition of three normal operators. (See Corollary 4.1.) Note that this is a characteristic that it shares with Pörn's logic of 'bringing it about'. (See our short presentation in Chapter 1. We shortly argued that using a non-normal operator was making it more likely to have interesting properties of agency.)

*Alternating-time Temporal Logic is consistent with philosophy of action.* Assuming discrete time and the grand coalition power to determine a unique outcome the logic of the strategic stit ability is more expressive than ATL. (Corollary 6.1.) ATL is a significant logic in computer science, but can then be justified as a relevant fragment of a more general logic in philosophy of action.

*It is helpful to interpret a brute choice as a choosing to perform a set of actions.* In the context of STIT, a choice at a moment  $w$  is simply a subset of histories passing through  $w$  that an agent choose (or can choose) to follow leaving apart some others. However, in a sufficiently rich framework, we can explain brute choice by the choice of triggering some set of actions. We can even reconstruct the relations of choice by collecting together histories along which exactly the same actions run. (See Sections 5.4.4 and 7.4.4.)

## 8.2 Towards rationality

Rationality has particularly been left aside. Brute choice has no component such as mental attitudes and we make use of mental aspects very seldom throughout this dissertation. We deal with epistemic notions in Section 4.7 just for the purpose of an application. The sole necessary use of mental attitude in our analysis of agency is done in our ontology of action via a predicate that associates an action to what its agent expects of it. It appeared to be essential to capture responsibility of an agent for a state

of affairs. Indeed, some notions of agency deserve to be studied beyond mere physical causality, as we did till now.

But why have we deliberately refrained from talking about rationality? The answer is simple but perhaps disappointing. Rationality is concerned with mental objects that trigger a behaviour of agents. We mention later epistemic notions that also play an obvious role in the picture. In this section, we are concerned with the incentive part of every action, namely the payoff function.

In the introduction, we said that we were going to abstract away from the payoff of game forms. The reason is that among the possible good candidates for their representation in logics would be *preference logics*. But there is no universally accepted approach. In [vB02], van Benthem argued that the simplest way was to add “a bunch of atomic propositions for value assertions” but described it as a bleak approach to the doing of intelligent agents. See also [vBL07] for a more advanced work.

In ATL-style frameworks, ongoing research is mostly influenced by the “Liverpudlian School”. In [WÅDvdH07], Wooldridge et al. review quickly the work done till now. But what we see as a Brobdingnagian challenge in importing preferences in logics of cooperation, is to understand the mechanisms of *preference aggregation* [ASS02] or more basically of *group preferences*. It is subject of much current research trying to merge preferences of individual agents to the level of coalitions. Researchers in preference aggregation are interested in extrapolating the preferences of a coalition of agents from those of its members, in such a way that it describes fairly the behaviour of the group.

Group preference is a challenge by itself, even if it does not involve precise notions of agency as we tried to make out. Integrating a satisfying account of group preference representation in a rich framework of agency could be the topic of another doctoral dissertation, or of a research of even longer term.

On the other hand, there is perhaps something worth studying concerning individual rationality. We have seen some intuitions in Chapter 1 about the links between normal form games and STIT moments. In addition, the formal framework of Chapter 4 introduces epistemic notions.

We thus see as an interesting path of research the issue of giving epistemic characterizations of equilibria in game theory. The problem is the following: Aumann and Brandenburger in [AB95] proved that some strong hypothesis were assumed in the definitions of solution concepts and in particular in Nash-equilibrium: in presence of many agents, mutual knowl-

edge of payoff function and of rationality plus common knowledge of eventual strategies of players are underlying assumptions of the definition of a Nash-equilibrium.

**Open problem.** Propose a formal framework based on STIT theory suitable for giving a pure logical proof of epistemic characterizations of solution concepts in game theory.

Boudewijn de Bruin in his doctoral dissertation also gave a very interesting analysis of the problem [dB04]. Thorsten Clausen for example did a similar work for characterizing Backward Induction in [Cla04].

