



**HAL**  
open science

# Lightweight approaches to reasoning about knowledge and belief

Elise Perrotin

► **To cite this version:**

Elise Perrotin. Lightweight approaches to reasoning about knowledge and belief. Library and information sciences. Université Paul Sabatier - Toulouse III, 2021. English. NNT : 2021TOU30265 . tel-03683642

**HAL Id: tel-03683642**

**<https://theses.hal.science/tel-03683642>**

Submitted on 31 May 2022

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

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



# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**  
Délivré par l'Université Toulouse 3 - Paul Sabatier

---

Présentée et soutenue par

**Elise PERROTIN**

Le 11 octobre 2021

**Approches légères pour le raisonnement sur les connaissances et les croyances**

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Informatique et Télécommunications**

Unité de recherche :  
**IRIT : Institut de Recherche en Informatique de Toulouse**

Thèse dirigée par  
**Andreas HERZIG et Emiliano LORINI**

Jury

M. Alexandru BALTAG, Rapporteur  
Mme Sonja SMETS, Examinatrice  
M. Frédéric MARIS, Examineur  
M. Andreas HERZIG, Directeur de thèse  
M. Emiliano LORINI, Co-directeur de thèse  
M. Paul EGRÉ, Président

Lightweight approaches to reasoning about  
knowledge and belief

Elise Perrotin

## Acknowledgements

The road leading up to this thesis has been cleared for me by so many people and institutions it would simply be a lie to present all of this as resulting solely of my own hard work.

It is difficult to know where to start, but perhaps I should go for the Ecole Normale Supérieure de Lyon, which has introduced me to logic, made it possible and comfortable for me to spend a year in JAIST in the Tojo lab where I first discovered that logic could be applied to reasoning about actual people's minds, and then allowed me to further explore that interest with the LMFI master in Paris and internships in Sapporo with Sano-sensei, in Amsterdam with Sonja Smets and Fernando Velázquez-Quesada, and in Nottingham with Natasha Alechina and Rustam Galimullin. The ENSL has opened so many doors for me as well as funding all these years of studying and internships *and* my PhD, and I am so incredibly grateful that such a school exists, and that I was able to be a student there and enjoy all of these benefits. And for *this* I must also thank all my professors at the Lycée Pierre de Fermat, who helped me succeed at the entrance exam to the ENSL.

I must also of course thank my PhD supervisors, Andreas Herzig, Emiliano Lorini and Frédéric Maris, who have always been kind and encouraging, always ready to help me figure things out and include me into projects and praise the work that I did. You obviously can't work three years with someone without there being at least a few times of frustration, but and I am also grateful for those, since if my supervisors, who are recognized researchers in my field, can be human and make mistakes, that means it must also be okay for *me* to be human and make mistakes, no matter what any impostor syndrome might try to say. Thank you also to the rest of my defense jury, Paul, Alexandru and Sonja, who have borne good-heartedly with all the administrative nonsense and have been very kind and positive about my work while still challenging the rougher edges of it.

My supervisors and jury are not the only fellow researchers I am indebted to. I have mentioned those from my previous years of studying and internships, and I must now mention all the other doctoral students that have become my friends, Pierre-François who was so very kind from the very beginning, Julien who listened and helped me through all my ranting at various work- and administration-related things and helped me look at my work through different perspectives, Saúl, always enthusiastic and unquestioningly supportive, Cédric who spent afternoons with me reorganizing and cleaning our workspace to make it more welcoming, Maryam, Victor, Mickaël, Xinghan, and probably more that I am now forgetting. I must also thank my other coauthors and co-teachers and the rest of the members of the AI department who always made me feel welcome and happy to work in such a friendly environment: Martin, Florence, Dominique, Olivier, Umberto, etc. etc. Thank you also to everyone I was lucky enough to meet and chat with at the MAFTEC workshops and at the conferences I got to attend, and to Sébastien Konieczny and the team at Lens for being so eager to welcome me for a postdoc and help me settle things on that front so that I had some peace of mind to focus on the end of my PhD time.

No work can be done if the downtime isn't equally fulfilling, and though it may seem like I spend a lot of time holed up in my own home, I am never

truly alone so long as I can reach out to my friends. Thank you to Quentin and Céline for never giving up on inviting me out despite how many times I didn't feel up to it, to Sawasaki for making online conferences so much more fun and somehow always being awake and ready to chat, to Hoshio and Nomoto for being the kindest friends I could have, to Dina Norlund and her community for always inspiring me and motivating me to keep making art, to Cantin, to Florette, to Valentin, who always knows just what to say to make me feel better and push me to expand my comfort zone little by little.

The last of the acknowledgements go to my family, my sister Mylène who came down to Toulouse just for my defense despite having no interest whatsoever in my field of work, my dad who is always so proud of me, my mom who took care of me for these three years just like all the years before and took on way more than her share of chores just so I could be comfortable and focus on my work. I would certainly not have arrived here so easily without everything you've all done for me my whole life.

## Résumé

Il existe de nombreuses manières de représenter les connaissances et les croyances ainsi que leur évolution dans des systèmes multi-agents. Parmi celles-ci, la Logique Epistémique Dynamique (DEL) a tendance à prendre le devant de la scène. Dans DEL, les modèles sont constitués de mondes possibles ainsi que de relations entre ces mondes indiquant lesquels chaque agent sait distinguer ou non. Mais la sémantique concise et la grande expressivité de DEL ont un prix : en particulier, le problème de l'existence d'une solution à un problème de planification en DEL est indécidable dès que plus de deux agents sont en jeu. Plus précisément, étant donné une situation initiale, un ensemble d'actions autorisées et un but, il n'est pas possible en général de décider s'il existe ou non une séquence d'actions menant de l'état initial à un état dans lequel le but est satisfait.

Il convient alors de se demander si l'on peut trouver des approches plus simples, qui donnent de meilleurs résultats en termes de complexité tout en restant suffisamment expressives pour modéliser des situations réelles intéressantes. Dans cette thèse, nous développons et étudions certaines de ces logiques restreintes de connaissances et de croyances ainsi que leur application à des problèmes de planification épistémique. Nous ne sommes pas les premiers à nous intéresser à de telles simplifications de DEL: typiquement, les autres approches existantes restreignent les types d'actions qui peuvent être exécutés à des actions n'augmentant pas (ou peu) l'incertitude, ou limitent les capacités cognitives des agents de sorte que la connaissance commune (tout le monde sait que tout le monde sait que tout le monde sait que... à l'infini) ne soit pas considérée, les agents ne puissent pas raisonner sur les connaissances des autres agents, ou bien les agents ne puissent pas savoir que d'autres agents connaissent des informations qu'ils ne connaissent pas eux-mêmes.

L'approche principale développée dans cette thèse, la Logique Epistémique d'Observation (EL-O), repose sur le concept de "savoir si", ou "voir si", plutôt que le plus traditionnel "savoir que". Cela nous permet d'avoir un niveau d'expressivité plus élevé que d'autres simplifications existantes de DEL. Dans cette thèse nous développons EL-O et étudions un certain nombre de ses propriétés et variants possibles. En particulier, nous montrons que le problème de satisfiabilité est NP-complet, et que EL-O est un fragment de la logique standard S5. Nous définissons ensuite une sémantique pour les actions et problèmes de planification EL-O, et montrons que le problème de l'existence d'un plan pour un problème donné est non seulement décidable, mais aussi PSPACE-complet, ce qui est la même classe de complexité qu'en planification classique (c'est-à-dire non-épistémique). Ce résultat vaut à la fois pour la planification séquentielle, où une action est effectuée à la fois, et parallèle, où plusieurs actions peuvent être effectuées en même temps. Nous étudions enfin des pistes pour adapter notre approche à une logique de croyances, et montrons que toutes ces approches ainsi que d'autres de la littérature peuvent être traduites dans le cadre plus général de la Logique des Attitudes Doxastiques de Lorini.

## Abstract

There are many ways to represent knowledge and belief and the evolution thereof in multiagent systems. Among those, one framework that has taken the forefront of the scene is Dynamic Epistemic Logic (DEL), in which models consist of possible worlds that are valuations over propositional variables and relations between those worlds which indicate which worlds agents can or cannot tell apart. The concise semantics and high expressive power of DEL come at a cost: in particular, planning in DEL is undecidable as soon as more than two agents are involved. That is, given an initial situation, a set of allowed actions, and a goal, it is not possible in general to decide whether or not there exists a series of actions which will lead one from the initial situation to a situation in which the goal is achieved.

The question is then whether one can find some simpler approaches which give better complexity results while remaining expressive enough to model a variety of real-life situations. In this thesis we develop and study some such restricted logics of knowledge and belief, and consider their applicability in particular when it comes to epistemic planning. We are not the first to do this: other approaches exist which typically either heavily restrict the types of actions that can be executed, limit the agents' reasoning power, or both. For agents, uncertainty on actions might not be allowed, common knowledge (everybody knowing that everybody knows that everybody knows that... up to infinity) might not be considered, or agents might not be allowed to reason about other agents' knowledge or know that another agent knows something that they don't know themselves.

The approach presented in this thesis is rooted in logics of observation and relies on 'knowing whether', or 'seeing', as a basic concept, rather than the more traditional 'knowing that'. This allows our logic to be more expressive than other existing restricted logics which are based on 'knowing that'. We refine the already existing Epistemic Logic of Observation (EL-O), discuss a number of its properties (including NP-completeness of the EL-O satisfiability problem and the fact that EL-O is a fragment of standard S5 logic) and possible variations, and show that it naturally leads to a framework for epistemic planning which is not only decidable, but PSPACE complete. We then discuss how this approach could be extended to reason about beliefs, and finally compare this approach to other symbolic epistemic and doxastic logics through the more general framework of Lorini's Logic of Doxastic Attitudes.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Background</b>	<b>11</b>
2.1	Standard classical and epistemic logics . . . . .	12
2.2	Shared and common knowledge . . . . .	16
2.3	Classical and DEL planning . . . . .	23
<b>3</b>	<b>EL-O: Epistemic Logic of Observation</b>	<b>30</b>
3.1	Syntax and semantics of EL-O . . . . .	31
3.2	Some properties of EL-O . . . . .	35
3.3	EL-O as a fragment of epistemic logic $S5 \setminus GFP$ . . . . .	37
3.4	Adding constants: EL-OC . . . . .	40
3.5	Discussion: joint vision for arbitrary groups of agents . . . . .	43
3.6	Conclusion . . . . .	44
<b>4</b>	<b>On common knowledge whether</b>	<b>46</b>
4.1	An alternative axiomatization of S5 common knowledge . . . . .	50
4.2	Commonly knowing whether . . . . .	52
4.3	Discussion: epistemic logics weaker than S5 . . . . .	58
4.4	A second version of EL-O . . . . .	59
4.5	EL-O+ <i>Ind</i> as a fragment of epistemic logic S5 . . . . .	64
4.6	Conclusion . . . . .	67
<b>5</b>	<b>Planning with EL-O</b>	<b>68</b>
5.1	Epistemic planning with conditional effects . . . . .	69
5.2	Planning over finite states and complexity . . . . .	74
5.3	Applications . . . . .	77
5.4	Conclusion . . . . .	83
<b>6</b>	<b>Around EL-O planning</b>	<b>84</b>
6.1	Discussion: planning with other versions of EL-O . . . . .	85
6.2	A dynamic version of EL-OC: DEL-PPAOC . . . . .	87
6.3	Relation to planning with DEL event models . . . . .	91
6.4	Other related works . . . . .	98
6.5	Conclusion . . . . .	101



<b>7 True and mere belief about a proposition</b>	<b>103</b>
7.1 ‘Belief-about’ and epistemic-doxastic situations . . . . .	106
7.2 Reduction of ‘about’ modalities . . . . .	109
7.3 Conclusion . . . . .	113
<b>8 On knowledge and belief bases</b>	<b>114</b>
8.1 An overview of LDA . . . . .	116
8.2 EL-O: Vision bases . . . . .	118
8.3 Relation to knowledge structures . . . . .	123
8.4 Relation to PEKBs . . . . .	124
8.5 True and mere beliefs . . . . .	124
8.6 Discussion and conclusion . . . . .	129
<b>9 Conclusion</b>	<b>130</b>

# Chapter 1

## Introduction

Imaginons un scénario dans lequel une lettre serait apportée à deux agents, la destinataire du courrier et son ami. Cette lettre contient une information inconnue des deux agents ; les deux sont donc curieux d'en connaître le contenu. La destinataire du pli préfère peut-être lire celui-ci en privé, mais si son compagnon la voit faire, ce dernier *saura qu'elle sait* ce que renferme le courrier. La destinataire pourrait aussi être d'accord pour partager l'information et lire la lettre à voix haute. Dans ce cas, il y aura *connaissance commune* du contenu du pli entre les deux agents : ils sauront tous deux ce que dit la lettre, sauront tous deux qu'ils le savent tous deux, sauront tous deux qu'ils savent tous deux qu'ils le savent tous deux, etc.

Les logiques épistémiques et doxastiques ont pour but de représenter ce type d'information sur les connaissances et les croyances des agents. La logique épistémique fut introduite par Hintikka dans [76], et popularisée à l'aune de l'intelligence artificielle dans des travaux tels que [52]. Ces logiques ont une grande variété de domaines d'application et permettent aussi bien d'implémenter des robots possédant une théorie de l'esprit [46] que d'analyser des protocoles blockchain [67] ou des jeux à information incomplète telles que le Cluedo ou Hanabi [100].

Nous nous intéressons dans cette thèse à l'utilisation de ces représentations pour la planification épistémique : étant donné une situation initiale, un certain nombre d'actions possibles, et un but, peut-on trouver une suite d'actions qui nous mèneront de l'état initial à notre but ? L'on peut par exemple souhaiter que tous les agents apprennent les secrets de tous les autres agents au bout d'une série d'appels téléphoniques comme dans le problème de bavardage [41], ou bien vouloir planifier les déplacements et communications des agents afin que certains d'entre eux prennent connaissance d'informations que d'autres ne doivent pas apprendre comme dans le problème de communication sélective de [80], ou bien encore essayer d'apprendre le contenu d'une lettre destinée à autrui comme dans notre exemple d'ouverture. Si la planification classique (c'est-à-dire non-épistémique) est un domaine de recherche riche et actif, il n'y a à ce jour pas de consensus clair sur la meilleure manière d'ajouter de l'incertitude en termes de croyances ou de connaissances des agents, et en particulier sur le meilleur compromis entre expressivité d'une part et complexité calculatoire d'autre part.

Traditionnellement, la connaissance et la croyance sont représentées à l'aide

de modèles de Kripke contenant des mondes possibles pour chaque agent [51]. Un agent connaît une certaine information si cette information reste inchangée dans tous les mondes que l'agent considère possible, de légers ajustements au niveau de la sémantique permettant de représenter différentes sortes de connaissance et de croyance. En particulier, il est souvent supposé que les agents ont une capacité d'introspection, autrement dit qu'ils ont conscience de ce qu'ils savent et même de ce qu'ils ne savent pas. La logique épistémique correspondante est appelée **S5**.

Des opérateurs dynamiques furent ajoutés à ce formalisme, tout d'abord sous la forme des annonces publiques de Plaza ([111], travaux d'abord publiés en 1989) ainsi que dans d'autres travaux sur des types particuliers d'actions [59, 60, 47] avant d'arriver au cadre général des modèles à événements [15] dans ce qui est désormais connu sous le nom de Logique Epistémique Dynamique (**DEL**). Ces modèles à événements utilisent des modèles de Kripke afin de représenter des événements possibles en plus des mondes possibles. (Voir [51] pour une présentation plus approfondie de **DEL**.) Ici encore, des restrictions sur ces modèles permettent de s'assurer que la connaissance a toutes les propriétés souhaitées ; on peut par exemple citer les travaux sur les modèles de plausibilité pour représenter les croyances et leur révision [122, 16].

Plusieurs présentations de ce genre de logique épistémique existent, prenant comme primitive soit l'opérateur 'savoir que' ( $K_i\varphi$  se lit "l'agent  $i$  sait que  $\varphi$  est vrai"), soit l'opérateur 'savoir si' ( $Kif_i\varphi$  se lit "l'agent  $i$  sait si  $\varphi$  est vrai ou non"). L'approche avec l'opérateur 'savoir si' a son origine dans les logiques de non-contingence [106, 78] et a été développée en tant qu'alternative à l'approche 'savoir que' dans des travaux tels que [54]. Les deux opérateurs sont interdéfinissables : un agent sait si une formule  $\varphi$  est vraie ou non s'il sait soit *que*  $\varphi$  est vrai, soit *que*  $\varphi$  est faux, et il sait que  $\varphi$  est vrai si  $\varphi$  est vrai et il sait *si*  $\varphi$  est vrai. Autrement dit,  $Kif_i\varphi \leftrightarrow K_i\varphi \vee K_i\neg\varphi$  et  $K_i\varphi \leftrightarrow \varphi \wedge Kif_i\varphi$ . Les présentations des logiques épistémiques standard avec l'opérateur 'savoir si' ont donc la même expressivité que celles avec l'opérateur 'savoir que'. D'autres logiques mettant en avant des opérateurs tels que 'savoir ce qu'est' et 'savoir comment' ont aussi été développées [125, 126, 14]. L'idée qu'il est particulièrement intéressant, lorsque l'on considère des applications en planification, de pouvoir raisonner sur la connaissance de valeurs de constantes telles que des numéros de téléphone ou des codes pour ouvrir des portes par le biais de logiques avec opérateur 'savoir ce qu'est' est défendue dans [125].

Si la logique épistémique standard présente une grande expressivité, cela vient au prix de complexités plus élevées que celles de la logique propositionnelle : le problème de satisfiabilité dans **S5** est **PSPACE**-complet dès qu'il y a plus d'un agent, et **EXPTIME**-complet si l'on ajoute un opérateur de connaissance commune. L'extension dynamique **DEL** présente aussi un coût élevé : en particulier, la planification **DEL** est indécidable [26] même lorsqu'on y impose des restrictions relativement strictes [27]. Il est donc utile de considérer des manières plus simples de représenter la connaissance et les actions, afin d'obtenir des résultats de complexité plus intéressants. De telles approches ont été considérées par exemple dans [45, 118, 77, 110, 119, 108, 84, 80, 107]. Elles prennent le parti soit de restreindre les actions de manière à ce que l'incertitude ne puisse pas 'trop' augmenter lors de l'application d'un plan [77, 80], soit de simplifier la logique statique [118, 107], lorsqu'elles n'optent pas pour les deux stratégies à la fois [108, 119]. Les logiques statiques simplifiées considérées

dans la littérature sont limitées à un seul agent, n'incorporent pas de notion de connaissance commune, ou bien restreignent les capacités de raisonnement des agents, soit en plaçant une borne supérieure sur la profondeur d'imbrication des opérateurs de connaissance (éliminant souvent toute connaissance d'ordre supérieure), soit en limitant l'applicabilité de ces opérateurs. En particulier, Lakemeyer et Lespérance [84] proposent un langage dans lequel l'opérateur 'savoir que' ne peut s'appliquer qu'à des formules d'une forme particulière de sorte que les agents ne peuvent connaître des disjonctions. Plus précisément, le langage consiste en des combinaisons booléennes de *littéraux modaux restreints*, définis par :

$$\lambda ::= p \mid \neg\lambda \mid K_i\lambda$$

Cela signifie que les situations dans lesquelles un agent sait qu'un autre agent en sait plus qu'elle ne peuvent être représentées dans cette approche : par exemple, l'agent 1 ne peut pas savoir que l'agent 2 sait si  $p$  est vrai ou non ( $K_1(K_2p \vee K_2\neg p)$ ) à moins de connaître elle-même la valeur de vérité de  $p$  (nous montrons ceci formellement dans le chapitre 6). Ceci représente un inconvénient majeur, car de telles situations sont fondamentales lorsque l'on souhaite représenter des agents pouvant communiquer entre eux : la situation décrite ci-dessus mènerait naturellement l'agent 1 à interroger l'agent 2 au sujet de  $p$  ; de manière plus générale, le fait de savoir à qui demander de l'information ainsi que le fait de savoir si quelqu'un a besoin d'une information particulière sont d'importants déclencheurs d'instances de communication.

Une logique épistémique s'inspirant de l'approche mentionnée ci-dessus fut proposée dans [70, 40]. Dans cette logique, l'applicabilité des opérateurs de connaissance est restreinte d'une manière similaire, mais l'opérateur 'savoir si' remplace l'opérateur 'savoir que'. Si ces deux opérateurs engendrent la même expressivité dans le cas général, cela n'est plus le cas lorsqu'on considère le langage restreint. En particulier, la formule  $K_1(K_2p \vee K_2\neg p)$  peut être exprimée à l'aide d'opérateurs 'savoir si' par  $Kif_2p \wedge Kif_1Kif_2p$ .