Nevertheless, if such an issue deserves more work on some notions of rationality, it also involves a big piece of work in the analysis of temporal aspects in game forms.

### 8.3 Towards extensive games

We have seen the balance between time and agency in the last chapters of this dissertation. It appears that if our aim is to extend the concepts of agency to some ingredients of rationality, the nature of time is worth to be studied. Indeed, we rapidly feel the need to improve the ontology of time. We can illustrate this claim by briefly presenting a work in which we have identified difficulties w.r.t. the time-agency setting.

In [LTHC07] we have investigated *individual intentions* in a STIT-like semantics. Models are roughly  $BT + AC$  structures with one relation of belief and one relation of preference for every agent. What we consider to be the weakest point is actually the operator of agency that we use, that is Chellas's stit. We have already discussed in Section 5.2 a difficulty due to the lack of temporality brought by this operator.

Intention is not an end by itself: the purpose of [LTHC07] was to propose a logic of *delegation*. In particular, we identified as a crucial element of active delegation<sup>1</sup> was the ability of influencing another agent. *Influence* of an agent  $a$  on a distinct agent  $b$  for achieving  $\varphi$  should be naturally captured in the deliberative STIT theories by the formula  $[a\ cstit : [b\ cstit : \varphi]]$ . However, and because Chellas's stit has no temporal counterpart, there is no causal precedence of  $a$ 's underlying action over  $b$ 's one. Influence in this situation is incompatible with free-will of agents and with the STIT postulate of independence of agents. In fact,  $a$  can influence this way that

<sup>1</sup>It corresponds to *mild* delegation in [FC98].

$b$  ensures  $\varphi$  if and only if  $\varphi$  is inevitable. Admittedly, this is quite a singular notion of influence. Hence, we used a trick, and modeled the above influence by  $[a\ cstit : \mathbf{F}[b\ cstit : \varphi]]$ , that is by inserting an ‘artificial’ future statement to describe that  $a$ ’s action has some duration. We feel like an operator for action with duration would fit more adequately to the requirements than a brute choice operator as a mark of actual agency.

This problem is somewhat related to the one raised by Horty and that we briefly mentioned in our introduction to a strategic ability version of STIT in Section 2.3. It is the problem of treating agency in time (by means of a sequence of choices) and not only possible agency in time.

We can here use Thomas Müller’s argument in [Mül05]. He observes that STIT theory deals with Davidsonian sentences like “Jones buttered the toast”, but fails to grasp Anscombe-style sentences, e.g., “he is making tee”. The verbal *aspects* differ. STIT is suitable for modeling actions in the perfective aspect, or finished actions. However, it lacks some mechanisms in order to model actions in the imperfective aspect, or occurring actions. Müller says that “a full account of agency needs to consider Anscombe-type examples of continuous actions, too.” [Mül05, p. 195].

**Open problem.** Propose a logical framework for *actual* agency in time.

After having given in Chapter 4 an account of a logic that is close to normal form of games in game theory (and adequate to reason about uniform ‘one step’ strategies) a natural path for further investigation would be to adapt it to extensive forms of games. However, it is important to insist that we are interested in *actual* agency and not only in possible agency. This latter simply corresponds to the proposition of Alternating-time Temporal Logic that we have seen, does not support conveniently the addition of epistemic reasoning, and is likely to lack versatility in a research agenda in rationality.

The general remarks of Chapter 5 provided some understanding of the few temporal aspects captured by the logic of Chellas’s stit and tried to identify its situation in the literature of philosophy of action. A challenge for further research would then be to provide a logic of agency showing capabilities to capture complex sorts of interaction between agents and coalitions. We think that working on a logic of agency in time is worth considering, and Chapter 7 aimed at giving it a first specification.

More specifically, we think that *independence of agents* has been the central principle that allowed to capture the *causal structure of a moment* (or of a normal form game with abstract utilities). It constrains the needed structure of a moment for furnishing all relevant information about possible outcomes and how agents of the system can play in order to achieve them. Analogously, we regard the assumption of *no choice between undivided histories* as fundamental to capture the *causal structure of an extensive game*. It indeed forces that an agent or a group of agents can choose between one or another history only if those histories are divided.

For now, perhaps we have *stitted* enough. Still, we have to raise a puzzle to interested readers.

Lemma 3.1 page 26 concerns the validity of an alternative class of axiom schemas for independence of agents and called  $(AAIA_k)$ . Their theoremhood follows by Xu's completeness proof. Intriguingly, we have not been able to find a syntactic proof of it which revealed itself a riddle. Even the very simple instance  $(AAIA_1)$  is a challenge: can you find a derivation of the formula  $\diamond\varphi \rightarrow \langle 0 \rangle \langle 1 \rangle \varphi$  in Xu's axiomatic system described in Section 3.2.2?

---

# Bibliography

- [AB95] R. Aumann and A. Brandenburger, *Epistemic Conditions for Nash Equilibrium*, *Econometrica* **63** (1995), no. 5, 1161–1180.
- [AH99] R. Alur and T. A. Henzinger, *Reactive modules*, *Formal Methods in System Design: An International Journal* **15** (1999), no. 1, 7–48.
- [AHK97] R. Alur, T. A. Henzinger, and O. Kupferman, *Alternating-time temporal logic*, *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, IEEE Computer Society Press, 1997, pp. 100–109.
- [AHK99] ———, *Alternating-time temporal logic*, *Compositionality: The Significant Difference*, *Lecture Notes in Computer Science* 1536, Springer, 1999, pp. 23–60.
- [AHK02] ———, *Alternating-time temporal logic*, *Journal of the ACM* **49** (2002), 672–713.
- [AHKV98] R. Alur, T. Henzinger, O. Kupferman, and M. Vardi, *Alternating refinement relations*, *International Conference on Concurrency Theory*, 1998, pp. 163–178.
- [AHM<sup>+</sup>98] R. Alur, T. A. Henzinger, F. Y. C. Mang, S. Qadeer, S. K. Rajamani, and S. Tasiran, *MOCHA: Modularity in model checking*, *Computer Aided Verification*, 1998, pp. 521–525.
- [Ans57] E. Anscombe, *Intention*, Cornell University Press, Ithaca, NY, 1957.
- [ASS02] K. J. Arrow, A. K. Sen, and K. Suzumura (eds.), *Handbook of social choice and welfare*, *Handbook of Social Choice and Welfare*, vol. 1, Elsevier, June 2002.
- [BdRV01] P. Blackburn, M. de Rijke, and Y. Venema, *Modal logic*, Cambridge University Press, 2001.

- [Bel91] N. Belnap, *Backwards and Forwards in the Modal Logic of Agency*, *Philosophy and Phenomenological Research* **LI** (1991), no. 4, 777–807.
- [BG01] P. Blackburn and V. Goranko, *Hybrid Ockhamist Temporal Logic*, Proceedings of the 8th Int. Symp. on Temporal Representation and Reasoning (TIME-01) (Bettini, C. and Montanari, A, ed.), IEEE Computer Society Press, 2001, pp. 183–188.
- [BGH<sup>+</sup>07] P. Balbiani, O. Gasquet, A. Herzig, F. Schwarzenruber, and N. Troquard, *Coalition games over Kripke semantics*, Festschrift in Honour of Shahid Rahman (C. Dégrémont, L. Keiff, and H. Rückert, eds.), College Publications, 2007.
- [BHT06a] J. Broersen, A. Herzig, and N. Troquard, *A STIT-extension of ATL*, Tenth European Conference on Logics in Artificial Intelligence (JELIA'06), Liverpool, England, UK, Lecture Notes in Artificial Intelligence, vol. 4160, Springer, 2006, pp. 69–81.
- [BHT06b] ———, *Embedding Alternating-time Temporal Logic in strategic STIT logic of agency*, *Journal of Logic and Computation* **16** (2006), no. 5, 559–578.
- [BHT06c] ———, *From Coalition Logic to STIT*, Third International Workshop on Logic and Communication in Multi-Agent Systems (LCMAS 2005), Edinburgh, Scotland, UK (W. van der Hoek, A. Lomuscio, E. de Vink, and M. Wooldridge, eds.), *Electronic Notes in Theoretical Computer Science*, vol. 157:4, Elsevier, 2006, pp. 23–35.
- [BHT07a] P. Balbiani, A. Herzig, and N. Troquard, *Alternative axiomatics and complexity of deliberative STIT theories*, 2007, arXiv:0704.3238v1. Submitted.
- [BHT07b] J. Broersen, A. Herzig, and N. Troquard, *Normal simulation of coalition logic and an epistemic extension*, Proceedings of TARK 2007 (Brussels, Belgium), ACM DL, 2007.
- [BM04] A. Baltag and L. S. Moss, *Logics for epistemic programs*, *Synthese* **139** (2004), 165–224.