La logique qui en résulte est la Logique Epistémique de l'Observation, ou EL-O. Par analogie avec les logiques d'observation, dans lesquelles les connaissances des agents proviennent de ce qu'ils observent, les opérateurs dans EL-O ne s'appellent pas  $Kif_i$  et  $CKif_i$  mais  $S_i$  et  $JS$  :  $S_i\alpha$  se lit "l'agent  $i$  voit la valeur de vérité de  $\alpha$ ", et  $JS\alpha$  se lit "les agents *voient conjointement* la valeur de vérité de  $\alpha$ " (autrement dit, il y a connaissance commune de la valeur de vérité de  $\alpha$ ). Ces *atomes d'observabilité* sont définis selon la grammaire suivante :

$$\alpha ::= p \mid S_i\alpha \mid JS\alpha$$

La logique EL-O est suffisamment expressive pour capturer bon nombre de problèmes du monde réel : elle est multi-agent, a une forme de connaissance commune, et la connaissance est moins restreinte que dans les logiques dans lesquelles la profondeur épistémique est bornée, ou l'applicabilité des opérateurs limitée aux littéraux modaux restreints de [84, 107].

Dans cette thèse nous proposons une version mise à jour de EL-O, exposons les principales propriétés de cette logique, et montrons en particulier que l'expressivité mentionnée ci-dessus ne vient pas au prix d'une grande complexité calculatoire : le problème de satisfiabilité EL-O est NP-complet, c'est-à-dire la même classe de complexité que pour le calcul propositionnel classique. Du côté dynamique, nous définissons un cadre pour faire de la planification avec EL-O

dans lequel les actions ont des préconditions et des effets conditionnels, et montrons que le problème de décision de l'existence d'une solution à un problème de planification ainsi défini est PSPACE-complet, à la fois pour la recherche de plans séquentiels et de plans parallèles. Ceci est encore une fois la même classe de complexité que les problèmes correspondants en planification classique. La preuve repose sur des traductions polynomiales des problèmes de planification EL-O vers des problèmes de planification classiques.

Au cœur de cette efficacité, nous retrouvons les capacités restreintes de raisonnement des agents ainsi que l'approche symbolique : les états EL-O sont simplement des ensembles d'atomes d'observabilité indiquant les vérités ontiques ainsi que ce que les agents savent dans l'état considéré. D'autres cadres de planification épistémique simple utilisent aussi des bases de connaissance [109, 119], qui sont aussi prisées pour leur efficacité dans la littérature de vérification de modèles [120, 21, 22]. L'approche symbolique est encore plus courante lorsque l'on considère la représentation de la croyance [3, 82, 68, 113, 20, 114, 94], pour laquelle une des préoccupations principales est le maintien de la cohérence des croyances d'un agent alors qu'il apprend de nouvelles informations. Cependant, contrairement à la connaissance, la croyance ne se prête pas si facilement que cela au passage de l'opérateur 'croire que' à un opérateur 'croire si', ou 'avoir une croyance à propos de'. Il est toutefois intéressant de considérer comment les principes du cadre EL-O pourraient être adaptées au raisonnement avec de la croyance en plus de la connaissance. Nous abordons ceci à l'aide de deux nouveaux opérateurs : l'opérateur de 'croyance correcte'  $TBA_i$  et celui de 'simple croyance'  $MBA_i$ .

Si nous explorons tout au long de cette thèse la relation entre EL-O et DEL, la question de la relation entre EL-O et d'autres logiques avec bases de connaissances ou de croyances se pose également. Parmi ces dernières, la logique des attitudes doxastiques (LDA) de Lorini [94] s'avère fournir un cadre très général dans lequel EL-O ainsi qu'un certain nombre d'autres approches peuvent être exprimés. Cela nous permet de mieux déterminer les liens entre ces différentes logiques.

La thèse est organisée de la manière suivante : dans le chapitre 2, nous donnons un aperçu de notions préexistantes qui nous seront utiles tout au long du reste de la thèse. Dans le chapitre 3, nous présentons EL-O et prouvons ses principales propriétés, y compris la NP-complétude de son problème de satisfiabilité. Nous montrons également que EL-O est un fragment de la logique S5 standard dans laquelle le principe d'induction pour la connaissance commune n'est pas satisfait. Dans le chapitre 4, nous étudions avec un peu plus de recul les axiomatisations de la connaissance commune et de la connaissance commune avec opérateur 'savoir si', et prenons la nouvelle axiomatisation que nous proposons comme inspiration pour définir une version de EL-O qui est un vrai fragment de S5, suivant le principe d'induction pour la connaissance commune. Dans le chapitre 5, nous définissons les actions et tâches de planification EL-O et montrons que le problème de décision de l'existence d'une solution à un problème de planification EL-O est PSPACE-complet, à la fois pour la planification séquentielle et pour la planification parallèle. Nous illustrons ce formalisme sur un éventail d'exemples d'applications, certains étant nouveaux et d'autres provenant de la littérature. Dans le chapitre 6, nous explorons un certain nombre de propriétés ayant rapport à la planification EL-O, comme l'extension des résultats obtenus aux autres versions de EL-O présentées dans les chapitres

précédents, l'ajout d'opérateurs dynamiques au langage, et enfin la relation entre le formalisme EL-O et DEL ainsi qu'entre EL-O et d'autres approches pour la planification épistémique simplifiée. Dans le chapitre 7, nous abordons une manière d'adapter les principes de EL-O afin de définir une logique simple de connaissance et de croyance avec opérateurs de 'croyance correcte à propos de' et 'simple croyance à propos de'. Enfin, dans le chapitre 8 nous montrons comment intégrer EL-O ainsi que d'autres logiques similaires au cadre plus général des logiques LDA.

La liste des publications sur lesquelles est basée cette thèse est donnée à la fin de ce chapitre.

---

Imagine a situation in which a letter is brought to two agents. The letter is addressed to one of the agents, and contains a piece of information. The recipient of the letter and their companion alike are curious to know the contents of the letter. The former may wish to read the letter privately, but if their friend sees them doing so, this friend will then *know that they know* the contents of the letter, and start asking questions. If this turns out to be to no avail, the companion might turn to reading the letter in secret, making sure that the recipient does not see them and *does not know that they know* what is in the letter. But the recipient might also be open to sharing, and read the letter out loud: in this case, there will be *common knowledge* between the two agents of the piece of information. That is, not only will they both know this information, but they will both know that they both know it, both know that they both know that they both know it, and so on.

Representing this type of information about knowledge and belief of agents is the subject of epistemic and doxastic logics. Epistemic logic was introduced by Hintikka in [76] and popularized from an AI point of view in works such as [52]. The range of applications of these logics varies greatly, from providing representations of knowledge for intelligent robots [46] to investigating blockchain protocols [67] or analyzing games with imperfect information such as Cluedo or Hanabi [100].

One way of using these representations that is of particular interest to us is through epistemic planning: given an initial situation, some set of allowed actions, and a goal, can we find a sequence of actions that will bring us from our initial situation to our goal? This can mean everybody learning everybody else's secrets through series of phone calls as in the gossip problem [41], planning movements and announcements so that only desired agents learn crucial information while others stay ignorant of it as in the selective communication problem of [80], or indeed figuring out the contents of a letter not destined to us as in our opening example. While classical (i.e., non-epistemic) planning is a very rich and active field, there is as of yet no clear consensus on the best way of adding uncertainty to the mix in terms of agents' knowledge or beliefs, or on the best balance between expressivity and computational complexity.

Traditionally, knowledge and belief are represented using Kripke models with possible worlds for each agent [51]. An agent knows a piece of information if that information is the same throughout the worlds which that agent considers possible; slight adjustments in the semantics allow for representation of different kinds and strengths of knowledge or belief. In particular, a common assumption is that agents should have a capacity of introspection, that is, to be aware of

what they do know and even of what they don't know. The corresponding logic of knowledge is called S5.

Dynamics were added to this formalism starting with Plaza's Public Announcement Logic ([111], work originally published in 1989) as well as other works on specific types of updates [59, 60, 47], before being generalized through the use of event models [15], resulting in what is now known as Dynamic Epistemic Logic (DEL). These event models similarly uses Kripke models to represent possible events as well as possible worlds. (We refer the reader to [51] for a more in-depth overview of DEL.) Once again, specific restrictions on these models allow for reasoning about different types of knowledge and belief (see e.g. the work on plausibility models for beliefs [122, 16]).

Several presentations of this type of epistemic logic exist, which take either 'knowing that' ( $K_i\varphi$  reads "agent  $i$  knows that  $\varphi$  is true") or 'knowing whether' ( $Kif_i\varphi$  reads "agent  $i$  knows whether or not  $\varphi$  is true") as a primitive. The 'knowing whether' approach has roots in non-contingency logics [106, 78] and has been developed as an alternative to 'knowing that' in works such as [54]. The two operators are interdefinable: an agent knows whether or not a formula  $\varphi$  is true if they either know *that*  $\varphi$  is true or they know *that*  $\varphi$  is false, and they know *that*  $\varphi$  is true if  $\varphi$  is true and they know *whether* or not  $\varphi$  is true. That is,  $Kif_i\varphi \leftrightarrow K_i\varphi \vee K_i\neg\varphi$  and  $K_i\varphi \leftrightarrow \varphi \wedge Kif_i\varphi$ . Therefore presentations of the standard epistemic logic with 'knowing whether' operators have the same expressivity as those with 'knowing that'. In addition to these, other logics have been developed arguing for operators such as 'knowing what' or 'knowing how' [125, 126, 14]. As argued in [125], being able to reason about knowledge of the value of constants like a telephone number or the code to open a door through logics of 'knowing what' is of particular interest when considering planning applications.

While it is very powerful in terms of expressivity, reasoning in standard epistemic logic is strictly more complex than reasoning in propositional logic: deciding satisfiability of a formula in S5 is PSPACE-complete if there is more than one agent, and EXPTIME-complete if the operator of common knowledge is present [66]. Reasoning with the dynamic extension DEL also proves to be impractical in many cases: in particular, planning in DEL is undecidable [26] even under severe restrictions [27]. It is therefore useful to consider simpler ways of representing knowledge and actions, leading to reduced complexity results. Such approaches have been considered e.g. in [45, 118, 77, 110, 119, 108, 84, 80, 107]. These approaches either restrict the range of actions so that uncertainty cannot grow 'too much' throughout the application of a plan [77, 80], simplify the static logic [118, 107], or both [108, 119]. The simple static logics considered either are limited to a single agent, lack common knowledge, or restrict the reasoning capabilities of agents, either by placing an upper bound on the number of nested knowledge operators (often dropping higher order knowledge altogether) or by limiting the scope of these operators. In particular, Lakemeyer and Lespérance [84] propose a language of 'knowing that' in which the scope of the knowledge operator is restricted so that agents cannot know disjunctions. More precisely, formulas are boolean combinations of *restricted modal literals*, defined as follows:

$$\lambda ::= p \mid \neg\lambda \mid K_i\lambda$$

This means that one cannot represent situations in which one agent knows that another agent knows more than they do: agent 1 can not know that agent 2

knows whether or not  $p$  is true ( $K_1(K_2p \vee K_2\neg p)$ ) without knowing the truth value of  $p$  themselves (we will show this formally in Chapter 6). This is a major drawback because such formulas are fundamental in communication and more generally in any form of interaction: the situation described would naturally lead agent 1 to ask agent 2 about  $p$ ; more generally, knowing who to ask about information as well as knowing whether someone is in need of more information are major factors in prompting communication.

Drawing inspiration from this last approach, an epistemic logic was proposed in [70, 40] in which the scope of knowledge operators is similarly restricted, but shifting from ‘knowing that’ to ‘knowing whether’ operators. While the two presentations are equivalent in the general case, this is no longer true when considering the restricted language: in particular, we can express  $K_1(K_2p \vee K_2\neg p)$  with ‘knowing whether’ operators as  $Kif_2p \wedge Kif_1Kif_2p$ .

The resulting logic is the Epistemic Logic of Observation, or EL-O. Drawing a parallel to logics of observation, in which agents’ knowledge come from what they see, the operators in EL-O are not called  $Kif_i$  and  $CKif$  but  $S_i$  and  $JS$ :  $S_i \alpha$  reads “agent  $i$  sees the truth value of  $\alpha$ ”, and  $JS \alpha$  reads “all agents *jointly* see the truth value of  $\alpha$ ”. The grammar of these so-called *observability atoms* is:

$$\alpha ::= p \mid S_i \alpha \mid JS \alpha$$

EL-O is expressive enough to capture a number of real life problems: it is multi-agent, has a form of common knowledge, and offers agents a wider scope of knowledge than logics in which knowledge is either bounded in depth or restricted to the modal literals of [84].

In this thesis we provide an updated version of EL-O, give a number of its properties, and show in particular that this expressivity does not come at a high complexity price: the satisfiability problem is NP-complete, that is, the same complexity class as satisfiability in Classical Propositional Calculus. On the dynamic side, we define a framework for epistemic planning with EL-O in which actions have preconditions and conditional effects and show that deciding the existence of a solution to a planning problem is PSPACE-complete, for both sequential and parallel plan search. Once again, this is the same complexity class as the classical planning counterpart of this problem. The proof involves defining polynomial translations to the corresponding classical problems.

At the core of this efficiency are the limited reasoning ability of agents as well as the symbolic approach: EL-O states are simply sets of epistemic atoms indicating which propositional facts and which knowledge holds at that state. This feature of reasoning with so-called knowledge bases is shared by a few other frameworks for simple epistemic planning [109, 119] and is also studied with regards to model checking efficiency [120, 21, 22]. The symbolic approach with bases is perhaps more widespread when considering representations of beliefs [3, 82, 68, 113, 20, 114, 94], in which a core concern is that of maintaining consistency of an agent’s beliefs even when receiving new information. However, unlike knowledge, it is not that straightforward to go from a logic of ‘belief that’ to a logic ‘belief whether’, or more naturally ‘belief about’. Nevertheless, it is interesting to consider how the principles of the EL-O framework could be adapted to reason about beliefs as well as knowledge. We will be discussing this through two new operators: the operator of ‘true belief’  $TBA_i$  and that of ‘mere belief’  $MBA_i$ .



While we discuss the relationship between EL-O and DEL throughout this thesis, another question is that of comparing EL-O to other logics of knowledge and belief bases. Among these, we find the Logic of Doxastic Attitudes (LDA) proposed by Lorini [94] to be a very general framework in which EL-O as well as many of these related approaches can be embedded. This allows us to better characterize the relationship between all of these logics.

The thesis is organized as follows: in Chapter 2, we give an overview of preexisting notions that will be useful to us throughout the rest of the thesis. In Chapter 3, we present EL-O and prove a number of its properties, including NP-completeness of its satisfiability problem. We also show that EL-O is a fragment of the standard S5 logic without the induction principle for common knowledge. In Chapter 4, we take a step back to discuss axiomatizations of common knowledge and ‘common knowledge whether’, then take the new axiomatization that we propose as inspiration to define a version of EL-O that is a true fragment of S5, with the induction principle for common knowledge. In Chapter 5, we define EL-O actions and planning tasks and show that deciding the existence of a solution to an EL-O planning task is PSPACE-complete both for sequential and parallel planning. We illustrate the framework with a number of application problems, both new and from the literature. In Chapter 6, we discuss some properties around EL-O planning, such as its extension to variants of EL-O proposed in previous chapters, the addition of dynamic operators to the language and finally its relation to DEL and other approaches to lightweight epistemic planning. In Chapter 7, we discuss how to take ideas from EL-O to define a simple logic of knowledge and belief with operators of ‘true belief about’ and ‘mere belief about’. Finally, in Chapter 8 we show how to embed EL-O as well as other related logics into the framework of LDA.

This thesis is mainly based on the following publications:

- Cooper, M. C., Herzig, A., Maffre, F., Maris, F., Perrotin, E., & Régnier, P. (2020). *A lightweight epistemic logic and its application to planning*. Artificial Intelligence, 103437.
- Herzig, A. & Perrotin, E. (2020). *On the axiomatization of common knowledge*. Advances in Modal Logic (AiML 2020) (13:309–328).
- Cooper, M., Herzig, A., Maris, F., Perrotin, E., & Vianey, J. (2020). *Lightweight Parallel Multi-Agent Epistemic Planning*. Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR 2020) (Vol. 17, No. 1, pp. 274-283).
- Herzig, A. & Perrotin, E. (2021). *True belief and mere belief about a proposition and the classification of epistemic-doxastic situations*. Filosofiska Notiser (1-Special issue on modal logic) (pp.103-117).
- Herzig, A., Maris, F., & Perrotin, E. (2021). *A dynamic epistemic logic with finite iteration and parallel composition*. Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR 2021) (Vol. 18, No. 1, pp. 676-680).
- Herzig, A. & Perrotin, E. (2022, forthcoming). *Efficient reasoning about knowledge and common knowledge*. Many-Valued and Modal Logics (Essays in honour of Yu. V. Ivlev).

# Chapter 2

## Background

### Contexte

Dans ce chapitre nous rappelons un certain nombre de notions que nous serons amenés à utiliser tout au long de ce manuscrit. Nous donnons d’abord les langages, sémantiques et axiomatisations des logiques standard de connaissance, avec ou sans croyance ou connaissance commune. Nous passons ensuite à l’aspect dynamique et rappelons les bases de la planification classique ainsi que les modèles à événements de la logique épistémique dynamique standard.

---

In this chapter we go over some preexisting notions which we will be building upon in the rest of this thesis. We start with recalling the language, semantics and axiomatizations of a number of standard static epistemic logics without and with common knowledge. We then move on to dynamics and go over some basics of classical planning as well as event models in DEL.

### Contents

---

<b>2.1</b>	<b>Standard classical and epistemic logics . . . . .</b>	<b>12</b>
2.1.1	Boolean languages and Classical Propositional Calculus CPC . . . . .	12
2.1.2	The standard epistemic language . . . . .	13
2.1.3	Kripke models and usual epistemic logics . . . . .	13
2.1.4	Three epistemic-doxastic logics . . . . .	14
<b>2.2</b>	<b>Shared and common knowledge . . . . .</b>	<b>16</b>
2.2.1	Kripke models with common knowledge . . . . .	17
2.2.2	Shared knowledge . . . . .	17
2.2.3	Two axiomatizations of common knowledge . . . . .	18
2.2.4	Equivalence of the two axiomatics . . . . .	20
2.2.5	From ‘knowing that’ to ‘knowing whether’ . . . . .	22
2.2.6	Bisimulations . . . . .	22
<b>2.3</b>	<b>Classical and DEL planning . . . . .</b>	<b>23</b>
2.3.1	Action descriptions and planning tasks . . . . .	23
2.3.2	Semantics and solvability of planning tasks . . . . .	24
2.3.3	The case of classical sequential planning . . . . .	25

2.3.4	The case of classical parallel planning . . . . .	26
2.3.5	DEL planning: event models and product updates . . . . .	27

---

## 2.1 Standard classical and epistemic logics

In this section we give an overview of the standard logics we will be discussing in the rest of the thesis. We start with a general presentation of boolean languages and Classical Propositional Calculus (CPC), and then move on to logics of knowledge and belief, with a focus on **S5**.

In the entire thesis we will be using *Prop* to denote a countable set of propositional variables with typical members  $p, q$ , etc., and *Ag* to denote a finite set of agents with typical members  $i, j$ , etc.

### 2.1.1 Boolean languages and Classical Propositional Calculus CPC

The language  $\mathcal{L}_{\text{bool}}(\mathbb{V})$  of boolean formulas over some vocabulary  $\mathbb{V}$  combines the elements of  $\mathbb{V}$  by means of the boolean operators. It is defined by the grammar

$$\mathcal{L}_{\text{bool}}(\mathbb{V}) : \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi$$

where  $p$  ranges over  $\mathbb{V}$ . The boolean operators  $\top, \perp, \vee, \rightarrow$ , and  $\leftrightarrow$  are defined in the standard way, and we use the standard conventions for omitting parentheses. The *vocabulary of*  $\varphi$  is defined inductively by:  $\text{Voc}(p) = \{p\}$ , for  $p \in \mathbb{V}$ ;  $\text{Voc}(\neg\varphi) = \text{Voc}(\varphi)$ ;  $\text{Voc}(\varphi \wedge \varphi') = \text{Voc}(\varphi) \cup \text{Voc}(\varphi')$ .

The vocabulary  $\mathbb{V}$  is usually supposed to be a countable set of propositional variables *Prop* that are not analysed further. However, in the lightweight epistemic logic EL-O that will be presented over the next few chapters  $\mathbb{V}$  has some structure: its observability atoms involve propositional variables and agents. Related to that, instead of defining the length of an atomic formula to be 1 we suppose here that each  $p \in \text{Prop}$  has some length  $\ell(p) \geq 1$ . The length of complex formulas is then defined inductively from the length of atomic formulas by:  $\ell(\neg\varphi) = \ell(\varphi) + 1$  and  $\ell(\varphi \wedge \varphi') = \ell(\varphi) + \ell(\varphi') + 1$ . For example, if  $\ell(p) = 1$  then  $\ell(\top) = 5$  because  $\top$  abbreviates  $\neg(p \wedge \neg p)$ .

A *state* is a subset of the vocabulary  $\mathbb{V}$ . We denote states by  $s, s'$ , etc. The number of elements of a finite state  $s$  is denoted by  $|s|$ . In *Classical Propositional Calculus* (CPC) the formulas of  $\mathcal{L}_{\text{bool}}(\mathbb{V})$  are interpreted in the set of all states  $2^{\mathbb{V}}$  according to the following truth conditions:

$$\begin{array}{ll} s \models p & \text{iff } p \in s; \\ s \models \neg\varphi & \text{iff } s \not\models \varphi; \\ s \models \varphi \wedge \varphi' & \text{iff } s \models \varphi \text{ and } s \models \varphi'. \end{array}$$

for every  $p \in \mathbb{V}$ . A formula  $\varphi$  is CPC *satisfiable* iff  $s \models \varphi$  for every  $s \in 2^{\mathbb{V}}$ , and it is CPC *valid* iff  $s \models \varphi$  for some  $s \in 2^{\mathbb{V}}$ . It is known that CPC-satisfiability of a formula is NP-complete and that CPC-validity of a formula is coNP-complete.

### 2.1.2 The standard epistemic language

The grammar of the standard language  $Fml_{EL}$  is:

$$Fml_{EL} : \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid CK\varphi$$

where  $p$  ranges over the set of propositional variables  $Prop$ . The formula  $K_i\varphi$  reads “agent  $i$  knows that  $\varphi$ ” and  $CK\varphi$  reads “it is common knowledge that  $\varphi$ ”. From there we can define the operator  $EK$  of *shared knowledge*:  $EK\varphi$  is defined as  $\bigwedge_{i \in Agt} K_i\varphi$ , and reads “everybody knows that  $\varphi$ ”.

**Example 2.1** (Gossip). In the gossip problem [41], each agent has a secret, and can call other agents in order to share their own secret as well as any other secret they might know. While many variants of this problem exist featuring various goals and modalities for sharing information, the goal of the original gossip problem is simply to obtain shared knowledge of all secrets. This is expressed in the standard epistemic language  $Fml_{EL}$  by  $\bigwedge_{i,j \in Agt} (K_j s_i \vee K_j \neg s_i)$ . Shared knowledge up to depth 2 is expressed by  $\bigwedge_{i,j,r \in Agt} (K_r K_j s_i \vee K_r K_j \neg s_i)$ . In the generalised gossip problem [70, 73] the goal is shared knowledge up to depth  $k$ , written  $\bigwedge_{i \in Agt} \bigwedge_{\vec{K} \in \{K_j : j \in Agt\}^k} (\vec{K} s_i \vee \vec{K} \neg s_i)$  in  $Fml_{EL}$ , where  $\{K_j : j \in Agt\}^k$  is the set of all sequences of individual ‘knowing that’ operators of length  $k$ .

### 2.1.3 Kripke models and usual epistemic logics

Let us first set aside common knowledge and focus on individual knowledge; we will come back to common knowledge in section 2.2.

Semantically, a *Kripke model* [24], also called *possible worlds model*, is a tuple  $M = \langle W, \{R_i\}_{i \in Agt}, V \rangle$  where  $W$  is a nonempty set of possible worlds; every  $R_i$  is a binary relations on  $W$ ; and  $V : W \rightarrow 2^{Prop}$  is a valuation. The interpretation of  $Fml_{EL}$  formulas with no common knowledge operator in a pointed Kripke model  $(M, w)$  is:

$$\begin{aligned} (M, w) \models p & \quad \text{iff } p \in V(w); \\ (M, w) \models \neg\varphi & \quad \text{iff } (M, w) \not\models \varphi; \\ (M, w) \models \varphi \wedge \psi & \quad \text{iff } (M, w) \models \varphi \text{ and } (M, w) \models \psi; \\ (M, w) \models K_i\varphi & \quad \text{iff } (M, w') \models \varphi \text{ for every } w' \text{ such that } wR_iw'. \end{aligned}$$

A formula  $\varphi$  is *valid in  $M$*  if  $(M, w) \models \varphi$  for every  $w \in W$ ; if  $\mathbf{C}$  is a class of Kripke models,  $\varphi$  is *valid in  $\mathbf{C}$*  if  $\varphi$  is valid in every Kripke model in  $\mathbf{C}$ .

The standard classes of Kripke models which we will be considering are defined via required properties, called *frame properties*, of the binary relations  $R_i$ . First, the logic of all Kripke frames is  $K(K_i)$ , an axiomatization of which is given in Table 2.1. The specific frame properties which we will be considering and the corresponding axioms are listed in Table 2.2. A number of standard logics of interest to us are listed in Table 2.3. Logics of knowledge are usually considered to be at least  $KT(K_i)$ : agents’ knowledge should be truthful.  $S4(K_i)$  requires agents to have *positive introspection*: if they know something, then they should know that they know it. The logic we base most of this thesis on is  $S5(K_i)$ , in which agents have both positive and *negative introspection*: they know exactly what they do and don’t know. Semantically, in  $S5(K_i)$  Kripke models all relations  $R_i$  are equivalence relations. Finally, the logic  $S4.2(K_i)$  is interesting as a logic of knowledge and belief; we expand on this in Section 2.1.4.

$CPC$	axiomatics of classical propositional calculus
$RN(K_i)$	from $\varphi$ , infer $K_i\varphi$
$K(K_i)$	$K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$

Table 2.1: Axioms for the logic  $\mathsf{K}(K_i)$ 

Axiom name	Property of $R_i$	Axiom
$T(K_i)$	Reflexivity	$K_i\varphi \rightarrow \varphi$
$4(K_i)$	Transitivity	$K_i\varphi \rightarrow K_iK_i\varphi$
$.2(K_i)$	Confluence	$\neg K_i\neg K_i\varphi \rightarrow K_i\neg K_i\neg\varphi$
$5(K_i)$	Euclideaness	$\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$

Table 2.2: Some standard frame properties for Kripke models and the corresponding axioms

### 2.1.4 Three epistemic-doxastic logics

In this section we recall three epistemic-doxastic logics of ‘knowledge that’ and ‘belief that’: one basic system and two possible extensions.

In epistemology there is a long-standing debate about the relation between knowledge and belief. For long it was taken for granted that knowledge can be reduced to belief by defining ‘knowledge that  $\varphi$ ’ as ‘justified true belief that  $\varphi$ ’. However, Gettier’s counterexample [61] sparked new debates on the subject that are still ongoing today, and there is no consensus about whether such a reduction is possible and how it should be defined. A more cautious enterprise is to take both  $B_i$  and  $K_i$  as primitives and to study the interaction between these two modal operators. This however has to be done with care, as we will see later in this section.

The traditional language of epistemic-doxastic logic, which we denote here by  $Fml_{EDL}$ , is defined by the following grammar:

$$Fml_{EDL} : \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_i p \mid K_i p$$

where  $p$  ranges over a countable set  $Prop$  of propositional variables and  $i$  ranges over a finite set  $Agnt$  of agents. With the language  $Fml_{EDL}$  we are going to study three different epistemic-doxastic logics.

The weakest logic EDL is axiomatized by the first four lines of Table 2.5, i.e., by  $KD5(B_i)+S4(K_i)+KiB+BiKB$ . An axiomatization of  $KD5(B_i)$  is given in Table 2.4, and the axiomatization of  $S4(K_i)$  is the one given in Section 2.1.3.

Logics	Axiomatization
$KT(K_i)$	$\mathsf{K}(K_i)+ T(K_i)$
$S4(K_i)$	$KT(K_i)+ 4(K_i)$
$S4.2(K_i)$	$S4(K_i)+ .2(K_i)$
$S5(K_i)$	$KT(K_i)+ 5(K_i)$

Table 2.3: Some standard modal logics

$RN(B_i)$	from $\varphi$ , infer $B_i\varphi$
$K(B_i)$	$B_i(\varphi \rightarrow \psi) \rightarrow (B_i\varphi \rightarrow B_i\psi)$
$D(B_i)$	$B_i\varphi \rightarrow \neg B_i\neg\varphi$
$5(B_i)$	$\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$

Table 2.4: An axiomatization of  $KD5(B_i)$ 

$KD5(B_i)$	the principles of modal logic $KD5$ for $B_i$
$S4(K_i)$	the principles of modal logic $S4$ for $K_i$
$KiB$	$K_i\varphi \rightarrow B_i\varphi$
$BiKB$	$B_i\varphi \rightarrow K_iB_i\varphi$
$BiBK$	$B_i\varphi \rightarrow B_iK_i\varphi$
$5(K_i)$	$\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$

Table 2.5: Principles of the three logics  $EDL$  (first four lines),  $EDL+BiBK$ , and  $EDL+5(K_i)$ .

The two remaining axioms are interaction axioms governing the interplay of knowledge and belief:  $KiB$  says that knowledge implies belief;  $BiKB$  is the axiom of positive introspection of belief.

*Remark 2.1.* The positive introspection axiom  $4(B_i)$ :  $B_i\varphi \rightarrow B_iB_i\varphi$  is not in our list but can be derived from  $BiKB$  and  $KiB$ . Our logic of belief is therefore  $KD45(B_i)$ .

Similarly, the negative introspection axiom  $\neg B_i\varphi \rightarrow K_i\neg B_i\varphi$  is not in our list but can be derived from  $KD5(B_i)$  and positive introspection of belief.<sup>1</sup>

We will moreover be considering two extensions of  $EDL$  the axioms of which are stated in the last two lines of Table 2.5. For the first, which we call  $EDL+BiBK$ , we add an axiom of strong belief  $BiBK$  to  $EDL$ : if  $i$  believes that  $\varphi$  then  $i$  believes that she knows that  $\varphi$ . The second, which we call  $EDL+5(K_i)$ , does not feature the axiom  $BiBK$ , but has negative introspection for knowledge instead.

*Remark 2.2.* The strong belief axiom  $BiBK$  makes the following equivalences provable in  $EDL+BiBK$ :

$$\begin{aligned} K_i(B_i\varphi \rightarrow K_i\psi) &\leftrightarrow B_i\varphi \rightarrow K_i\psi, \\ K_i(\varphi \rightarrow B_i\psi) &\leftrightarrow K_i\neg\varphi \vee B_i\psi, \\ K_i(\varphi \rightarrow \neg B_i\psi) &\leftrightarrow K_i\neg\varphi \vee \neg B_i\psi. \end{aligned}$$

*Remark 2.3.* It is known that the extension of  $EDL$  by both  $BiBK$  and  $5(K_i)$  is not very interesting: in that logic belief implies knowledge [102]. To see that,

<sup>1</sup>Here is a formal proof:

- |   |                                       |
|---|---------------------------------------|
| 1. $\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$       | $5(B_i)$                              |
| 2. $B_i\neg B_i\varphi \rightarrow K_iB_i\neg B_i\varphi$ | $BiKB$                                |
| 3. $K_iB_i\neg B_i\varphi \rightarrow K_i\neg B_i\varphi$ | by $4(B_i)$ , $D(B_i)$ , $K_i$ normal |

it suffices to prove that  $(B_i\varphi \wedge \neg K_i\varphi) \rightarrow \perp$  becomes a theorem. Intuitively, *BiBK* characterizes agents not being able to tell apart their beliefs from their knowledge, while *5* characterizes agents knowing exactly what is knowledge and what is merely belief.<sup>2</sup> We refer the reader to the work of Voorbraak for further results and discussions [124, 62].

It was established by Lenzen that the logic *EDL+BiBK* is strongly related to the modal logic *S4.2*.

**Proposition 2.1** ([88]). *The equivalence  $B_i\varphi \leftrightarrow \neg K_i\neg K_i\varphi$  is a theorem of *EDL+BiBK*. The logic *S4.2* together with the axiom  $B_i\varphi \leftrightarrow \neg K_i\neg K_i\varphi$  is an equivalent axiomatization of *EDL+BiBK*.*

We have seen that *S4.2(K<sub>i</sub>)* Kripke models are models  $\langle W, \{R_i\}_{i \in \text{Agt}}, V \rangle$  such that the individual accessibility relations  $R_i$  are reflexive, transitive and confluent. In [34] *S4.2(K<sub>i</sub>)* is described as the logic of reflexive and transitive Kripke frames with a final cluster. That is, for every  $i \in \text{Agt}$ , there is a set of worlds  $W_i \subseteq W$  such that for any  $w \in W$  and  $w' \in W_i$  it is the case that  $wR_iw'$ . Proposition 2.1 then tells us that what agent  $i$  believes is exactly what is true in this final cluster:

$$(M, w) \models B_i\varphi \text{ iff } (M, w') \models \varphi \text{ for every } w' \in W_i.$$

Beyond Lenzen's papers, an excellent survey of extensions of *EDL* and their properties can be found in Aucher's papers [7] and [8]. (The second publication extends the conference version in [6].)

Proposition 2.1 allows us to settle the complexity of deciding provability of a formula  $\varphi$  in *EDL+BiBK*. If we replace all subformulas  $B_i\psi$  of  $\varphi$  by  $\neg K_i\neg K_i\psi$  then the resulting formula  $\varphi'$  only contains modal operators  $K_i$ , and its length is linear in the length of the original  $\varphi$ . Following Proposition 2.1,  $\varphi'$  is equivalent to  $\varphi$  in *EDL+BiBK*. Moreover and again by Proposition 2.1,  $\varphi'$  is a theorem of *EDL+BiBK* if and only if  $\varphi'$  is a theorem of *S4.2*. The results in [116, 35, 34] that the problem of deciding provability in *S4.2* is *PSPACE*-complete can be generalised to multi-agent *S4.2*, as confirmed in [35]. Hence both provability and consistency in *EDL+BiBK* are *PSPACE*-complete. We do not know whether complexity results for the logics *EDL* and *EDL+5(K<sub>i</sub>)* exist; we conjecture that provability is *PSPACE*-complete in these logics as well.

## 2.2 Shared and common knowledge

We now consider the properties of shared and common knowledge.

---

<sup>2</sup>A proof is:

- |  |                          |
|--|--------------------------|
| 1. $B_i\varphi \rightarrow B_iK_i\varphi$                        | <i>BiBK</i>              |
| 2. $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$              | <i>5(K<sub>i</sub>)</i>  |
| 3. $K_i\neg K_i\varphi \rightarrow B_i\neg K_i\varphi$           | <i>KiB</i>               |
| 4. $(B_iK_i\varphi \wedge B_i\neg K_i\varphi) \rightarrow \perp$ | <i>KD(B<sub>i</sub>)</i> |
| 5. $(B_i\varphi \wedge \neg K_i\varphi) \rightarrow \perp$       | from 1,2,3,4             |

### 2.2.1 Kripke models with common knowledge

A *Kripke model with common knowledge* is a tuple  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  where  $W$  is a nonempty set of possible worlds;  $R_{Agt}$  and every  $R_i$  are binary relations on  $W$  such that the transitive closure of the union of all  $R_i$  is a subset of  $R_{Agt}$ ; and  $V : W \rightarrow 2^{Prop}$  is a valuation. To obtain the full truth conditions for formulas of  $Fml_{EL}$  in pointed Kripke models, we combine the semantics given in Section 2.1.3 with the condition for the common knowledge operator:

$$(M, w) \models CK\varphi \quad \text{iff } (M, w') \models \varphi \text{ for every } w' \text{ such that } wR_{Agt}w'.$$

Just as before, a formula  $\varphi$  is *valid in*  $M$  if  $(M, w) \models \varphi$  for every  $s \in W$ ; if  $C$  is a class of Kripke models with common knowledge,  $\varphi$  is *valid in*  $C$  if  $\varphi$  is valid in every Kripke model in  $C$ . Just as with individual knowledge, standard properties of common knowledge are reflected through corresponding properties of  $R_{Agt}$ .

### 2.2.2 Shared knowledge

We now consider logics that are at least  $KT(K_i)$ . The operator  $EK$  is a normal modal operator: it obeys the modal schema  $K$  and the rule of necessitation  $RN$ . Moreover it obeys:

$$T(EK) \quad EK\varphi \rightarrow \varphi.$$

It is straightforward to prove that the following holds for logics of individual knowledge from  $KT(K_i)$  on:

**Proposition 2.2.** *The formula*

$$Def2(EKif) \quad (EK\varphi \vee EK\neg\varphi) \leftrightarrow \bigwedge_{i \in Agt} (K_i\varphi \vee K_i\neg\varphi)$$

*is a theorem of the  $KT(K_i)$  axiomatics.*

When individual knowledge is  $S5$ , it is an established result that the operator  $K_i$  obeys the axiom  $B$ :

$$B(K_i) \quad \neg\varphi \rightarrow K_i\neg K_i\varphi$$

In this case, despite the fact that the shared knowledge operator  $EK$  neither obeys positive nor negative introspection, it also obeys the  $B$  axiom:

**Proposition 2.3.** *The formula*

$$B(EK) \quad \neg\varphi \rightarrow EK\neg EK\varphi$$

*is a theorem of the  $S5(K_i)$  axiomatics.*

*Proof.* The proof is simple, but we give it here as we did not find it in the literature:

1.  $\varphi \rightarrow K_i\neg K_i\neg\varphi$   $B(K_i)$
2.  $\neg K_i\neg\varphi \rightarrow \neg EK\neg\varphi$
3.  $K_i\neg K_i\neg\varphi \rightarrow K_i\neg EK\neg\varphi$  from (2),  $K_i$  normal
4.  $\varphi \rightarrow K_i\neg EK\neg\varphi$  from (1), (3)
5.  $\varphi \rightarrow EK\neg EK\neg\varphi$  from (4)

□



<i>GFP</i> -based axiomatics	
At least $\text{KT}(K_i)$	see Table 2.3
$RN(CK)$	from $\varphi$ , infer $CK\varphi$
$K(CK)$	$CK(\varphi \rightarrow \psi) \rightarrow (CK\varphi \rightarrow CK\psi)$
$T(CK)$	$CK\varphi \rightarrow \varphi$
$*5(CK)$	$\neg CK\varphi \rightarrow CK\neg CK\varphi$
$FP'$	$CK\varphi \rightarrow EKCK\varphi$
$GFP$	$CK(\varphi \rightarrow EK\varphi) \rightarrow (\varphi \rightarrow CK\varphi)$
<i>RGFP</i> -based axiomatics	
At least $\text{KT}(K_i)$	see Table 2.3
$FP$	$CK\varphi \rightarrow EK(\varphi \wedge CK\varphi)$
$RGFP$	from $\varphi \rightarrow EK(\psi \wedge \varphi)$ , infer $\varphi \rightarrow CK\psi$

Table 2.6: Two axiomatizations of common knowledge for logics that are at least  $\text{KT}$ : the *GFP* axiomatics with an induction axiom of [65] (top) and the *RGFP* axiomatics with an induction rule of [66, 52] (bottom). For  $\text{S5}$  common knowledge, add the starred axiom  $*5(CK)$  to the *GFP*-based axiomatics.

### 2.2.3 Two axiomatizations of common knowledge

We now come back to common knowledge. An overview of the different axiomatizations of logics of common knowledge can be found in [91] where the relation between the underlying logic of individual knowledge and the resulting logic of common knowledge is studied in depth.

Common knowledge is generally considered to be a fixpoint of shared knowledge: if there is common knowledge that  $\varphi$ , then everybody knows that  $\varphi$ , everybody knows that everybody knows that  $\varphi$ , and so on. This means that common knowledge is always at least  $\text{S4}$ , and is reflected in the requirement in Kripke models that the transitive closure of the union of all  $R_i$  be included in  $R_{\text{Agt}}$ .

Whether the converse holds is less generally agreed upon: if it is true that  $EK\varphi$ ,  $EKEK\varphi$ ,  $EKEKEK\varphi$ , and so on, that is, if  $EK^k\varphi$  is true for every integer  $k$ , should it be the case that  $CK\varphi$ ? We call this property the *induction principle for common knowledge*. It has been argued by some that this would make common knowledge too strong [95, 69], and logics in which this does not hold have been developed, such as that by Genesereth and Nilsson [58], used in modal extensions of logic programming in the 90s [13, 12], in which common knowledge is seen as knowledge of a ‘fictitious knower’ or ‘any fool’.