- [BP88] N. Belnap and M. Perloff, *Seeing to it that: a canonical form for agentives*, *Theoria* **54** (1988), 175–199.
- [BPX01] N. Belnap, M. Perloff, and M. Xu, *Facing the future: agents and choices in our indeterminist world*, Oxford, 2001.
- [Bra87] M. E. Bratman, *Intention, plans, and practical reason*, Harvard University Press, Cambridge, MA, 1987.
- [Bro88] M. A. Brown, *On the logic of ability*, *Journal of Philosophical Logic* **17** (1988).
- [Cas03] C. Castelfranchi, *The micro-macro constitution of power*, *Protosociology*, no. 18-19, 2003.
- [CE01] E. M. Clarke and E. A. Emerson, *Synthesis of synchronization skeletons for branching time temporal logic*, *Lecture Notes in Computer Science*, Springer Verlag, 2001, pp. 52–71.
- [Che69] B. F. Chellas, *The Logical Form of Imperatives*, Ph.D. thesis, Philosophy Department, Stanford University, 1969.
- [Che92] ———, *Time and modality in the logic of agency.*, *Studia Logica* **51** (1992), no. 3/4, 485–518.
- [Cla04] T. Clausing, *Doxastic Conditions for Backward Induction*, *Theory and Decision* **54** (2004), no. 4, 315–336.
- [CP01] J. Carmo and O. Pacheco, *Deontic and action logics for organized collective agency, modeled through institutionalized agents and roles.*, *Fundamenta Informaticae* **48** (2001), no. 2-3, 129–163.
- [CV96] R. Casati and A. Varzi (eds.), *Events*, Dartmouth Publishing, Aldershot, 1996.
- [Dav91] D. Davidson, *Essays on actions and events*, Clarendon Press, Oxford, 1991.
- [dB04] B. de Bruin, *Explaining Games: On the Logic of Game Theoretic Explanations*, Ph.D. thesis, University of Amsterdam, 2004.
- [Dég06] C. Dégremont, *Dialogical Deliberative Stit*, Master's thesis, University of Lille 3, 2006.

- [Dou76] W. Douglas, *Logical form and agency*, *Philosophical Studies* **29** (1976), 75–89.
- [Elg93] D. Elgesem, *Action theory and modal logic*, Ph.D. thesis, Department of philosophy, University of Oslo, 1993.
- [Elg97] ———, *The modal logic of agency*, *Nordic Journal of Philosophical Logic* **2** (1997), no. 2, 1–46.
- [FC98] R. Falcone and C. Castelfranchi, *Towards a theory of delegation for agent-based systems*, *Robotics and Autonomous Systems* **24** (1998), 141–157.
- [GB96] R. P. Goldman and M. S. Boddy, *Expressive planning and explicit knowledge*, *Proceedings of the 3rd International Conference on Artificial Intelligence Planning Systems (AIPS-96)*, AAAI press, 1996, pp. 110–117.
- [GG95] N. Guarino and P. Giaretta, *Ontologies and knowledge bases: Towards a terminology clarification*, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing* (N. Mars, ed.), IOS Press, 1995, pp. 25–32.
- [GGMO03] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari, *Sweetening wordnet with dolce*, *AI Magazine* **24** (2003), no. 3, 13–24.
- [GJ04] V. Goranko and W. Jamroga, *Comparing semantics of logics for multi-agent systems*, *Synthese* **139** (2004), no. 2, 241–280.
- [GKWZ03] D. M. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyashev, *Many-dimensional modal logics: Theory and applications*, *Studies in Logic and the Foundations of Mathematics*, no. 148, Elsevier, North-Holland, 2003.
- [Gol70] A. Goldman, *A theory of human action*, Prentice Hall, Englewood Cliffs, N.J., 1970.
- [Gor01] V. Goranko, *Coalition games and alternating temporal logics*, *TARK '01: Proceedings of the 8th conference on Theoretical aspects of rationality and knowledge* (San Francisco, CA, USA), Morgan Kaufmann Publishers Inc., 2001, pp. 259–272.

- [GvD06] V. Goranko and G. van Drimmelen, *Decidability and complete axiomatization of the alternating-time temporal logic*, *Theoretical Computer Science* **353** (2006), no. 1-3, 93–117.
- [Hen67] P. D. Henry, *The logic of st. anselm*, Oxford University Press, 1967.
- [Hor79] T. Horgan, *Action theory and social science: some formal models*. By Ingmar Pörn., *The Philosophical Reviews* **88** (1979), no. 2, 308–311.
- [Hor01] J. F. Horty, *Agency and deontic logic*, Oxford University Press, Oxford, 2001.
- [HT06] A. Herzig and N. Troquard, *Knowing How to Play: Uniform Choices in Logics of Agency*, 5th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-06), Hakodate, Japan (G. Weiss and P. Stone, eds.), ACM Press, 2006, pp. 209–216.
- [JÅ06] W. Jamroga and T. Ågotnes, *Constructive knowledge: What agents can achieve under incomplete information*, 5th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-06), Hakodate, Japan (G. Weiss and P. Stone, eds.), ACM Press, 2006, pp. 232–234.
- [JD05] W. Jamroga and J. Dix, *Do agents make model checking explode (computationally)?*, *CEEMAS*, 2005, pp. 398–407.
- [JS93] A. Jones and M. Sergot, *On the characterization of law and computer systems: The normative systems perspective*, *Deontic Logic in Computer Science: Normative System Specification* (J.-J. C. Meyer and R. J. Wieringa, eds.), Wiley, New York, 1993, pp. 275–307.
- [JS96] ———, *A formal characterization of institutionalized power*, *Journal of the IGPL* **4** (1996), no. 3, 429–445.
- [JvdH04] A. Jamroga and W. van der Hoek, *Agents that know how to play*, *Fundamenta Informaticae* (2004).
- [Kan72] S. Kanger, *Law and logic*, *Theoria* **38** (1972).