Nevertheless, most standard logics of common knowledge do uphold this equivalence. We give in Table 2.6 two prevalent axiomatizations of such logics of common knowledge for logics that are at least  $\text{KT}$  [65, 66, 52]. Other axiomatics can be found in [87, 83]. Such axiomatizations are popular in Dynamic Epistemic Logics [23, 51]. Common knowledge implying any nesting of shared knowledge is captured by  $K(CK)$ ,  $T(CK)$ , and  $FP'$  on the left side, and by  $FP$  on the right. The other direction is captured by the *induction axiom GFP* on the

$S5(K_i)$	see Table 2.3
$RN(CK)$	from $\varphi$ , infer $CK\varphi$
$K(CK)$	$CK(\varphi \rightarrow \psi) \rightarrow (CK\varphi \rightarrow CK\psi)$
$T(CK)$	$CK\varphi \rightarrow \varphi$
$5(CK)$	$\neg CK\varphi \rightarrow CK\neg CK\varphi$
$FP'$	$CK\varphi \rightarrow EKCK\varphi$

Table 2.7: Axiomatization of the logic  $S5 \setminus GFP$  of S5-based common knowledge without the induction principle for common knowledge

left and the *induction rule*  $RGFP$  on the right. When considering S5 common knowledge, the axiom  $5(CK)$  is added in the  $GFP$ -based axiomatics, while in the  $RGFP$ -based axiomatics common knowledge is S5 iff individual knowledge is S5. We will call these axiomatics “KT-based” when referring to the version in which individual knowledge is  $KT(K_i)$ , and “S5-based” when referring to the version in which individual knowledge is  $S5(K_i)$ .

We will also call  $S5 \setminus GFP$  the logic described by all axioms of the S5-based  $GFP$  axiomatized except for  $GFP$ , as given in Table 2.7.  $S5 \setminus GFP$  Kripke models are Kripke models  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  in which  $R_{Agt}$  and all  $R_i$  are equivalence relations and  $(\bigcup_{i \in Agt} R_i)^* \subseteq R_{Agt}$ , where  $(\bigcup_{i \in Agt} R_i)^*$  is the transitive closure of the union of all  $R_i$ . In this thesis we will refer to Kripke models with S5 individual and common knowledge which obey the induction principle for common knowledge simply as S5 Kripke models. These are models  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  in which  $R_{Agt}$  and all  $R_i$  are equivalence relations and  $(\bigcup_{i \in Agt} R_i)^* = R_{Agt}$ .

Let us analyze the two axiomatics of Table 2.6 a bit further. It is a standard result in normal modal logics that axiom  $4$  can be proved from  $T$  and  $5$ . In the case of common knowledge,  $4(CK)$  is already a theorem of the KT-based  $GFP$  axiomatics thanks to the induction axiom schema:

**Proposition 2.4.** *The formula*

$$4(CK) \quad CK\varphi \rightarrow CKCK\varphi$$

*is a theorem of the KT-based GFP axiomatics.*

*Proof.* 1.  $CK(CK\varphi \rightarrow EKCK\varphi)$  from  $FP'$  and  $RN(CK)$   
 2.  $CK(CK\varphi \rightarrow EKCK\varphi) \rightarrow (CK\varphi \rightarrow CKCK\varphi)$   $GFP$   
 3.  $CK\varphi \rightarrow CKCK\varphi$  from (1) and (2)  $\square$

**Proposition 2.5.** *Axiom  $5(CK)$  is redundant in the S5-based GFP axiomatics.*

*Proof.* 1.  $\neg CK\varphi \rightarrow K_i\neg K_iCK\varphi$   $B(K_i)$   
 2.  $CK\varphi \rightarrow K_iCK\varphi$  from  $FP'$   
 3.  $K_i\neg K_iCK\varphi \rightarrow K_i\neg CK\varphi$  from (2),  $K_i$  normal  
 4.  $\neg CK\varphi \rightarrow K_i\neg CK\varphi$  from (1), (3)

5.  $\neg CK\varphi \rightarrow EK\neg CK\varphi$  from (4)
  6.  $CK(\neg CK\varphi \rightarrow EK\neg CK\varphi)$  from (5) by  $RN(CK)$
  7.  $CK(\neg CK\varphi \rightarrow EK\neg CK\varphi) \rightarrow (\neg CK\varphi \rightarrow CK\neg CK\varphi)$   $GFP$
  8.  $\neg CK\varphi \rightarrow CK\neg CK\varphi$  from (6) and (7)
- Note the use of the axiom  $B(K_i)$ , valid in  $S5(K_i)$  but not in general.  $\square$

The axiomatics with the induction rule  $RGFP$  are due to [66, 52]; the induction rule can actually be traced back to the analysis of common knowledge in the philosophical literature [117]. Interestingly and contrasting with the  $GFP$  axiomatics, the  $S5$  axioms and rules for  $CK$  are implicit here:

**Proposition 2.6.** *The formulas  $K(CK)$ ,  $T(CK)$ ,  $4(CK)$ , are theorems and the rule  $RN(CK)$  is derivable in the  $KT$ -based  $RGFP$  axiomatics. Moreover, the formula  $*5(CK)$  is a theorem of the  $S5$ -based  $RGFP$  axiomatics.*

*Proof.* The proofs are simple, but we give them here for completeness.  $K(CK)$  can be proved by substituting  $\varphi$  by  $CK\varphi \wedge CK(\varphi \rightarrow \psi)$  in  $RGFP$ , using  $FP$  and the fact that  $EK$  is a normal modal operator.  $T(CK)$  can be proved from  $FP$  and  $T(EK)$ .  $4(CK)$  can be proved by substituting both  $\varphi$  and  $\psi$  by  $CK\varphi$  in  $RGFP$ , using  $FP$  and the fact that  $EK$  is a normal modal operator. The rule  $RN(CK)$  can be derived with  $RGFP$  if we substitute  $\top$  for  $\varphi$  and  $\varphi$  for  $\psi$  and use the fact that  $EK$  is a normal modal operator. It is only the proof of  $*5(CK)$  which is a bit longer:

1.  $\varphi \rightarrow EK\neg EK\neg\varphi$   $B(EK)$
2.  $EK\neg EK\neg\varphi \rightarrow EK\neg CK\neg\varphi$  from  $FP$ ,  $EK$  normal
3.  $EK\neg CK\neg\varphi \rightarrow EKEK\neg EKCK\neg\varphi$  from  $B(EK)$ ,  $EK$  normal
4.  $EKEK\neg EKCK\neg\varphi \rightarrow EKEK\neg CK\neg\varphi$  from  $FP$ ,  $EK$  normal
5.  $EK\neg CK\neg\varphi \rightarrow EK(\neg CK\neg\varphi \wedge EK\neg CK\neg\varphi)$  from (3), (4),  $EK$  normal
6.  $EK\neg CK\neg\varphi \rightarrow CK\neg CK\neg\varphi$  from (5) by  $RGFP$
7.  $\varphi \rightarrow CK\neg CK\neg\varphi$  from (1), (2), (6)

$\square$

## 2.2.4 Equivalence of the two axiomatics

The  $RGFP$  axiomatics and the  $GFP$  axiomatics are both complete for the same semantics. Therefore all axioms in one system must be derivable in the other, and the inference rules of one system are admissible in the other. We are however not aware of a direct equivalence proof in the respective systems in the literature, so we give it below.<sup>3</sup> We prove the two directions:

<sup>3</sup>The paper by Bucheli et al. [32] establishes that  $RGFP$  is derivable from a variant of  $GFP$ ,  $CK(\varphi \rightarrow EK\varphi) \rightarrow (EK\varphi \rightarrow CK\varphi)$  (which they have to choose instead of  $GFP$  because they take  $\mathbb{K}$  as the logic of individual knowledge; see Remark 2.4). However their proof is indirect, making use of an intermediate system.

1. in the *KT*-based *RGFP* axiomatics,  $K(CK)$ ,  $T(CK)$ ,  $FP'$ ,  $GFP$  are theorems and  $RN(CK)$  is derivable, and in the *S5*-based *RGFP* axiomatics,  $5(CK)$  is a theorem;
2. in the *GFP* axiomatics,  $FP'$  is a theorem and *RGFP* is derivable.

We have already established in Section 2.2.3 that  $K(CK)$ ,  $T(CK)$ , and  $5(CK)$  are theorems of the relevant *RGFP* axiomatics. Second and quite obviously, as  $EK$  is a normal modal operator, we have that  $FP'$  is provable from  $FP$  and that, the other way round,  $FP$  is provable from  $FP'$  and  $T(CK)$ . It remains to prove the equivalence of the induction axiom and the induction rule.

**Proposition 2.7.** *The induction axiom  $GFP$  is a theorem of the *KT*-based *RGFP* axiomatics (and a fortiori of the *S5*-based *RGFP* axiomatics).*

- Proof.*
1.  $CK(\varphi \rightarrow EK\varphi) \rightarrow EKCK(\varphi \rightarrow EK\varphi)$  from  $FP$ ,  $EK$  normal
  2.  $(CK(\varphi \rightarrow EK\varphi) \wedge \varphi) \rightarrow EK\varphi$  from  $T(CK)$
  3.  $(CK(\varphi \rightarrow EK\varphi) \wedge \varphi) \rightarrow (EK\varphi \wedge EKCK(\varphi \rightarrow EK\varphi))$  from (1) and (2)
  4.  $(CK(\varphi \rightarrow EK\varphi) \wedge \varphi) \rightarrow EK((\varphi \wedge CK(\varphi \rightarrow EK\varphi)) \wedge \varphi)$  from (3),  $EK$  normal
  5.  $(CK(\varphi \rightarrow EK\varphi) \wedge \varphi) \rightarrow CK\varphi$  from (4) by *RGFP*  $\square$

**Proposition 2.8.** *The induction rule *RGFP* is derivable in the *GFP* axiomatics.*

- Proof.*
1.  $\varphi \rightarrow EK(\psi \wedge \varphi)$  hypothesis
  2.  $CK(\psi \wedge \varphi \rightarrow EK(\psi \wedge \varphi))$  from (1) by  $RN(CK)$
  3.  $CK(\psi \wedge \varphi \rightarrow EK(\psi \wedge \varphi)) \rightarrow (\psi \wedge \varphi \rightarrow CK(\psi \wedge \varphi))$  *GFP*
  4.  $\psi \wedge \varphi \rightarrow CK(\psi \wedge \varphi)$  from (2), (3)
  5.  $\varphi \rightarrow \psi \wedge \varphi$  from (1) by  $T(EK)$
  6.  $\varphi \rightarrow CK\psi$  from (5), (4),  $CK$  normal  $\square$

*Remark 2.4.* Notice how both proofs make use of the truth axiom  $T$ . This is due to the fact that  $T$  is ‘hidden’ in *GFP* in order for the presentation of *GFP* to be more concise. A more general presentation of the induction axiom found in the literature is the axiom  $CK(\varphi \rightarrow EK\varphi) \rightarrow (EK\varphi \rightarrow CK\varphi)$ . This axiom is equivalent to *GFP* in the presence of  $T$ , and equivalent to *RGFP* even without  $T$ .<sup>4</sup>

<sup>4</sup>We call *GFP'* this alternative axiom. Equivalence of *GFP* and *GFP'* in the presence of  $T$  is straightforward. A direct proof of the equivalence with *RGFP* is as follows.

From *RGFP*:

1.  $EK\varphi \wedge CK(\varphi \rightarrow EK\varphi) \rightarrow EK\varphi \wedge (EK\varphi \rightarrow EKEK\varphi)$  from  $FP$ ,  $EK$  normal
2.  $EK\varphi \wedge CK(\varphi \rightarrow EK\varphi) \rightarrow EK(\varphi \wedge EK\varphi \wedge CK(\varphi \rightarrow EK\varphi))$  from (1),  $EK$  normal
3.  $EK\varphi \wedge CK(\varphi \rightarrow EK\varphi) \rightarrow CK\varphi$  from (2) and *RGFP*

### 2.2.5 From ‘knowing that’ to ‘knowing whether’

An alternative presentation of logics of knowledge (that is, logics that are at least KT) takes ‘knowing whether’ rather than ‘knowing that’ as a core concept. The language thus becomes:

$$Fml_{EL}^{if} : \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid Kif_i\varphi \mid CKif\varphi$$

where  $p$  ranges over *Prop*.  $Kif_i\varphi$  reads “agent  $i$  knows whether or not  $\varphi$  is true”, and  $CKif\varphi$  reads “it is common knowledge whether or not  $\varphi$  is true”. The operator of shared knowledge whether,  $EKif$ , can be defined from individual knowledge whether:  $EKif\varphi$  is defined as  $\bigwedge_{i \in Agt} Kif_i\varphi$ . These operators and the standard ‘knowing that’ operators are interdefinable: in  $Fml_{EL}^{if}$ ,  $K_i\varphi$  can be defined as  $\varphi \wedge Kif_i\varphi$ , and  $CK\varphi$  as  $\varphi \wedge CKif\varphi$ ; and in  $Fml_{EL}$ ,  $Kif_i\varphi$  can be defined as  $K_i\varphi \vee K_i\neg\varphi$  and  $CKif\varphi$  and  $CK\varphi \vee CK\neg\varphi$ . Alternative possibilities of defining the ‘knowing whether’ operators for shared and common knowledge are explored in [53], but they turn out to be equivalent when the considered operators are at least KT, as they are here.

The semantics of the ‘knowing whether’ operators  $Kif_i$  and  $CKif$  are as follows. Given a Kripke model  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$ ,

$$\begin{aligned} (M, w) \models Kif_i\varphi & \text{ iff } (M, w) \text{ and } (M, w') \text{ agree on } \varphi \text{ for every } w' \text{ s.t. } wR_iw'; \\ (M, w) \models CKif\varphi & \text{ iff } (M, w) \text{ and } (M, w') \text{ agree on } \varphi \text{ for every } w' \text{ s.t. } wR_{Agt}w'; \end{aligned}$$

where agreement on a formula is defined naturally:  $(M, w)$  and  $(M, w')$  agree on  $\varphi$  when either  $(M, w) \models \varphi$  and  $(M, w') \models \varphi$ , or  $(M, w) \not\models \varphi$  and  $(M, w') \not\models \varphi$ .

One way to axiomatize logics of ‘knowing whether’ is to simply take an axiomatization of the corresponding logic of ‘knowing that’ and replace all instances of  $K_i\varphi$  with  $\varphi \wedge Kif_i\varphi$ , and all instances of  $CK\varphi$  with  $\varphi \wedge CKif\varphi$ . A more direct axiomatization of the logic without common knowledge is given in [54], with the definition and axiomatization of common knowledge whether being left as an open question. The matter is discussed and an axiomatization for S5 given in [53]; we recall it in Table 2.8. We will be discussing this further and proposing a different axiomatization in Chapter 3.

### 2.2.6 Bisimulations

In order to compare the EL-O framework to standard DEL we will be making use of bisimulation properties. A *bisimulation* [104] between two Kripke models  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  and  $M' = \langle W', R'_{Agt}, \{R'_i\}_{i \in Agt}, V' \rangle$  is a relation  $Z \subseteq W \times W'$  such that for all  $w \in W$  and  $w' \in W'$  such that  $wZw'$ :  $V(w) = V'(w')$  (atomic); if  $wR_{Agt}u$  for some  $u \in W$  then  $uZu'$  and  $w'R'_{Agt}u'$  for some  $u' \in W'$ , and if  $wR_iu$  for some  $u \in W$  then  $uZu'$  and  $w'R'_iu'$  for some  $u' \in W'$  (forth); if  $w'R'_{Agt}u'$  for some  $u' \in W'$  then  $uZu'$  and  $wR_{Agt}u$  for some  $u \in W$ .

---

From *GFP'*:

- |  |                                      |
|--|--------------------------------------|
| 1. $\varphi \rightarrow EK(\varphi \wedge \psi)$                 | hypothesis                           |
| 2. $CK(\varphi \wedge \psi \rightarrow EK(\varphi \wedge \psi))$ | from (1) and <i>RN</i> ( <i>CK</i> ) |
| 3. $EK(\varphi \wedge \psi) \rightarrow CK(\varphi \wedge \psi)$ | from (2) and <i>GFP</i>              |
| 4. $\varphi \rightarrow CK(\varphi \wedge \psi)$                 | from (1) and (3)                     |

<i>CPC</i>	Axiomatics of classical propositional calculus
<i>Con(Kif<sub>i</sub>)</i>	$Kif_i(\varphi \rightarrow \psi) \wedge Kif_i(\neg\psi \rightarrow \varphi) \rightarrow Kif_i\varphi$
<i>Dis(Kif<sub>i</sub>)</i>	$Kif_i\varphi \rightarrow Kif_i(\varphi \rightarrow \psi) \vee Kif_i(\neg\varphi \rightarrow \chi)$
<i>T(Kif<sub>i</sub>)</i>	$Kif_i\varphi \wedge Kif_i(\varphi \rightarrow \psi) \wedge \varphi \rightarrow Kif_i\psi$
<i>5(Kif<sub>i</sub>)</i>	$\neg Kif_i\varphi \rightarrow Kif_i\neg Kif_i\varphi$
$\leftrightarrow$ ( <i>Kif<sub>i</sub></i> )	$Kif_i\varphi \leftrightarrow Kif_{\neg}\varphi$
<i>RN(Kif<sub>i</sub>)</i>	from $\varphi$ , infer $Kif_i\varphi$
<i>RE(Kif<sub>i</sub>)</i>	from $\varphi \leftrightarrow \psi$ , infer $Kif_i\varphi \leftrightarrow Kif_i\psi$
<i>Con(CKif)</i>	$CKif(\varphi \rightarrow \psi) \wedge CKif(\neg\psi \rightarrow \varphi) \rightarrow CKif\varphi$
<i>Dis(CKif)</i>	$CKif\varphi \rightarrow CKif(\varphi \rightarrow \psi) \vee CKif(\neg\varphi \rightarrow \chi)$
<i>T(CKif)</i>	$CKif\varphi \wedge CKif(\varphi \rightarrow \psi) \wedge \varphi \rightarrow CKif\psi$
<i>Mix(CKif)</i>	$CKif\varphi \rightarrow EKif\varphi \wedge EKifCKif\varphi$
<i>Ind(CKif)</i>	$CKif(\varphi \rightarrow EKif\varphi) \rightarrow (\varphi \rightarrow CKif\varphi)$
<i>RN(CKif)</i>	from $\varphi$ , infer $CKif\varphi$
<i>RE(CKif)</i>	from $\varphi \leftrightarrow \psi$ , infer $CKif\varphi \leftrightarrow CKif\psi$

Table 2.8: The axiomatization of the logic S5 with ‘knowing whether’ of [53].

$W$ , and if  $w'R_iu'$  for some  $u' \in W'$  then  $uZu'$  and  $wR_iu$  for some  $u \in W$  (back). Two pointed models  $(M, w)$  and  $(M', w')$  are *bisimilar* if there exists a bisimulation  $Z$  between  $M$  and  $M'$  such that  $wZw'$ . In that case, we have  $(M, w) \models \varphi$  iff  $(M', w') \models \varphi$  for every formula  $\varphi \in Fml_{EL}$  [24].

## 2.3 Classical and DEL planning

We now define classical planning tasks [57] and more generally planning tasks over a vocabulary  $\mathbb{V}$ . Actions have preconditions and conditional effects, and are assumed to be deterministic. We define solvability of planning tasks via sequential plans, in which actions are performed one at a time, and via parallel plans, in which several actions might be performed simultaneously at each step. We finish the section by describing the dynamics of DEL planning.

### 2.3.1 Action descriptions and planning tasks

Just as for the definition of boolean formulas in Section 2.1.1, we consider an abstract vocabulary  $\mathbb{V}$  with which actions are described. An *action description* over  $\mathbb{V}$  (or *action* for short) is a pair  $\mathbf{a} = \langle pre(\mathbf{a}), eff(\mathbf{a}) \rangle$  where  $pre(\mathbf{a}) \in \mathcal{L}_{bool}(\mathbb{V})$  and  $eff(\mathbf{a}) \subseteq \mathcal{L}_{bool}(\mathbb{V}) \times 2^{\mathbb{V}} \times 2^{\mathbb{V}}$ . The formula  $pre(\mathbf{a})$  is the precondition of  $\mathbf{a}$ , describing when the action may be applied, and  $eff(\mathbf{a})$  are the conditional effects of  $\mathbf{a}$ , describing which atomic formulas the action may add or remove from the current state under additional conditions. For each conditional effect

$$ce = \langle cnd(ce), ceff^+(ce), ceff^-(ce) \rangle$$

in  $eff(\mathbf{a})$ ,  $cnd(ce)$  is the condition of  $ce$ ,  $ceff^+(ce)$  are the added and  $ceff^-(ce)$  are the deleted atomic formulas. The vocabulary and the length of an action

description are defined by:

$$\begin{aligned} \text{Voc}(\mathbf{a}) &= \text{Voc}(\text{pre}(\mathbf{a})) \cup \bigcup_{ce \in \text{eff}(\mathbf{a})} \left( \text{Voc}(\text{cnd}(ce)) \cup \text{ceff}^+(ce) \cup \text{ceff}^-(ce) \right); \\ \ell(\mathbf{a}) &= \ell(\text{pre}(\mathbf{a})) + \sum_{ce \in \text{eff}(\mathbf{a})} \left( \ell(\text{cnd}(ce)) + \left( \sum_{p \in \text{ceff}^+(ce) \cup \text{ceff}^-(ce)} \ell(p) \right) \right). \end{aligned}$$

Remember that each  $p \in \mathbb{V}$  has its length  $\ell(p) \geq 1$ .

**Example 2.2** (Lights on). The action of an agent  $i$  flipping a light switch in a room can be represented as:

$$\begin{aligned} \text{flip}_i &= \langle \text{In}_i, \{ \langle \text{On}, \emptyset, \{ \text{On} \} \rangle \\ &\quad \langle \neg \text{On}, \{ \text{On} \}, \emptyset \rangle \rangle \end{aligned}$$

where  $\text{In}_i$  is a propositional variable that is true when the agent is in the room, and  $\text{On}$  is a propositional variable representing the light being on. That is, the action  $\text{flip}_i$  has as a precondition that agent  $i$  is in the room, and has two conditional effects: first, if the light is on, then  $\text{On}$  should become false, hence we have  $\{ \text{On} \}$  as a negative effect and no positive effects. Second, if the light is off (i.e.  $\text{On}$  is false), then  $\text{On}$  should become true, hence in that case we have  $\{ \text{On} \}$  as a positive effect and no negative effects.

A *planning task over*  $\mathbb{V}$  is a triple  $\mathcal{P} = \langle \text{Act}, s_0, \text{Goal} \rangle$  where  $\text{Act}$  is a set of action descriptions,  $s_0 \in 2^{\mathbb{V}}$  is the finite initial state, and  $\text{Goal} \in \mathcal{L}_{\text{bool}}(\mathbb{V})$  is the goal formula. Its vocabulary and length are:

$$\begin{aligned} \text{Voc}(\mathcal{P}) &= \left( \bigcup_{\mathbf{a} \in \text{Act}} \text{Voc}(\mathbf{a}) \right) \cup s_0 \cup \text{Voc}(\text{Goal}); \\ \ell(\mathcal{P}) &= |s_0| + \ell(\text{Goal}) + \sum_{\mathbf{a} \in \text{Act}} \ell(\mathbf{a}). \end{aligned}$$

### 2.3.2 Semantics and solvability of planning tasks

An action  $\mathbf{a}$  is interpreted in terms of a deterministic partial function  $\tau_{\mathbf{a}}$  on states. Given some set of actions  $\text{Act}$  and corresponding action-interpreting functions  $\tau = \{ \tau_{\mathbf{a}} : \mathbf{a} \in \text{Act} \}$ , we say that the state  $s$  is *reachable from the state*  $s_0$  *via*  $\tau$  *by a sequential plan* if there exists an integer  $m \geq 0$ , a sequence of states  $s_0, \dots, s_m$  from  $2^{\mathbb{V}}$ , and a sequence of actions  $\mathbf{a}_1, \dots, \mathbf{a}_m$  from  $\mathcal{R}$  such that  $s_0 = s_0$ ,  $s = s_m$ , and for every  $k$  such that  $1 \leq k \leq m$ ,  $\tau_{\mathbf{a}_k}(s_{k-1})$  is defined and  $\tau_{\mathbf{a}_k}(s_{k-1}) = s_k$ . Given a such interpretation of actions  $\tau$ , a planning task  $\langle \text{Act}, s_0, \text{Goal} \rangle$  is *solvable by a sequential plan via*  $\tau$  if there exists a state  $s$  satisfying the goal formula that is reachable from the initial state  $s_0$  via  $\tau$  by a sequential plan.

A parallel plan is a sequence of *steps* each of which is a set of actions that are executed simultaneously. A set of actions  $\mathbf{A} = \{ \mathbf{a}_1, \dots, \mathbf{a}_m \}$  determines a partial function  $\tau_{\mathbf{A}}$  on states. Given a set of actions  $\text{Act}$  and corresponding action-interpreting functions  $\tau = \{ \tau_{\mathbf{A}} : \mathbf{A} \subseteq \text{Act} \}$ , a state  $s$  is *reachable from the state*  $s_0$  *via*  $\tau$  *by a parallel plan* if there is a sequence  $\mathbf{A}_1, \dots, \mathbf{A}_m$  of steps

and a sequence of states  $s_0, \dots, s_m$  with  $m \geq 0$  such that  $s = s_m$  and for every  $k$  such that  $1 \leq k \leq m$ ,  $\tau_{A_k}(s_{k-1})$  is defined and  $\tau_{A_k}(s_{k-1}) = s_k$ . Given a such interpretation  $\tau$  of sets of actions in  $Act$ , a planning task  $\langle s_0, Goal \rangle$  is *solvable by a parallel plan via  $\tau$*  if there is at least one state  $s$  satisfying  $Goal$  that is reachable by a parallel plan from  $s_0$  via  $\tau$ ; otherwise it is unsolvable by a parallel plan. Solvability by a sequential plan is the special case where the parallel plan is a sequence of singletons.

### 2.3.3 The case of classical sequential planning

Two conditional effects of an action description  $\mathbf{a}$  should not conflict when  $pre(\mathbf{a})$  and their triggering conditions are jointly satisfiable. In the case of classical planning, that is, planning in CPC, we say that  $\mathbf{a}$  is *classically consistent* if and only if for every  $ce_1, ce_2 \in eff(\mathbf{a})$ , if  $pre(\mathbf{a}) \wedge cnd(ce_1) \wedge cnd(ce_2)$  is CPC satisfiable then  $ceff^+(ce_1)$  and  $ceff^-(ce_2)$  are disjoint.

**Example 2.3.** For an example of inconsistency consider the following description of the action  $swap_{p,q}$  swapping the truth values of  $p$  and  $q$ . Suppose this action can always be executed, so its precondition is  $pre(swaps_{p,q}) = \top$ , and suppose its conditional effects are naively described by  $eff(swaps_{p,q}) = \{\langle p, \{q\}, \{p\} \rangle, \langle q, \{p\}, \{q\} \rangle\}$ . These two conditional effects conflict because  $\top \wedge p \wedge q$  is CPC-satisfiable. Observe that we can make the description classically consistent by replacing the precondition by  $\neg(p \leftrightarrow q)$ : the action is executable only if the truth values of  $p$  and  $q$  differ.

A classically consistent action description determines a partial function  $\tau_{\mathbf{a}}^{CPC}$  as follows: given a state  $s \subseteq \mathbb{V}$ ,  $\tau_{\mathbf{a}}^{CPC}$  is defined at  $s$  if  $s \models^{CPC} pre(\mathbf{a})$ . In that case we say that  $\mathbf{a}$  is executable at  $s$  and stipulate:

$$\tau_{\mathbf{a}}^{CPC}(s) = \left( s \setminus \bigcup_{\substack{ce \in eff(\mathbf{a}), \\ s \models cnd(ce)}} (ceff^-(ce)) \right) \cup \bigcup_{\substack{ce \in eff(\mathbf{a}), \\ s \models cnd(ce)}} ceff^+(ce).$$

That is, if the precondition of  $\mathbf{a}$  is satisfied then  $\mathbf{a}$  removes negative effects of all those conditional effects  $ce$  that ‘fire’, i.e., whose triggering conditions are satisfied, and it adds the positive effects of those  $ce$ . As  $\mathbf{a}$ ’s description is consistent it does not matter in which order we apply negative and positive effects.

**Example 2.4** (Lights on, ctd.). Let  $s = \{In_1, On\}$ , and  $flip_1$  be the action of agent 1 flipping a light switch described above. Then  $flip_1$  is executable at  $s$ ,  $\tau_{flip_1}(s) = \{In_1\}$ , and  $\tau_{flip_1}(\tau_{flip_1}(s)) = \{In_1, On\}$ . If  $s = \{On\}$ , the agent is not in the room, and cannot flip the switch: the precondition of  $flip_1$  is not satisfied. Hence  $flip_1$  is not executable at the state  $\{On\}$ .

A planning task is *classically solvable by a sequential plan* if a state satisfying the goal formula is reachable from the initial state by a sequential plan via  $\{\tau_{\mathbf{a}}^{CPC} : \mathbf{a} \in Act\}$ , i.e., via the CPC-interpretation of the actions in  $Act$ . It is known that classical solvability of a planning task is a PSPACE-complete reasoning problem [33].



### 2.3.4 The case of classical parallel planning

Parallel planning comes with a few more conditions for consistency of actions. We follow the notion of interference from [112], which guarantees that non-interfering actions can be interleaved in any order: in reality, “parallel” actions are rarely executed exactly at the same time, and the effects of these actions should not be changed depending on which of them happens to start first. Intuitively, in any parallel plan, no effect of an action should be destroyed by an effect of another action executed in parallel, and no precondition of an action should be destroyed by an effect of another action executed in parallel. Let us define these two consistency criteria formally.

Actions  $\mathbf{a}_1$  and  $\mathbf{a}_2$  that are executable at  $s$  *have no contradictory effects at  $s$*  if: for every  $ce_1 \in \text{eff}(\mathbf{a}_1)$  and  $ce_2 \in \text{eff}(\mathbf{a}_2)$ , if  $s \models \text{cnd}(ce_1) \wedge \text{cnd}(ce_2)$  then  $\text{ceff}^+(ce_1) \cap \text{ceff}^-(ce_2) = \emptyset$ . It follows that the description of the individual action  $\mathbf{a}$  is consistent iff  $\mathbf{a}$  has no contradictory effects with itself in every  $s$  such that  $s \models \text{pre}(\mathbf{a})$ .

We say that two different actions  $\mathbf{a}_1$  and  $\mathbf{a}_2$  that are executable at  $s$  *have no cross-interaction at  $s$*  if the following hold:

1.  $s$  and  $\tau_{\mathbf{a}_1}(s)$  agree on  $\text{pre}(\mathbf{a}_2)$  and on the condition  $\text{cnd}(ce_2)$  of every conditional effect  $ce_2 \in \text{eff}(\mathbf{a}_2)$ ;
2.  $s$  and  $\tau_{\mathbf{a}_2}(s)$  agree on  $\text{pre}(\mathbf{a}_1)$  and on the condition  $\text{cnd}(ce_1)$  of every conditional effect  $ce_1 \in \text{eff}(\mathbf{a}_1)$ .

**Example 2.5** (Lights on, ctd.). Consider instead of  $\text{flip}_i$  the actions  $\text{switchOn}_i$  and  $\text{switchOff}_i$  defined as follows:

$$\begin{aligned}\text{switchOn}_i &= \langle \top, \{\langle \top, \{\text{On}\}, \emptyset \rangle\} \rangle \\ \text{switchOff}_i &= \langle \top, \{\langle \top, \emptyset, \{\text{On}\} \rangle\} \rangle\end{aligned}$$

That is, both actions have no preconditions and one effect with no conditions.  $\text{switchOn}_i$  simply ensures that the light is on after the action is executed, and  $\text{switchOff}_i$  ensures that the light is off after the action is executed. Both actions are executable in every state, but have contradictory effects at every state: if they are executed at the same time, the light will be turned on and off at the same time.

If we add a precondition on the conditional effects of both actions so that these effects only ‘fire’ when the light is not in the right position, the actions become:

$$\begin{aligned}\text{switchOn}'_i &= \langle \top, \{\langle \neg \text{On}, \{\text{On}\}, \emptyset \rangle\} \rangle \\ \text{switchOff}'_i &= \langle \top, \{\langle \text{On}, \emptyset, \{\text{On}\} \rangle\} \rangle\end{aligned}$$

These actions no longer have contradictory effects, as the preconditions of both effects are not simultaneously satisfiable. However, they have cross-interactions at every state: if the light is on at  $s$  ( $\text{On} \in s$ ), then  $s$  and  $\tau_{\text{switchOff}'_i}(s)$  disagree on the condition of the conditional effect of  $\text{switchOn}'_i$ , and if the light is off at  $s$  ( $\text{On} \notin s$ ), then  $s$  and  $\tau_{\text{switchOn}'_i}(s)$  disagree on the conditional effect of  $\text{switchOff}'_i$ . Hence the order in which these actions are executed would change which conditional effects would fire for each action.

Putting things together, we say that a set of actions  $A = \{a_1, \dots, a_m\}$  such that all  $a_i$  are executable at  $s$  is *consistent in  $s$*  if for every  $a_i, a_j \in A$  such that  $a_i \neq a_j$ ,

1.  $a_i$  and  $a_j$  have no contradictory effects in  $s$ ;
2.  $a_i$  and  $a_j$  have no cross-interaction in  $s$ .

In classical planning, the function  $\tau_A^{\text{CPC}}$  is defined at  $s$  if every  $a_i \in A$  is executable at  $s$  and  $A$  is consistent in  $s$ . When  $\tau_A^{\text{CPC}}$  is defined at  $s$  then:

$$\tau_A^{\text{CPC}}(s) = \left( s \setminus \bigcup_{\substack{a \in A, ce \in \text{eff}(a), \\ s \models \text{cnd}(ce)}} (\text{ceff}^-(ce)) \right) \cup \bigcup_{\substack{a \in A, ce \in \text{eff}(a), \\ s \models \text{cnd}(ce)}} \text{ceff}^+(ce).$$

When  $\{a_1, a_2\}$  is consistent in  $s$  then the actions can be interleaved arbitrarily: we have  $\tau_{\{a_1, a_2\}}(s) = \tau_{a_2}(\tau_{a_1}(s)) = \tau_{a_1}(\tau_{a_2}(s))$ .

**Example 2.6.** Consider the actions of agents entering and exiting a room:

$$\begin{aligned} \text{enter}_i &= \langle \neg \text{In}_i, \{\langle \top, \{\text{In}_i\}, \emptyset \rangle\} \rangle \\ \text{exit}_i &= \langle \text{In}_i, \{\langle \top, \emptyset, \{\text{In}_i\}\} \rangle \rangle \end{aligned}$$

Consider a setting with two agents 1 and 2, and the state  $s = \{\text{In}_1\}$ . Then  $\tau_{\{\text{exit}_1, \text{enter}_2\}}(s) = \{\text{In}_2\}$ .

A planning task is *classically solvable by a parallel plan* if a state satisfying the goal formula is reachable from the initial state by a parallel plan via  $\{\tau_A^{\text{CPC}} : A \subseteq \text{Act}\}$ . Note that due to the consistency requirements, a planning task is classically solvable by a parallel plan iff it is classically solvable by a sequential plan.

### 2.3.5 DEL planning: event models and product updates

Dynamic Epistemic Logic (DEL) is the dynamic extension of the epistemic logics presented in Section 2.1 and 2.2. DEL-based planning has been studied extensively in the literature in the last ten years [26, 9, 29, 37, 28, 27]. Just like classical planning tasks, DEL planning tasks consist of an initial situation, some actions, and a goal formula. The initial situation is a standard Kripke model with possible worlds; all that is left is to describe DEL actions.

An S5 event model is a tuple  $\text{Evt} = \langle W^{\text{Evt}}, R_{\text{Agt}}^{\text{Evt}}, \{R_i^{\text{Evt}}\}_{i \in \text{Agt}}, \text{pre}^{\text{Evt}}, \text{post}^{\text{Evt}} \rangle$  where  $W^{\text{Evt}}$  is a finite set of *events*;  $R_{\text{Agt}}^{\text{Evt}}$  and the  $R_i^{\text{Evt}}$  are equivalence relations on  $W^{\text{Evt}}$  such that every  $R_i^{\text{Evt}}$  is a subset of  $R_{\text{Agt}}^{\text{Evt}}$ ;  $\text{pre}^{\text{Evt}} : W^{\text{Evt}} \rightarrow \text{Fml}_{\text{EL}}$  maps events to their preconditions; and  $\text{post}^{\text{Evt}} : W^{\text{Evt}} \rightarrow (\text{Prop} \rightarrow \text{Fml}_{\text{EL}})$  maps events to partial functions such that for every  $e \in W^{\text{Evt}}$ ,  $\text{post}^{\text{Evt}}(e)$  is undefined almost everywhere. Just like with standard Kripke models, the properties of  $R_i^{\text{Evt}}$  and  $R_{\text{Agt}}^{\text{Evt}}$  can be adjusted to correspond to different types of knowledge and belief. A *pointed event model* is a pair  $(\text{Evt}, e)$  where  $e \in W^{\text{Evt}}$  is the designated event (or the actual event). A *multipointed event model* is a pair  $(\text{Evt}, E)$  where  $E \subseteq W^{\text{Evt}}$  is such that for any two distinct events  $e, e' \in E$ , the preconditions of  $e$  and  $e'$  are incompatible, that is,  $\text{pre}^{\text{Evt}}(e) \wedge \text{pre}^{\text{Evt}}(e')$  is unsatisfiable.

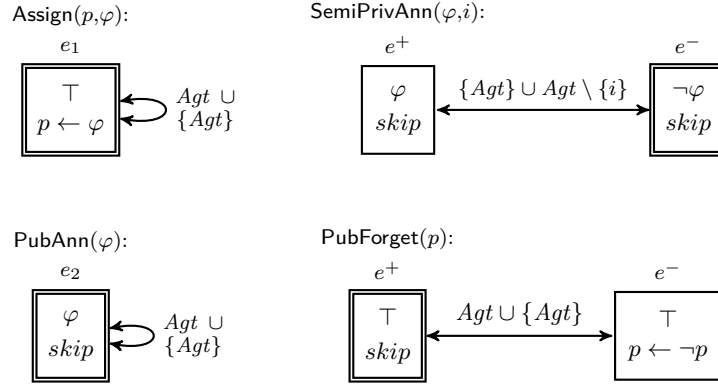


Figure 2.1: DEL event models for the public assignment of  $\varphi$  to  $p$  (top left), the truthful public announcement of  $\varphi$  (bottom left), the truthful semi-private announcement of  $\neg\varphi$  to agent  $i$  (top right), and the public forgetting of  $p$  (bottom right).

Here are some examples and their representations as graphs, in which nodes are events, labelled arrows represent the equivalence relations, the actual event is denoted via a double outline, and pre- and postconditions are given respectively in the top and bottom line in the nodes. An undefined postcondition is denoted by *skip*. We will often omit transitive and reflexive arrows for clarity.

1. The public assignment of  $\varphi$  to  $p$ , i.e., the event where  $p$  publicly gets the truth value of  $\varphi$ , is modeled by an event model  $\text{Assign}(p, \varphi)$  with a single point  $e_1$  whose precondition is  $\top$ , with total relations, i.e., such that  $R_{Agt}^{\text{Evt}} = R_i^{\text{Evt}} = \{\langle e_1, e_1 \rangle\}$ , and with the postcondition function  $post^{\text{Evt}}(e_1)$  such that  $post^{\text{Evt}}(e_1)(p) = \varphi$  and undefined for all  $q \neq p$ . This event model is shown on the top left-hand side in Figure 2.1.
2. The truthful public announcement of a formula  $\varphi$  is modeled by an event model  $\text{PubAnn}(\varphi)$ , represented on the bottom right-hand side of Figure 2.1, with a single point  $e_2$  whose precondition is  $\varphi$ , with total relations and with an undefined postcondition function.
3. The truthful semi-private announcement whether  $\varphi$  to agent  $i$  is when  $i$  learns whether  $p$  and the other agents only learn that  $i$  learns whether  $p$  without learning whether  $p$  themselves. This is modeled by an event model  $\text{SemiPrivAnn}(\varphi, i)$  with two points  $e^+$  and  $e^-$  where  $pre^{\text{Evt}}(e^+) = \varphi$  and  $pre^{\text{Evt}}(e^-) = \neg\varphi$ , with an undefined postcondition function, and with  $R_{Agt}^{\text{Evt}} = R_j^{\text{Evt}} = \{e^+, e^-\} \times \{e^+, e^-\}$  for every  $j \neq i$  and  $R_i^{\text{Evt}} = \{\langle e^+, e^+ \rangle, \langle e^-, e^- \rangle\}$ . When we want to model that  $i$  learns that  $\varphi$  then  $e^+$  is the designated event; otherwise it is  $e^-$ , as is the case in the top right-hand model of Figure 2.1.
4. The public forgetting of a propositional variable  $p$  by all agents [50] can be captured by an event model  $\text{PubForget}(p)$  with two points  $e^+$  and  $e^-$  where  $pre^{\text{Evt}}(e^+) = pre^{\text{Evt}}(e^-) = \top$ , where  $R_{Agt}^{\text{Evt}} = R_i^{\text{Evt}} = \{e^+, e^-\} \times \{e^+, e^-\}$

for every  $i$ , and where  $post^{Evt}$  is such that  $post^{Evt}(e^+)(p) = skip$ , and  $post^{Evt}(e^-)(p) = \neg p$ . This event model is represented on the bottom right-hand side in Figure 2.1.