- [KS92] H. Kautz and B. Selman, *Planning as satisfiability*, Proceedings 10th European Conference on AI, Wiley, 1992, pp. 359–363.
- [LTHC07] E. Lorini, N. Troquard, A. Herzig, and C. Castelfranchi, *Delegation and mental states*, 6th International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS-07), Honolulu, Hawaii, USA (Edmund H. Durfee and Makoto Yokoo, eds.), ACM Press, May 2007.
- [Lut04] C. Lutz, *An improved nexttime-hardness result for description logic alc extended with inverse roles, nominals, and counting*, Tech. report, University of Dresden, 2004.
- [LWWW06] C. Lutz, D. Walther, F. Wolter, and M. Wooldridge, *ATL is indeed EXPTIME-complete*, Journal of Logic and Computation **16** (2006), no. 6, 765–787.
- [Man96] M. Manzano, *Extensions of first order logic*, Cambridge University Press, 1996.
- [Mar99] M. Marx, *Complexity of products of modal logics*, Journal of Logic and Computation **9** (1999), no. 2, 221–238.
- [Mel96] A. Mele, *Springs of action: Understanding intentional behavior*, Oxford University Press, New York, 1996.
- [MM01] M. Marx and S. Mikulas, *Products, or how to create modal logics of high complexity*, Logic Journal of the IGPL **9** (2001), 77–88.
- [Mül05] T. Müller, *On the formal structure of continuous action*, Advances in Modal Logic (5), King’s College Publications, 2005, pp. 191–209.
- [MvN44] O. Morgenstern and J. von Neumann, *Theory of games and economic behavior*, Princeton University Press, 1944.
- [Ohl98] H. J. Ohlbach, *Combining hilbert style and semantic reasoning in a resolution framework*, CADE-15, Lecture Notes in Artificial Intelligence, vol. 1421, 1998, pp. 205–219.
- [OR94] M. J. Osborne and A. Rubinstein, *A course in game theory*, The MIT Press, 1994.

- [Par02] R. Parikh, *Social software*, Synthese **132** (2002), no. 3, 187–211.
- [Pau01] M. Pauly, *Logic for social software*, Ph.D. thesis, University of Amsterdam, 2001, ILLC Dissertation Series 2001-10.
- [Pau02] ———, *A modal logic for coalitional power in games.*, Journal of Logic and Computation **12** (2002), no. 1, 149–166.
- [Pie00] P. M. Pietroski, *Causing actions*, Oxford University Press, 2000.
- [Pör70] I. Pörn, *The logic of power*, Blackwell, Oxford, 1970.
- [Pör77] ———, *Action theory and social science: some formal models*, D. Reidel, Dordrecht, 1977.
- [Pra76] V. R. Pratt, *Semantical Considerations on Floyd-Hoare Logic*, Proc. 17th Annual IEEE Symposium on Foundations of Computer Science, 1976, pp. 109–121.
- [Pri67] A. N. Prior, *Past, present, and future*, Clarendon Press, 1967.
- [Roy00] L. Royakkers, *Combining deontic and action logics for collective agency*, Legal Knowledge and Information Systems (Jurix 2000) (J. Breuker and R. Leenes and R. Winkels, ed.), IOS Press, 2000.
- [Sch94] A. Schaerf, *Reasoning with individuals in concept languages*, Data and Knowledge Engineering **13** (1994), no. 2, 141–176.
- [Sch04] P. Y. Schobbens, *Alternating-time logic with imperfect recall*, Electronic Notes in Theoretical Computer Science **85** (2004), no. 2.
- [Sea01] J. Searle, *Rationality in action*, MIT Press, Cambridge, MA, 2001.
- [Seg92] K. Segerberg, *Getting started: Beginnings in the logic of action.*, Studia Logica **51** (1992), no. 3/4, 347–378.
- [Swa07] F. Swarzentruher, *Décidabilité et complexité de la logique normale des coalitions*, Master’s thesis, Université Toulouse 3, 2007.