Given a Kripke model  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  as defined in Section 2.2.1 and an event model  $Evt$ , the *product update* of  $M$  by  $Evt$  is the Kripke model  $M \otimes Evt = \langle W', R'_{Agt}, \{R'_i\}_{i \in Agt}, V' \rangle$  with

- $W' = \{(s, e) : s \in W, e \in W^{Evt}, \text{ and } (M, s) \models pre^{Evt}(e)\}$ ;
- $(s, e)R'_{Agt}(t, f)$  iff  $sR_{Agt}t$  and  $eR_{Agt}^{Evt}f$ ;
- $(s, e)R'_i(t, f)$  iff  $sR_i t$  and  $eR_i^{Evt}f$ , for every  $i \in Agt$ ;
- $V'((s, e)) = \{p : post^{Evt}(e)(p) \text{ is undefined and } (M, s) \models p\} \cup \{p : post^{Evt}(e)(p) \text{ is defined and } (M, s) \models post^{Evt}(e)(p)\}$ .

The product update of a pointed Kripke model  $(M, s)$  with a pointed event model  $(Evt, e)$  is the pointed Kripke model  $(M \otimes Evt, (s, e))$ , defined only when  $(M, s) \models pre^{Evt}(e)$ . The product update of a pointed model  $(M, s)$  with a multipointed event model  $(Evt, E)$  is the pointed model  $(M \otimes Evt, (s, e))$  where  $e \in E$  is such that  $(M, s) \models pre^{Evt}(e)$ , defined when such an  $e$  exists in  $E$ . In that case the hypothesis in the definition of a multipointed event model that the preconditions of different events are incompatible ensures that it is unique.

## Chapter 3

# EL-O: Epistemic Logic of Observation

### La logique épistémique de l’observation : EL-O

La logique EL-O est un fragment de la logique DEL-PAO qui fut introduite dans [70]. La thèse [99] présente une vue d’ensemble du travail précédemment effectué sur EL-O. Cette logique est fondée sur un langage simple : les formules sont des combinaisons booléennes d’atomes d’observation, ces derniers étant définis de la manière suivante :

$$ATM \ni \alpha ::= p \mid S_i \alpha \mid JS \alpha$$

où  $p$  représente une variable propositionnelle,  $S_i \alpha$  se lit “l’agent  $i$  voit  $\alpha$ ” (c’est à dire que l’agent  $i$  sait si  $\alpha$  est vrai ou non) et  $JS \alpha$  se lit “les agents voient conjointement  $\alpha$ ” (c’est à dire qu’il y a connaissance commune de la valeur de vérité de  $\alpha$ ).

Dans ce chapitre nous présentons une version perfectionnée de EL-O reposant sur une nouvelle sémantique : les états sont des ensembles quelconques d’atomes d’observation, et les principes d’introspection sont simulés dans la sémantique. Cela nous permet d’obtenir une propriété de modèles finis : toute formule satisfiable l’est dans un état fini. Nous montrons aussi que le problème de satisfiabilité EL-O est NP-complet, et donnons une axiomatisation de cette logique. Nous étudions ensuite la relation entre EL-O et la logique épistémique standard et montrons que EL-O est un fragment de la logique  $S5 \setminus GFP$ . Nous considérons enfin plusieurs possibilités pour étendre EL-O. Tout d’abord, nous montrons qu’il est naturel d’ajouter des constantes au langage afin de généraliser la notion de ‘savoir la valeur de vérité de’ à celle de ‘savoir la valeur de’ (un code, un numéro de téléphone, etc.). En revanche, il est moins simple d’étendre l’opérateur de vision jointe  $JS$  à des groupes arbitraires d’agents, et nous discutons brièvement des problèmes que cela soulèverait.

---

The logic EL-O is the static fragment of the logic DEL-PAO, introduced in [70]. An overview of previous work on EL-O can be found in [99]. In this chapter we present a more polished version of EL-O with updated semantics and study some of its properties. In particular, we give a finite model property and show that

checking EL-O satisfiability is NP-complete. We then investigate the relation between EL-O and standard epistemic logics and show that EL-O is a fragment of the logic  $S5 \setminus GFP$ . We finally discuss a few possibilities of extending EL-O. Mainly, we show that adding constants to the language and seeing the operator  $S_i$  as a ‘knowing what’ or ‘knowing the value of’ operator is easily done. Less straightforward is generalizing the joint vision operator  $JS$  to arbitrary groups of agents, and we end the chapter by discussing the issues this would bring up.

This chapter is mainly based on the following publications:

- Cooper, M. C., Herzig, A., Maffre, F., Maris, F., Perrotin, E., & Régnier, P. (2020). *A lightweight epistemic logic and its application to planning*. Artificial Intelligence, 103437.
- Herzig, A., Maris, F., & Perrotin, E. (2021). *A dynamic epistemic logic with finite iteration and parallel composition*. Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR 2021) (Vol. 18, No. 1, pp. 676-680).

## Contents

---

<b>3.1 Syntax and semantics of EL-O</b>	<b>31</b>
3.1.1 Observability atoms and introspection	32
3.1.2 Atomic consequence and introspective states	33
3.1.3 Language of EL-O	33
3.1.4 Semantics of EL-O	34
<b>3.2 Some properties of EL-O</b>	<b>35</b>
3.2.1 Finite model property	35
3.2.2 Complexity of EL-O satisfiability	36
3.2.3 Axiomatization of EL-O validities	36
<b>3.3 EL-O as a fragment of epistemic logic <math>S5 \setminus GFP</math></b>	<b>37</b>
3.3.1 The canonical model for EL-O	38
3.3.2 Completeness of the EL-O fragment of $Fml_{EL}^{if}$ w.r.t. $S5 \setminus GFP$ validity	39
3.3.3 Axiomatization of the $M^{EL-O}$ validities in $Fml_{EL}$	39
<b>3.4 Adding constants: EL-OC</b>	<b>40</b>
3.4.1 The language of EL-OC	41
3.4.2 Semantics and axiomatization	42
3.4.3 Applications	42
<b>3.5 Discussion: joint vision for arbitrary groups of agents</b>	<b>43</b>
<b>3.6 Conclusion</b>	<b>44</b>

---

## 3.1 Syntax and semantics of EL-O

In this section we describe the language and semantics of EL-O. While the observability atoms and introspection principles described in Sections 3.1.1 and 3.1.2 have remained the core of the EL-O framework, the actual language and semantics have known quite a bit of variation over time, with forays into possible

world semantics and infinite valuations as states [99]. Here worlds are arbitrary sets of atoms, and introspection principles are simulated via the nonstandard semantics described in Section 3.1.4. This new presentation will allow us to work with finite states, which many of the results in this thesis rely on.

### 3.1.1 Observability atoms and introspection

We now define the vocabulary we will be using in our epistemic logic. Let  $Prop$  be a countable set of *propositional variables* and let  $Agt$  be a finite set of *agents*. The set of *observability operators* is  $OBS = \{S_i : i \in Agt\} \cup \{JS\}$ , where  $S_i$  stands for individual observability of agent  $i$  and  $JS$  stands for joint observability of all agents. The set of all sequences of observability operators is noted  $OBS^*$  and the set of all non-empty sequences is noted  $OBS^+$ . We use  $\sigma, \sigma'$ , etc. to denote elements of  $OBS^*$  and reserve **nil** for the empty sequence. The length of a sequence of observability operators is defined inductively by:  $\ell(\mathbf{nil}) = 0$ ;  $\ell(S_i \sigma) = \ell(\sigma) + 1$ ; and  $\ell(JS \sigma) = \ell(\sigma) + 1$ .

*Observability atoms*, or atoms for short, are finite sequences of observability operators followed by a propositional variable. The set

$$ATM = \{\sigma p : \sigma \in OBS^*, p \in Prop\}.$$

is the set of all atoms. (This set depends on the set of propositional variables  $Prop$  and on the set of agents  $Agt$ ; we however leave these arguments implicit.) We use the small Greek letters  $\alpha, \alpha', \beta, \dots$  to denote atoms. Here are some examples:  $S_1 p$  reads “1 sees the value of  $p$ ”. Hence 1 knows whether  $p$  is true or false.  $JS S_2 q$  reads “all agents jointly see whether agent 2 sees the value of  $q$ ”. In other words, there is joint attention in the group of all agents concerning 2’s observation of  $q$ : agent 2 may or may not see the value of  $q$ , and in both cases this is jointly observed.  $S_1 S_2 S_3 p$  reads “1 sees whether 2 sees whether 3 sees  $p$ ”. The *length* of an observability atom is  $\ell(\sigma p) = \ell(\sigma) + 1$ . Therefore an atom of length 1 is nothing but a propositional variable. For example,  $\ell(JS S_2 p) = 3$ .

**Example 3.1** (Gossip, ctd.). In the initial state  $s_0^{G_1}$  of the gossip problem introduced in Example 2.1 of Section 2.1.2, each agent only knows their own secret. These secrets can be seen as rumors concerning each agent, which may or may not be true; each agent knows whether the rumor pertaining to themselves is true or not. Therefore  $s_0^{G_1} = \{S_i s_i : i \in Agt\} \cup S$  where  $S$  is some subset of  $\{s_i : i \in Agt\}$ .

Individual introspection is expressed with our ‘knowing whether’ operator as  $S_i S_i \alpha$ . That is, each agent knows whether or not they see  $\alpha$ , for any atom  $\alpha$ . Moreover, in S5 is common knowledge that this introspection takes place: this is expressed by  $JS S_i S_i \alpha$ . Joint introspection is expressed by  $JS JS \alpha$ . We have also stated in Section 2.2.3 that common knowledge is a fixpoint of shared knowledge, which means that joint vision should imply any nesting of individual vision. (We formalize this in Section 3.1.2). We therefore call an atom *introspective* if it contains two consecutive  $S_i$ , or a  $JS$  that is preceded by a non-empty sequence of observability operators. In other words, an atom is introspective if it is of the form  $\sigma S_i S_i \alpha$  for some  $\sigma \in OBS^*$ , or of the form  $\sigma JS \alpha$  for some  $\sigma \in OBS^+$ . The set of all introspective atoms is

$$\begin{aligned} I-ATM &= \{\sigma S_i S_i \alpha : \sigma \in OBS^* \text{ and } \alpha \in ATM\} \cup \\ &\quad \{\sigma JS \alpha : \sigma \in OBS^+ \text{ and } \alpha \in ATM\}. \end{aligned}$$

The set  $I\text{-}ATM$  is infinite and is a proper subset of  $ATM$ . Intuitively, it is the set of atoms that should always be true when considering agents capable of both positive and negative introspection.

### 3.1.2 Atomic consequence and introspective states

We now formalize the fact that joint observation is a fixpoint of shared individual observation. We define a relation of *atomic consequence* between observability atoms as follows:

$$\alpha \Rightarrow \beta \text{ iff } \alpha = \beta, \text{ or } \alpha = JS \alpha' \text{ and } \beta = \sigma \alpha' \text{ for some } \sigma \in OBS^+.$$

If this is the case then we say that  $\alpha$  is a *cause* of  $\beta$  and that  $\beta$  is a *consequence* of  $\alpha$ . For example,  $JS p \Rightarrow S_i p$  and  $JS p \Rightarrow JS S_i p$ . The relation  $\Rightarrow$  is clearly reflexive and transitive. Moreover, the set of introspective atoms is closed under atomic consequence: if  $\alpha \in I\text{-}ATM$  and  $\alpha \Rightarrow \beta$  then  $\beta \in I\text{-}ATM$ .

We let  $\alpha^{\Leftarrow}$  denote the set of causes of  $\alpha$  and  $\alpha^{\Rightarrow}$  the set of its consequences:

$$\alpha^{\Leftarrow} = \{\beta : \beta \Rightarrow \alpha\}, \quad \alpha^{\Rightarrow} = \{\beta : \alpha \Rightarrow \beta\}.$$

Here are some examples:

$$\begin{aligned} (p)^{\Leftarrow} &= \{p\}, & (p)^{\Rightarrow} &= \{p\}, \\ (S_i p)^{\Leftarrow} &= \{S_i p, JS p\}, & (S_i p)^{\Rightarrow} &= \{S_i p\}, \\ (JS p)^{\Leftarrow} &= \{JS p\}, & (JS p)^{\Rightarrow} &= \{\sigma p : \sigma \in OBS^+\}, \\ (S_i S_j p)^{\Leftarrow} &= \{S_i S_j p, JS S_j p, JS p\}, & (S_i S_j p)^{\Rightarrow} &= \{S_i S_j p\}. \end{aligned}$$

Observe that  $\alpha^{\Leftarrow}$  is always finite while  $\alpha^{\Rightarrow}$  is either infinite (namely when  $\alpha$  starts by  $JS$ ) or the singleton  $\{\alpha\}$  (namely when  $\alpha$  is a propositional variable or starts by some  $S_i$ ). When  $\beta \Rightarrow \alpha$  then the length of  $\beta$  is less than or equal to the length of  $\alpha$ . Moreover, the set of causes of  $\alpha$  has at most  $\ell(\alpha)$  elements:  $|\alpha^{\Leftarrow}| \leq \ell(\alpha)$ . It follows that the sum of the lengths of all causes of  $\alpha$  is at most quadratic in the length of  $\alpha$ :

**Proposition 3.1.** *For every  $\alpha$ ,  $\sum_{\beta: \beta \Rightarrow \alpha} \ell(\beta) \leq (\ell(\alpha))^2$ .*

We generalize atomic cause and consequence to states  $s \in 2^{ATM}$  by defining  $s^{\Leftarrow} = \bigcup_{\alpha \in s} \alpha^{\Leftarrow}$  and  $s^{\Rightarrow} = \bigcup_{\alpha \in s} \alpha^{\Rightarrow}$ . When  $s$  contains all its atomic consequences and all introspective atoms then we say that  $s$  is *introspectively closed* or, for short, *introspective*. The set of all introspective states is

$$I\text{-}STATES = \{s^{\Rightarrow} \cup I\text{-}ATM : s \in 2^{ATM}\}.$$

This is an infinite set. Each of its elements is infinite due to infinity of  $I\text{-}ATM$ .

### 3.1.3 Language of EL-O

The language of EL-O is the set  $\mathcal{L}_{\text{bool}}(ATM)$  of boolean formulas built from the set of observability atoms  $ATM$ . Putting things together:

$$\begin{aligned} ATM \ni \alpha &::= p \mid S_i \alpha \mid JS \alpha \\ \mathcal{L}_{\text{bool}}(ATM) \ni \varphi &::= \alpha \mid \neg \varphi \mid \varphi \wedge \varphi \end{aligned}$$



where  $p \in Prop$  and  $i \in Agt$ . We recall that the length of formulas is determined by the definition of length of atoms from Section 3.1.1. For example,  $\ell(JS S_1 p \wedge \neg S_2 q) = 7$ . Note that the vocabulary of formulas such as  $JS S_1 p$  does not contain the ‘sub-atoms’  $p$  and  $S_1 p$ : we have  $Voc(JS S_1 p) = \{JS S_1 p\}$ . Note also that the language  $\mathcal{L}_{bool}(Prop)$  is the fragment of  $\mathcal{L}_{bool}(ATM)$  without observability operators.

**Example 3.2** (Gossip, ctd.). The goal of the original gossip problem is for all agents to know all secrets:  $Goal^{G_1} = \bigwedge_{i,j \in Agt} S_j s_i$ . The goal of obtaining shared knowledge of depth 2 is expressed by  $Goal^{G_2} = \bigwedge_{i,j,r \in Agt} (S_j s_i \wedge S_r S_j s_i)$ , i.e., every agent  $r$  knows that every  $S_j s_i$  is true.

**Example 3.3** (Letter). In the example given in the introduction, there are two agents:  $Agt = \{1, 2\}$ . If both agents read the letter separately, then  $S_1 p \wedge S_2 p$  (where  $p$  represents the contents of the letter) will be true: both agents learn whether or not  $p$  is true. If one agent reads the letter aloud to the other agent, then  $JS p$  will become true. If agent 1 sees agent 2 reading the letter, then  $S_2 p \wedge S_1 S_2 p$  will be true, but if agent 2 manages to secretly read the letter, then  $S_2 p \wedge \neg S_1 S_2 p$  will be true.

### 3.1.4 Semantics of EL-O

We recall that a state is a subset of the vocabulary  $ATM$ . We wish for agents to have full positive and negative introspection, that is, to be aware of what they do and don’t see, and of what is and isn’t jointly seen. We also wish for joint vision to actually be a fixpoint of shared individual observation. For this we will be using the notions described in Sections 3.1.1 and 3.1.2. A way of guaranteeing these properties was proposed in [70] where formulas are interpreted exclusively in the set of introspectively closed states  $I-STATES$  defined in Section 3.1.2, that is, the states that contain all introspective atoms and are closed under atomic consequence. As such states are always infinite, it is not immediately clear how to define model checking (which requires finite states). One way out is to work with ‘sufficiently introspective states’, as done in [99, Chapter 2]. Here we work with finite models and instead interpret formulas in such a way that the desired properties are simulated. The EL-O truth conditions are just as those for CPC given in the previous chapter, except that for atomic formulas we stipulate:

$$s \models \alpha \text{ iff } \alpha \in s^{\Rightarrow} \cup I-ATM.$$

Hence  $\alpha$  is true in  $s$  if and only if  $\alpha$  is introspective or  $\beta \Rightarrow \alpha$  for some  $\beta \in s$ .

**Example 3.4** (Letter, cont.). If one of the agents reads the letter aloud to the other agent, then they get common knowledge of whether or not  $p$  is true: we are in state  $s = \{JS p\}$ . We then also have individual vision of  $p$ :  $s \models S_1 p \wedge S_2 p$ . If instead agent 1 sees agent 2 privately reading the letter without agent 2 noticing this, we will be in state  $s' = \{S_2 p, S_1 S_2 p\}$ . In this state agent 1 does not know the truth value of  $p$ , but agent 2 does:  $s' \models \neg S_1 p \wedge S_2 p$ .

**Example 3.5** (Gossip, ctd.). Let  $s_0^{G_1}$  be the initial state of the gossip problem from Example 3.1. Then  $s_0^{G_1} \models S_i s_i \wedge \bigwedge_{j \neq i} \neg S_i s_j$  for every  $i \in Agt$ .

The EL-O semantics coincides with the CPC semantics for the fragment  $\mathcal{L}_{bool}(Prop)$  of  $\mathcal{L}_{bool}(ATM)$ :

**Proposition 3.2.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(\text{Prop})$ . Then  $s \models \varphi$  iff  $s \models^{\text{CPC}} \varphi$ .*

It will sometimes be useful to consider introspective closures of states. The following proposition can easily be shown by induction on the structure of  $\varphi$ .

**Proposition 3.3.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$  be a formula and  $s \in 2^{\text{ATM}}$  a state. Then  $s \models \varphi$  iff  $s^{\rightarrow} \cup I\text{-ATM} \models \varphi$ .*

A formula  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$  is *EL-O satisfiable* iff  $s \models \varphi$  for some  $s \in 2^{\text{ATM}}$ ; it is *EL-O valid* iff  $s \models \varphi$  for every  $s \in 2^{\text{ATM}}$ . Clearly, an atom  $\alpha$  is *EL-O valid* if and only if it is introspective. Moreover, for atoms  $\alpha, \beta \in \text{ATM}$  we have that  $\alpha \rightarrow \beta$  is *EL-O valid* iff  $\alpha \Rightarrow \beta$  or  $\beta$  is introspective.

**Proposition 3.4.** *Let  $A, B \subseteq \text{ATM}$  be such that  $(\bigwedge_{\alpha \in A} \alpha) \wedge (\bigwedge_{\beta \in B} \neg \beta)$  is *EL-O unsatisfiable*. Then  $B$  contains some introspective  $\beta$ , or there are  $\alpha \in A$ ,  $\beta \in B$  such that  $\alpha \Rightarrow \beta$ .*

*Proof.* We prove the contraposition: suppose  $B$  contains no introspective atom and there are no  $\alpha \in A$  and  $\beta \in B$  such that  $\alpha \Rightarrow \beta$ . We show that the state  $s = A$  satisfies  $(\bigwedge_{\alpha \in A} \alpha) \wedge (\bigwedge_{\beta \in B} \neg \beta)$ . First,  $s$  satisfies every element of  $A$ . Second,  $s$  does not satisfy any  $\beta \in B$ : otherwise  $\beta$  would be introspective, or we would have  $\alpha \Rightarrow \beta$  for some  $\alpha \in A$ .  $\square$

*Remark 3.1.* When  $\text{Agt}$  is a singleton then  $S_i p \wedge \neg JS p$  is satisfiable. While this anomaly could be taken care of by a modification of the semantics, we do not do so for the sake of readability and content ourselves with the observation that the  $JS$  operator is superfluous when there is only one agent.

*Remark 3.2.* It would be interesting to generalise joint observability from  $\text{Agt}$  to arbitrary subsets of  $\text{Agt}$ . For instance, after a gossiping phone call between agents  $i$  and  $j$  all secrets previously known by one of them become jointly observable by the group  $\{i, j\}$ . This however presents some difficulties, which we discuss in Section 3.5.

In the rest of the section we establish finite model property, given an axiomatization, and prove NP-completeness of EL-O satisfiability.

## 3.2 Some properties of EL-O

We study some properties of the logic EL-O, starting with a finite model property which allows us to then establish NP-completeness of the satisfiability problem. We end this section with an axiomatization of EL-O validities.

### 3.2.1 Finite model property

A standard property of CPC is that  $s \models^{\text{CPC}} \varphi$  iff  $s \cap \text{Voc}(\varphi) \models^{\text{CPC}} \varphi$ . This does not hold in EL-O due to the non-standard truth condition for atoms. (It suffices to consider  $s = \{JS p\}$  and  $\varphi = S_i p$  to see this.) Nevertheless, we obtain a finite model property by using a closure of the state  $s$ .

First of all, let us say that two states  $s$  and  $s'$  *agree* on the set of atoms  $A \subseteq \text{ATM}$  when they agree on every element of  $A$ , i.e., when for every  $\alpha \in A$ ,  $s \models \alpha$  iff  $s' \models \alpha$ . This is not enough to guarantee that  $s \cap A = s' \cap A$ . To witness, consider  $s = \{JS p\}$ ,  $s' = \{S_i p\}$ , and  $A = \{S_i p\}$ :  $s$  and  $s'$  agree on  $A$ ,

but  $s \cap A = \emptyset$  while  $s' \cap A = \{S_i p\}$ . However by straightforward induction on the structure of formulas we get:

**Proposition 3.5.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  be a formula and  $s, s' \in 2^{ATM}$  two states agreeing on  $\text{Voc}(\varphi)$ . Then  $s \models \varphi$  iff  $s' \models \varphi$ .*

Combining Propositions 3.3 and 3.5 we obtain the finite model property for EL-O.

**Proposition 3.6.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  be a formula and  $s \in 2^{ATM}$  a state. Then  $s \models \varphi$  iff  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi) \models \varphi$ .*

*Proof.* We know by Proposition 3.3 that  $s \models \varphi$  iff  $s \Rightarrow \cup I-ATM \models \varphi$ . By Proposition 3.5, it now suffices to show that  $s \Rightarrow \cup I-ATM$  and  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi)$  agree on  $\text{Voc}(\varphi)$ . Consider  $\alpha \in \text{Voc}(\varphi)$  such that  $s \Rightarrow \cup I-ATM \models \alpha$ . Then  $\alpha \in s \Rightarrow \cup I-ATM$ , hence  $\alpha \in (s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi)$  and  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi) \models \alpha$ . Consider now  $\alpha \in \text{Voc}(\varphi)$  such that  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi) \models \alpha$ . Then either  $\alpha \in I-ATM$  or there exists  $\beta \in (s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi)$  such that  $\beta \Rightarrow \alpha$ . In the first case, clearly  $s \Rightarrow \cup I-ATM \models \alpha$ . In the second, we have  $\beta \in s \Rightarrow \cup I-ATM$ , hence  $s \Rightarrow \cup I-ATM \models \alpha$ . This ends the proof.  $\square$

### 3.2.2 Complexity of EL-O satisfiability

**Proposition 3.7.** *Let  $s \in 2^{ATM}$  be a state and  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  a formula. Then  $s \models \varphi$  iff  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi) \models^{\text{CPC}} \varphi$ .*

*Proof.* It suffices to prove that  $s \models \varphi$  iff  $s \Rightarrow \cup I-ATM \models^{\text{CPC}} \varphi$ . This is shown by a straightforward induction on the structure of  $\varphi$ .  $\square$

**Proposition 3.8.** *The problem of deciding satisfiability of  $\mathcal{L}_{\text{bool}}(ATM)$  formulas is NP-complete.*

*Proof.* Hardness is the case because EL-O satisfiability and CPC satisfiability coincide for the  $\mathcal{L}_{\text{bool}}(\text{Prop})$  fragment of  $\mathcal{L}_{\text{bool}}(ATM)$  (Proposition 3.2). For membership, guess some subset  $s \in 2^{\text{Voc}(\varphi)}$  and check in polynomial time:

1. that  $s$  contains all its relevant consequences: for every  $\alpha \in \text{Voc}(\varphi)$  and  $\beta \in s$ , check that if  $\beta \Rightarrow \alpha$  then  $\alpha \in s$ ;
2. that  $s$  contains all introspective atoms from  $\text{Voc}(\varphi)$ : for every  $\alpha \in \text{Voc}(\varphi)$ , check that if  $\alpha \in ATM$  then  $\alpha \in s$ ;
3. that  $s \models^{\text{CPC}} \varphi$ .

Checking that  $s \models^{\text{CPC}} \varphi$  is tantamount to checking that  $s \models \varphi$  because: (a) thanks to the first and second check,  $s$  equals  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi)$ , and therefore  $s \models^{\text{CPC}} \varphi$  iff  $(s \Rightarrow \cup I-ATM) \cap \text{Voc}(\varphi) \models^{\text{CPC}} \varphi$ ; (b) by Proposition 3.7, the latter is the case iff  $s \models \varphi$ .  $\square$

### 3.2.3 Axiomatization of EL-O validities

The EL-O validities of our language  $\mathcal{L}_{\text{bool}}(ATM)$  are axiomatized by the schemas of Table 3.1 together with CPC. We prove its completeness via CPC.

$Vis_1$	$S_i S_i \alpha$
$Vis_2$	$JS JS \alpha$
$Vis_3$	$JS S_i S_i \alpha$
$Vis_4$	$JS \alpha \rightarrow S_i \alpha$
$Vis_5$	$JS \alpha \rightarrow JS S_i \alpha$

Table 3.1: Axioms of EL-O

**Proposition 3.9.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  be a formula. Then  $\varphi$  is EL-O valid iff  $\varphi$  is provable in CPC from the axiom schemas  $Vis_1$ – $Vis_5$  of Table 3.1.*

*Proof.* We again take advantage of Proposition 3.7 and show that  $Vis_1$ – $Vis_5$  characterise the set of introspectively closed states  $I\text{-STATES} = \{s \Rightarrow \cup I\text{-ATM} : s \in 2^{ATM}\}$ . The right-to-left direction is clear: each of the five axiom schemas is valid in introspectively closed states. For the left-to-right direction, we show that every  $s$  satisfying  $Vis_1$ – $Vis_5$  is introspectively closed.

For closure under atomic consequence, let  $s \models \alpha$ . The interesting case is when  $\alpha = JS \alpha'$ . Then  $s \models JS S_{i_1} \alpha'$  for any agent  $i_1$  by axiom  $Vis_5$ , and also  $s \models JS S_{i_2} S_{i_1} \alpha'$  for any  $i_1$  and  $i_2$ , and so on: we can generate any  $s \models JS S_{i_m} \dots S_{i_1} \alpha'$  and then, by  $Vis_4$ , we can obtain  $s \models S_{i_m} \dots S_{i_1} \alpha'$ . Moreover, we have  $s \models JS JS \alpha'$  by  $Vis_2$ , and in the same way, we can generate  $s \models S_{i_m} \dots S_{i_1} JS \alpha'$  for any agents  $i_1, \dots, i_m$  and  $i$ . We therefore obtain that  $s$  satisfies every  $\sigma \alpha'$  for any  $\sigma \in OBS^+$ , that is, every atomic consequence of  $\alpha$ .

We use the same technique to show that a state  $s$  satisfying  $Vis_1$ – $Vis_5$  satisfies every  $\sigma S_i S_i \alpha$  for  $\sigma \in OBS^*$  and every  $\sigma JS \alpha$  for  $\sigma \in OBS^+$ : we obtain the first with  $Vis_1$  (for  $\sigma$  empty) and  $Vis_3$ ,  $Vis_4$ , and  $Vis_5$  (for  $\sigma$  non-empty), and the second with  $Vis_2$ ,  $Vis_4$ , and  $Vis_5$ .  $\square$

*Remark 3.3.* We do not require states to satisfy the infinitary constraint “if  $\sigma \alpha \in s$  for every  $\sigma \in \{S_i : i \in Agt\}^+$  then  $JS \alpha \in s$ ”, which is the EL-O counterpart of the induction principle for common knowledge. We will discuss in Chapter 3 how to integrate this additional constraint.

### 3.3 EL-O as a fragment of epistemic logic $S5 \setminus GFP$

We now situate our lightweight epistemic logic w.r.t. the standard epistemic logic  $S5 \setminus GFP$  with ‘knowing whether’ operators  $Kif_i$  and  $CKif$ . Recall that  $S5 \setminus GFP$ , described in Chapter 2, is the logic of  $S5$  individual and common knowledge which does not obey the induction principle for common knowledge. We are going to identify the operators  $S_i$  with  $Kif_i$  and  $JS$  with  $CKif$ . Then the language  $\mathcal{L}_{\text{bool}}(ATM)$  of EL-O becomes a fragment of  $Fml_{\text{EL}}^{\text{if}}$ . When we make this identification we are going to talk about the EL-O *fragment* of  $Fml_{\text{EL}}^{\text{if}}$ .

Our main result is that for the EL-O fragment of  $Fml_{\text{EL}}^{\text{if}}$ , the axioms of Table 3.1 are sound and complete w.r.t. validity in  $S5 \setminus GFP$  models (Proposition 3.12). It follows by Proposition 3.8 that our lightweight epistemic logic is a fragment of standard epistemic logic with an NP-complete satisfiability problem. This is an important result because satisfiability of  $Fml_{\text{EL}}^{\text{if}}$  formulas is PSPACE-complete

as soon as there are two agents, and this is the case even without the common knowledge operator [24].

### 3.3.1 The canonical model for EL-O

Let us establish that every EL-O valuation  $s$  can be identified with a particular pointed  $S5 \setminus GFP$  Kripke model  $(M^{\text{EL-O}}, s)$ . This actually calls back to previous iterations of EL-O semantics, as described in [99, Chapter 2]. The first thing we do is define how the states' observability information determines equivalence relations between states. For every  $s, s' \in 2^{\text{ATM}}$  we define:

$$\begin{aligned} sR_i^{\text{EL-O}}s' &\text{ iff } s \text{ and } s' \text{ agree on every } \alpha \text{ such that } s \models S_i \alpha; \\ sR_{\text{Agt}}^{\text{EL-O}}s' &\text{ iff } s \text{ and } s' \text{ agree on every } \alpha \text{ such that } s \models JS \alpha. \end{aligned}$$

As joint vision implies individual vision, it is clear that all  $R_i^{\text{EL-O}}$  are included in  $R_{\text{Agt}}^{\text{EL-O}}$ . While these relations are clearly reflexive, it is not immediately clear that they are also symmetric and transitive; so let us prove that.

**Proposition 3.10.** *All relations  $R_i^{\text{EL-O}}$  and  $R_{\text{Agt}}^{\text{EL-O}}$  are equivalence relations.*

*Proof.* We prove symmetry and transitivity of  $R_i^{\text{EL-O}}$  for an arbitrary  $i$ ; the proof for  $R_{\text{Agt}}^{\text{EL-O}}$  is analogous. For symmetry, suppose  $sR_i^{\text{EL-O}}s'$ . Then for every  $\alpha$ , if  $s \models S_i \alpha$  then  $s$  and  $s'$  agree on  $\alpha$ . Take an arbitrary  $\beta$  such that  $s' \models S_i \beta$ .  $s'$  agrees with  $s$  on  $S_i \beta$  because  $s \models S_i S_i \beta$ . Hence  $s \models S_i \beta$ , and  $s$  and  $s'$  agree on  $\beta$ , and therefore  $s'R_i^{\text{EL-O}}s$ . For transitivity, suppose  $sR_i^{\text{EL-O}}s'$  and  $s'R_i^{\text{EL-O}}s''$  and suppose  $s \models S_i \alpha$ . We show that  $s$  and  $s''$  agree on  $\alpha$ . By hypothesis,  $s$  and  $s'$  agree on  $\alpha$ . As observed above,  $s$  and  $s'$  also agree on  $S_i \alpha$  because  $s \models S_i S_i \alpha$ . Hence  $s' \models S_i \alpha$ , and so  $s'$  and  $s''$  agree on  $\alpha$ . Therefore  $s$  and  $s''$  agree on  $\alpha$ .  $\square$

We are ready to define the *canonical Kripke model for EL-O* as the tuple  $M^{\text{EL-O}} = (W^{\text{EL-O}}, R_{\text{Agt}}^{\text{EL-O}}, \{R_i^{\text{EL-O}}\}_{i \in \text{Agt}}, V^{\text{EL-O}})$  with

$$\begin{aligned} W^{\text{EL-O}} &= I\text{-STATES} = \{s^{\Rightarrow} \cup I\text{-ATM} : s \in 2^{\text{ATM}}\}, \\ R_{\text{Agt}}^{\text{EL-O}} &= R_{\text{Agt}}^{\text{EL-O}} \cap (W^{\text{EL-O}} \times W^{\text{EL-O}}), \\ R_i^{\text{EL-O}} &= R_i^{\text{EL-O}} \cap (W^{\text{EL-O}} \times W^{\text{EL-O}}), \\ V^{\text{EL-O}}(w) &= w \cap \text{Prop for every } w \in W^{\text{EL-O}}. \end{aligned}$$

*Remark 3.4.* In EL-O, the empty initial state  $s_0 = \emptyset$  models maximal ignorance: the agents do not know anything beyond tautologies. The simplicity of this modeling contrasts with the corresponding pointed Kripke model  $(M^{\text{EL-O}}, I\text{-ATM})$ : in that model, every agent can access *infinitely many* possible worlds from  $s_0$ . Actually the classical examples in introductory textbooks and articles about epistemic logic (such as the muddy children puzzle) are all modeled by finite Kripke models. As recently observed by Artemov [4], such finite models presuppose more or less tacitly a lot of common knowledge, which is too strong an assumption in many situations.

**Proposition 3.11.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$  and  $s \in W^{\text{EL-O}}$ . Then  $s \models \varphi$  iff  $(M^{\text{EL-O}}, s) \models \varphi$ .*

*Proof.* The proof is by induction on the structure of  $\varphi$ . The only interesting case is the base case of atoms  $\alpha \in ATM$ . We use induction on the length of  $\alpha$ .

When  $\alpha = p$  then  $(M^{\text{EL-O}}, s) \models p$  iff  $s \models p$  for any  $s \in W^{\text{EL-O}}$  by definition of  $V^{\text{EL-O}}$ . When  $\alpha = S_i \beta$  we prove the two directions of the equivalence.

- If  $s \models S_i \beta$  for some  $\beta \in ATM$  and some  $s \in W^{\text{EL-O}}$ : consider  $s' \in W^{\text{EL-O}}$  such that  $s R_i^{\text{EL-O}} s'$ . By definition of  $R_i^{\text{EL-O}}$ ,  $s$  and  $s'$  agree on all  $\alpha$  such that  $s \models S_i \alpha$ . Therefore if  $s \models \beta$  then  $s' \models \beta$ , and by the induction hypothesis,  $(M^{\text{EL-O}}, s') \models \beta$ ; hence  $(M^{\text{EL-O}}, s) \models K_i \beta$ . If  $s \models \neg \beta$ , by the same argument, we have that  $(M^{\text{EL-O}}, s) \models K_i \neg \beta$ . Therefore  $(M^{\text{EL-O}}, s) \models S_i \beta$ .
- If  $(M^{\text{EL-O}}, s) \models S_i \beta$  for some  $\beta \in ATM$  and some  $s \in W^{\text{EL-O}}$ , then either  $(M^{\text{EL-O}}, s) \models K_i \beta$  or  $(M^{\text{EL-O}}, s) \models K_i \neg \beta$ . Suppose that  $s \not\models S_i \beta$ . In particular  $\beta \notin I-ATM$ . If  $s \models \beta$ , consider  $s' = s \setminus \beta^{\leftarrow}$ , and if  $s \not\models \beta$ , consider  $s' = s \cup \beta^{\Rightarrow}$ . In both cases  $s' \in W^{\text{EL-O}}$ , and  $s$  and  $s'$  agree on all  $\alpha$  such that  $s \models S_i \alpha$ . Therefore  $s R_i^{\text{EL-O}} s'$ , but by the induction hypothesis,  $(M^{\text{EL-O}}, s') \models \beta$  iff  $s' \models \beta$  iff  $s \models \neg \beta$  iff  $(M^{\text{EL-O}}, s) \models \neg \beta$ : this contradicts the hypothesis that  $(M^{\text{EL-O}}, s) \models S_i \beta$ .

For the case  $\alpha = JS \beta$  we proceed similarly.  $\square$

### 3.3.2 Completeness of the EL-O fragment of $Fml_{\text{EL}}^{\text{if}}$ w.r.t. $S5 \setminus GFP$ validity

We now show that the EL-O semantics in terms of observability and the standard Kripke semantics have the same validities as far as the EL-O fragment is concerned.

**Proposition 3.12.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  be a formula. Then  $\varphi$  is valid in  $S5 \setminus GFP$  Kripke models iff  $\varphi$  is provable in CPC from the EL-O axioms of Table 3.1.*

*Proof.* For soundness we show that the schemas  $S_i S_i \alpha$ ,  $JS JS \alpha$ ,  $JS S_i S_i \alpha$ ,  $JS \alpha \rightarrow S_i \alpha$ ,  $JS \alpha \rightarrow JS S_i \alpha$  are valid in  $S5 \setminus GFP$  Kripke models. For the first it suffices to observe that  $S_i S_i \alpha$  is equivalent to  $K_i(K_i \alpha \vee K_i \neg \alpha) \vee K_i \neg(K_i \alpha \vee K_i \neg \alpha)$ . The latter is equivalent in  $S5$  to the propositionally valid  $K_i \alpha \vee K_i \neg \alpha \vee \neg(K_i \alpha \vee K_i \neg \alpha)$ . Validity of the other axiom schemas can be proved in a similar manner.

For completeness suppose that  $\varphi$  is not provable from the EL-O axioms. By Proposition 3.9 there exists  $s \subseteq ATM$  such that  $s \not\models \varphi$ . Hence  $s^{\Rightarrow} \cup I-ATM \not\models \varphi$  by Proposition 3.7 and, as  $s^{\Rightarrow} \cup I-ATM \in W^{\text{EL-O}}$ ,  $(M^{\text{EL-O}}, s^{\Rightarrow} \cup I-ATM) \not\models \varphi$  by Proposition 3.11. Hence  $\varphi$  cannot be valid in  $S5 \setminus GFP$  Kripke models.  $\square$

### 3.3.3 Axiomatization of the $M^{\text{EL-O}}$ validities in $Fml_{\text{EL}}$

We return to  $Fml_{\text{EL}}$ , that is, the standard language of ‘knowing that’. It follows from Propositions 3.11 and 3.12 that there is no formula in the EL-O fragment that is valid in  $M^{\text{EL-O}}$  without being valid in the set of all  $S5 \setminus GFP$  Kripke models. This fails to hold for the full language  $Fml_{\text{EL}}$ . This can be seen from the axiomatization of the  $Fml_{\text{EL}}$  validities in  $M^{\text{EL-O}}$  of [71] and [99, Chapter 2].<sup>1</sup>

<sup>1</sup>Our semantics differs from the semantics in these papers in that it is finitary. The only thing that changes is the soundness proof, which is routine.