- [Tho70] R. H. Thomason, *Indeterminist time and truth-value gaps*, *Theoria* **36** (1970), 264–81.
- [Tho77] J. J. Thomson, *Acts and other events*, Cornell University Press, Ithaca, N.Y., 1977.
- [Tho84] R. H. Thomason, *Combinations of tense and modality*, *Handbook of Philosophical Logic: Extensions of Classical Logic* (D. M. Gabbay and F. Guentner, eds.), Reidel, 1984, pp. 135–165.
- [Tob01] S. Tobies, *Complexity results and practical algorithms for logics in knowledge representation*, Ph.D. thesis, LuFG Theoretical Computer Science, Aachen, Germany, 2001.
- [Tro07] N. Troquard, *Some clarifications in logics of agency*, *Proceedings of ESSLLI'07 Student Session* (Dublin, Ireland), 2007.
- [TTV06] N. Troquard, R. Trypuz, and L. Vieu, *Towards an ontology of agency and action: From STIT to OntoSTIT+*, *International Conference on Formal Ontology in Information Systems*, Baltimore, Maryland, USA (Brandon Bennett and Christiane Felbaum, eds.), IOS Press, 2006, pp. 179–190.
- [TV06] N. Troquard and L. Vieu, *Towards a logic of agency and actions with duration*, *European Conference on Artificial Intelligence 2006 (ECAI'06)*, Riva del Garda, Italy, IOS Press, 2006, short paper, pp. 775–776.
- [TV07] R. Trypuz and L. Vieu, *Building an Ontology of Agents and Choices in Branching Time*, Submitted, 2007.
- [vB84] J. van Benthem, *Correspondence theory*, *Handbook of Philosophical Logic*, vol. II (D. M. Gabbay and F. Guentner, eds.), reidel, 1984.
- [vB02] ———, *Extensive games as process models*, *Journal of Logic, Language and Information* **11** (2002), no. 3, 289–313.
- [vB06] ———, *Open Problems in Logic Dynamics*, *Mathematical Problems from Applied Logic I* (D. M. Gabbay, S. S. Goncharov, and M. Zakharyashev, eds.), vol. 4, Springer New York, 2006.

- [vBL07] J. van Benthem and F. Liu, *Dynamic Logic of Preference Upgrade*, Journal of Applied Non-Classical Logic (2007), forthcoming.
- [vBP06] J. van Benthem and E. Pacuit, *The Tree of Knowledge in Action: Towards a Common Perspective*, Advances in Modal Logic (I. Hodkinson G. Governatori and Y. Venema, eds.), vol. 6, College Publications, 2006.
- [vD03] G. van Drimmelen, *Satisfiability in alternating-time temporal logic*, Proceedings of the Eighteenth Annual IEEE Symp. on Logic in Computer Science, LICS 2003 (P. G. Kolaitis, ed.), IEEE Computer Society Press, June 2003, pp. 208–217.
- [vdHLW06] W. van der Hoek, A. Lomuscio, and M. Wooldridge, *On the complexity of practical atl model checking*, AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems (New York, NY, USA), ACM Press, 2006, pp. 201–208.
- [vdHW02] W. van der Hoek and M. Wooldridge, *Tractable multiagent planning for epistemic goals*, AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems (New York, NY, USA), ACM Press, 2002, pp. 1167–1174.
- [vdHW05] ———, *On the dynamics of delegation, cooperation, and control: A logical account*, Proc. of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'05), 2005.
- [Ven67] Z. Vendler, *Verbs and times*, Philosophical Review **56** (1967), 143–160.
- [vK86] F. von Kutschera, *Bewirken*, Erkenntnis **24** (1986), no. 3, 253–281.
- [vK97] ———,  *$T \times W$ -Completeness*, Journal of Philosophical Logic **26** (1997), 241–250.
- [WÅDvdH07] M. Wooldridge, T. Ågotnes, P. E. Dunne, and W. van der Hoek, *Logic for Automated Mechanism Design – A Progress Report*, Proceedings of AAI 2007, 2007.

- [Wan06] H. Wansing, *Tableaux for multi-agent deliberative-stit logic*, *Advances in Modal Logic*, Volume 6 (G. Governatori, I. Hodkinson, and Y. Venema, eds.), King's College Publications, 2006, pp. 503–520.
- [Wöl04] S. Wölfl, *Qualitative action theory: A comparison of the semantics of alternating time temporal logic and the Kutschera-Belnap approach to agency*, *Proceedings Ninth European Conference on Logics in Artificial Intelligence (JELIA'04)* (J. Alferes and J. Leite, eds.), *Lecture Notes in Artificial Intelligence*, vol. 3229, Springer, 2004, pp. 70–81.
- [Woo00] M. Wooldridge, *Reasoning about rational agents*, The MIT Press, 2000.
- [Woo02] ———, *An introduction to multiagent systems*, John Wiley & Sons, 2002.