It is made up of: the EL-O axioms of Table 3.1; the axioms

$$\begin{aligned} K_i \alpha &\leftrightarrow \alpha \wedge S_i \alpha, & K_i \neg \alpha &\leftrightarrow \neg \alpha \wedge S_i \alpha, \\ CK \alpha &\leftrightarrow \alpha \wedge JS \alpha, & CK \neg \alpha &\leftrightarrow \neg \alpha \wedge JS \alpha \end{aligned}$$

relating  $K_i$  to  $S_i$  and  $CK$  to  $JS$ ; all S5 principles for  $K_i$  and  $CK$ ; plus two axioms distributing  $K_i$  and  $CK$  over clauses: if  $A^+$  and  $A^-$  are two sets of atoms,

$$\begin{aligned} K_i \left( \bigvee_{\alpha \in A^+} \alpha \vee \bigvee_{\alpha \in A^-} \neg \alpha \right) &\leftrightarrow \left( \bigvee_{\alpha \in A^+} K_i \alpha \right) \vee \left( \bigvee_{\alpha \in A^-} K_i \neg \alpha \right), \\ CK \left( \bigvee_{\alpha \in A^+} \alpha \vee \bigvee_{\alpha \in A^-} \neg \alpha \right) &\leftrightarrow \left( \bigvee_{\alpha \in A^+} CK \alpha \right) \vee \left( \bigvee_{\alpha \in A^-} CK \neg \alpha \right) \end{aligned}$$

if  $(A^+)^{\leftarrow} \cap A^- = \emptyset$ , and

$$\begin{aligned} K_i \left( \bigvee_{\alpha \in A^+} \alpha \vee \bigvee_{\alpha \in A^-} \neg \alpha \right) &\leftrightarrow \top, \\ CK \left( \bigvee_{\alpha \in A^+} \alpha \vee \bigvee_{\alpha \in A^-} \neg \alpha \right) &\leftrightarrow \top \end{aligned}$$

otherwise. These last four axioms are specific to observability-based knowledge and are typically invalid in normal modal logics. An instance of the first one is  $K_i(p \vee q) \leftrightarrow (K_i p \vee K_i q)$ , for different  $p$  and  $q$ . This is a strong principle: to give an example, if it is known that the butler or the gardener was the murderer then it is also known which of the two it was. The reason is that our logic is built on atomic observability information: what is modeled in  $M^{\text{EL-O}}$  are forms of individual and common knowledge that are respectively obtained via individual observation and joint observation of facts. This differs conceptually from the classical operators of individual and common knowledge as studied in epistemic logic [52]. The property of distribution over falsifiable disjunctions of literals is common in epistemic logics that are based on observability of propositional variables. It is shown in [22] that it can be avoided if one introduces new propositional variables.

The axiomatics allows us to reduce every  $Fml_{\text{EL}}$  formula to an equivalent (in  $M^{\text{EL-O}}$ )  $\mathcal{L}_{\text{bool}}(ATM)$  formula. For example, positive introspection  $K_i p \rightarrow K_i K_i p$  reduces to  $(p \wedge S_i p) \rightarrow (p \wedge S_i p \wedge S_i S_i p)$ . Observe that the latter is EL-O valid.

### 3.4 Adding constants: EL-OC

Consider the examples we have mentioned up until now: they were of agents knowing secrets, or the contents of messages. We can also mention here the Two Generals' Problem [1, 63], in which two generals send a messenger back and forth to make sure they agree on the time of an attack (we will come back to this example in Chapter 5). As is standard, we have modeled these secrets and messages by propositional variables, making them entities that could be true or false. However, this may seem counterintuitive: what does it mean for a secret to be false? Or for a time of attack to be modeled by a propositional variable? While we can contrive to find explanations for these modeling choices,

we choose here to remark that representing these entities as constants, rather than variables, and seeing the  $S_i$  operator applied to constants as a ‘knowing what’, or ‘knowing the value of’, operator, rather than a ‘knowing whether’ operator, would change almost nothing to the EL-O framework.

This notion of ‘knowing what’ is not new. Plaza introduces and axiomatizes the *logic of knowledge with designators* [111],  $\mathbf{LK}_m^d(P, D)$ , where  $m$  is the number of agents,  $P$  is a set of proposition symbols and  $D$  is a set of constants. The main feature of this logic is operator  $Kv$ ; the formula  $Kv_i d$  is read as “agent  $i$  knows the value of  $d$ ”. The semantics are then defined on Kripke models with two valuations, one for propositions and one for designators; agent  $i$  knows the value of  $d$  at a world if the value of  $d$  is the same across all accessible worlds for  $i$  at that world. This notion of knowing the value of a constant is then picked up as ‘knowing what’ by Fan and Wang in [126], and developed more by Wang in [125] as well as by Baltag and colleagues in [14, 17]. While the interest of adding ‘knowing what’ operators to logics of knowledge is extensively argued in [125] especially, particularly in the field of AI, the focus in [126] and [125] is mostly on dynamic versions of the logic in which there are public announcements, while that in [17] is to investigate the use of ‘knowing what’ operators to model dependencies between variables.

We finish this chapter by introducing a version of EL-O called the Epistemic Logic of Observation with Constants, abbreviated to EL-OC. Its language refines that of EL-O: agents may see not only the truth values of propositions, but also the value of constants. We show that this refinement poses no difficulty, and leads to an enlarged field of possibilities regarding the problems that can be modeled.

### 3.4.1 The language of EL-OC

We now consider not only the set *Prop* of propositional variables, but also a countable set of constants *Cst*. We do not change the observability operators  $S_i$  and  $JS$ , but we refine the set of atoms:

$$ATM-C = \{\sigma p : \sigma \in OBS^*, p \in Prop\} \cup \{\sigma c : \sigma \in OBS^+, c \in Cst\}.$$

We still use  $\alpha, \beta, \dots$  to denote atoms, unless specified as members of  $ATM-C \cup Cst$ . Here are some examples involving constants:  $S_1 c$  reads “1 sees the value of  $c$ ”.  $JS S_2 c$  reads “all agents jointly see whether agent 2 sees the value of  $c$ ”. Note that constants are always preceded by at least one observability operator as they have no truth value themselves.

We follow the principles of EL-O w.r.t. introspective atoms. The set of all introspective atoms is

$$I-ATM-C = \{\sigma S_i S_i \alpha : \sigma \in OBS^* \text{ and } \alpha \in ATM-C \cup Cst\} \cup \{\sigma JS \alpha : \sigma \in OBS^+ \text{ and } \alpha \in ATM-C \cup Cst\}.$$

We also adapt the definition of atomic consequence between atoms as follows:

$$\alpha \Rightarrow \beta \text{ iff } \alpha = \beta \text{ or there are } \alpha' \in ATM-C \cup Cst, \sigma \in OBS^+ \text{ such that } \alpha = JS \alpha' \text{ and } \beta = \sigma \alpha'.$$



$Vis_1$	$S_i S_i \alpha$
$Vis_2$	$JS JS \alpha$
$Vis_3$	$JS S_i S_i \alpha$
$Vis_4$	$JS \alpha \rightarrow S_i \alpha$
$Vis_5$	$JS \alpha \rightarrow JS S_i \alpha$

Table 3.2: Axioms for introspection in EL-OC, where  $\alpha \in ATM-C \cup Cst$ 

The language of EL-OC is simply  $\mathcal{L}_{\text{bool}}(ATM-C)$ , defined by the grammar

$$\varphi ::= \alpha \mid \neg\varphi \mid (\varphi \wedge \varphi)$$

where  $\alpha$  ranges over  $ATM-C$ . The set  $ATM(\varphi)$  of atoms of formula  $\varphi$  is defined as in EL-O. For example,  $ATM(S_1 JS p \wedge S_2 c) = \{S_1 JS p, S_2 c\}$ .

### 3.4.2 Semantics and axiomatization

States are subsets of the set of atoms  $ATM-C$ . The set of all states is  $2^{ATM-C}$ . The semantics of EL-OC follow those of EL-O, with the only non-standard case being that of atoms:

$$s \models_{\text{EL-OC}} \alpha \text{ iff } \alpha \in s \Rightarrow \cup I\text{-}ATM\text{-}C$$

for  $\alpha \in ATM-C$ .

**Example 3.6** (Gossip, ctd.). Consider the original gossip problem in which secrets are seen as constants:  $Cst = \{s_i : i \in Agt\}$ . In the initial state  $s_0^{G_1^C}$  of the problem every agent only knows their own secret. Therefore  $s_0^{G_1^C} = \{S_i s_i : i \in Agt\}$ . Then  $s_0^{G_1^C} \models_{\text{EL-OC}} S_i s_i$  and  $s_0^{G_1^C} \models_{\text{EL-OC}} \bigwedge_{j \neq i} \neg S_i s_j$  for every  $i \in Agt$ .

The valid EL-OC formulas are axiomatized by the schemas of Table 3.2. The axioms are the same as those for EL-O, except that  $\alpha$  is taken to be a member of  $ATM-C \cup Cst$ . The proof of completeness is virtually identical to that of Proposition 3.9.

All other properties of EL-O hold: the translation to classical propositional calculus is the same and the finite model property as well as other intermediate results still hold. It follows that the satisfiability problem in EL-OC is NPcomplete.

### 3.4.3 Applications

As stated in the beginning of this section, the practical usefulness of EL-OC is already clear just from looking standard application problems of epistemic logic and planning: in the gossip problem, it does not matter what the secrets are; in the Two Generals' problem, it does not matter what the message is. EL-OC allows us to model in a natural way situations in which agents must share knowledge about some fact, such as a message or a secret, without burdening ourselves with the limitation that this fact must be expressible as a proposition that is true or false.

While sometimes these facts *can* be expressed as propositions without affecting the modeling much, other times allowing agents this broader definition of knowledge enables us to model problems in a much more succinct way. If an agent must know what another agent's phone number is in order to call them, or if they must know a code in order to unlock a door, the only way to model this in terms of 'knowing that' or 'knowing whether' operators is to have a propositional variable for all possible options for that phone number or code, while in EL-OC it suffices to have a constant representing the value of that number or code.

**Example 3.7.** Agent 1 wishes to unlock a door that has a code; the code might be 00, 01, 10, or 11. The code is actually 11, which agent 1 does not know, but agent 2 knows, and agent 1 knows that agent 2 knows what the code is. Consider the variables  $p_{00}$ ,  $p_{01}$ ,  $p_{10}$ , and  $p_{11}$ , where  $p_{ij}$  represents the proposition that the right code is  $ij$ . Then the situation can be modeled in EL-O as <sup>2</sup> :

$$\begin{aligned} p_{11} \wedge \neg S_1 p_{00} \wedge \neg S_1 p_{01} \wedge \neg S_1 p_{10} \wedge \neg S_1 p_{11} \\ \wedge S_2 p_{00} \wedge S_2 p_{01} \wedge S_2 p_{10} \wedge S_2 p_{11} \\ \wedge S_1 S_2 p_{00} \wedge S_1 S_2 p_{01} \wedge S_1 S_2 p_{10} \wedge S_1 S_2 p_{11} \end{aligned}$$

On the other hand, consider the constant  $c$  representing the code of the door. Then the situation can be modeled in EL-OC as:

$$\neg S_1 c \wedge S_2 c \wedge S_1 S_2 c$$

Notice that the code is not made explicit in this representation; if the modeler wishes for the code to be explicit, the previous representation can also be used in EL-OC, while it is unclear what the converse (considering  $c$  as a propositional variable in EL-O) would mean.

**Example 3.8** (Gossip, ctd.). Some variants of the gossip problem include the fact that agents may or may not know other agents' telephone numbers, and must know another agent's number in order to call them. This is easily modeled by adding agents' telephone numbers to the set of constants:  $Cst = \{s_i : i \in Agt\} \cup \{t_i : i \in Agt\}$ , where  $s_i$  represents each agent's secret and  $t_i$  represents each agent's telephone number. Agent 1 knowing agent 2's number is then expressed as  $S_1 t_2$ .

### 3.5 Discussion: joint vision for arbitrary groups of agents

It seems natural to want to generalize the joint vision operator  $JS$  to an operator  $JS_G$  of joint vision for any group  $G \subseteq Agt$ . For example, calls in the gossip problem should lead to some common knowledge between the two agents in the call, even though common knowledge between all agents can never be achieved as soon as there are more than two agents.

<sup>2</sup>This does not actually fully represent the situation, as agents should also know that the code must be exactly one of the four possibilities, that is, that exactly one of the  $p_{i,j}$  must be true. This, however, cannot be expressed in EL-O, as the language of EL-O restricts agents from knowing these kinds of disjunctions.

However, this generalization is not that easily added into EL-O; this was investigated in [38], but the main issue, which we describe below, was overlooked. The reason for this difficulty is that introducing these generalized operators would mean requiring more complex reasoning from agents. We analyze this in this section.

Joint vision for a singleton is identified with individual vision:  $JS_{\{1\}} \alpha$  is the same thing as  $S_1 \alpha$ . First, we remark that common knowledge of  $\alpha$  within a group  $G$  should imply common knowledge of  $\alpha$  within all subgroups of  $G$ , i.e.,  $JS_G \alpha \rightarrow JS_H \alpha$  should be valid for any  $H \subseteq G \subseteq \text{Agt}$ . We hence generalize the atomic consequence  $\Rightarrow$  to  $\Rightarrow$ :

$$\alpha \Rightarrow \beta \text{ iff } \alpha = \beta \text{ or } \alpha = JS_G \alpha' \text{ and } \beta = \sigma \alpha' \text{ for some } \sigma \in OBS_G^+$$

where  $OBS_G = \{JS_H : H \subseteq G\}$ .

Suppose now that  $S_3 JS_{\{1,2\}} p$  is true at a state  $s$ . If  $JS_{\{1,2\}} p$  is true, then agent 3 knows that  $JS_{\{1,2\}} p$  is true. Hence agent 3 should then know that  $S_1 p$  and  $S_2 p$  are true. More generally, if  $JS_G \alpha$  is true in some state  $s$  for some group  $G$  and atom  $\alpha$ , then if  $\alpha$  is true in  $s$ ,  $JS_G \beta$  should also be true for any atom  $\beta$  such that  $\alpha \Rightarrow \beta$ . This is already a step up from reasoning in EL-O: some truths might be a consequence of two atoms rather than just one.

That is not all. Considering another state  $s'$  in which  $S_3 JS_{\{1,2\}} p$  is true, we now suppose that  $JS_{\{1,2\}} p$  is false. Hence here agent 3 knows that  $JS_{\{1,2\}} p$  is false. But then agent 3 should also know that  $JS_{\{1,2,4\}} p$  is false, as  $JS_{\{1,2,4\}} p$  would imply  $JS_{\{1,2\}} p$ . More generally, if  $JS_G \alpha$  is true in some state  $s$  for some group  $G$  and atom  $\alpha$ , then if  $\alpha$  is false in  $s$ ,  $JS_G \beta$  should also be true for any atom  $\beta$  such that  $\beta \Rightarrow \alpha$ .

Here lies the difficulty in figuring out semantics for an EL-O with this generalized joint knowledge operator: while it is relatively straightforward to add truths to a set of atoms, it is more difficult to pronounce atoms ‘definitely false’. Completing an arbitrary set of atoms into the corresponding set of truths, or even checking that a set of atoms is consistent to begin with, is not straightforward at all, and we leave it as an open problem to figure out whether this can be done in general in a finite number of steps. We could avoid the problem by requiring states to be infinite sets of atoms required to be closed under the rules described above; however, working with infinite states with no known way of turning back to finite states is not only impractical when considering problems such as model checking, but also not very interesting for concrete applications such as planning. We will therefore not be considering this generalization any further in this thesis.

### 3.6 Conclusion

We have updated the logic EL-O of [99], a simple epistemic logic with individual and joint visibility operators. We allow in particular for states to consist of arbitrary sets of atoms. This leads to a finite model property as well as NP-completeness of the EL-O satisfiability problem. We have given an axiomatization for EL-O and shown that it is a fragment of the logic  $S5 \setminus GFP$ , that is, the logic of S5 individual and common knowledge in which common knowledge does not follow the induction principle. Finally, we have shown that adding

constants to EL-O is straightforward, resulting in the logic EL-OC, while adding joint vision for arbitrary subgroups of agents remains an open problem.

## Chapter 4

# On common knowledge whether

### Connaissance commune et ‘savoir si’

Nous avons vu dans le chapitre précédent que la logique EL-O correspondait à un fragment de  $S5 \setminus GFP$ , la version de S5 qui ne satisfait pas le principe d’induction pour la connaissance commune. Les présentations usuelles de ce principe d’induction reposent soit sur un axiome de plus grand point fixe [87, 83, 65], soit sur une règle d’induction [66, 52] :

$$\begin{array}{l} GFP \quad CK(\varphi \rightarrow EK\varphi) \rightarrow (\varphi \rightarrow CK\varphi); \\ RGFP \quad \text{from } \varphi \rightarrow EK(\psi \wedge \varphi), \text{ infer } \varphi \rightarrow CK\psi. \end{array}$$

où  $CK$  se lit “il y a connaissance commune que” et  $EK$  se lit “tous les agents savent que”. Si ces principes sont intuitifs dans le cadre d’autres logiques telles la logique temporelle, leur interprétation est moins claire lorsqu’on considère la connaissance commune. Leur traduction en termes du langage EL-O n’est pas non plus évidente.

Dans ce chapitre nous prenons un peu de recul et proposons une axiomatisation alternative pour la connaissance commune S5, que nous trouvons plus intuitive que les axiomatisations standard et qui se traduit bien aux logiques avec l’opérateur ‘savoir si’. Cette axiomatisation repose sur le nouvel axiome  $GFP0$  :

$$GFP0 \quad CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi).$$

Nous montrons que ce nouvel axiome est un théorème de toute logique de connaissance (c’est à dire toute logique où l’opérateur de connaissance individuelle est au moins KT) et qu’il est équivalent à  $GFP$  si la connaissance commune est S5. Nous montrons également que  $GFP0$  n’est plus correct si l’on passe de la connaissance à la croyance.

À partir de cette nouvelle axiomatisation, il est naturel de définir une axiomatisation des logiques avec opérateurs ‘savoir si’, en prenant l’axiome suivant:

$$GFP2 \quad CKif\varphi \leftrightarrow (EKif\varphi \wedge CKifEKif\varphi).$$

Ce dernier axiome nous ramène enfin vers EL-O, et nous terminons ce chapitre en définissant une nouvelle version de EL-O dans lequel le principe d'induction pour la connaissance commune est vérifié. Cette logique, que nous nommons EL-O+*Ind*, repose sur une restriction de la sémantique de EL-O afin de se limiter aux états vérifiant explicitement l'axiome *GFP2*. Nous étudions les propriétés de EL-O+*Ind* et montrons qu'il s'agit bien d'un fragment de la logique S5 standard.

We have seen in the previous chapter that the logic EL-O corresponds to a fragment of  $S5 \setminus GFP$ , that is, a version of S5 which does not follow the induction principle for common knowledge. When looking at the usual presentations of this induction principle, it is not immediately obvious how to translate them into the language of EL-O. In this chapter we take a step back and propose an alternative axiomatization for S5 common knowledge that is, in our opinion, more intuitive than the standard axiomatizations, and translates well to logics of knowing whether. After a brief discussion on the applicability of this new axiomatizations to logics weaker than S5, we take inspiration from the new axiom to define a second version of EL-O which is a true fragment of S5.

This chapter is mainly based on the following publications:

- Herzig, A. & Perrotin, E. (2020). *On the axiomatization of common knowledge*. Advances in Modal Logic (AiML 2020) (13:309–328).
- Herzig, A. & Perrotin, E. (2022, forthcoming). *Efficient reasoning about knowledge and common knowledge*. Many-Valued and Modal Logics (Essays in honour of Yu. V. Ivlev).

## Contents

<b>4.1</b>	<b>An alternative axiomatization of S5 common knowledge</b>	<b>50</b>
4.1.1	Soundness of the <i>GFP0</i> axiomatics	50
4.1.2	Completeness of the <i>GFP0</i> axiomatics for S5 knowledge	51
<b>4.2</b>	<b>Commonly knowing whether</b>	<b>52</b>
<b>4.3</b>	<b>Discussion: epistemic logics weaker than S5</b>	<b>58</b>
4.3.1	Epistemic logics between KT and S5	58
4.3.2	<i>GFP0</i> is not appropriate for belief	59
<b>4.4</b>	<b>A second version of EL-O</b>	<b>59</b>
4.4.1	Language and semantics of EL-O+ <i>Ind</i>	60
4.4.2	Completion of a finite state and finite model property	60
4.4.3	Complexity of EL-O+ <i>Ind</i> satisfiability	62
4.4.4	Axiomatization of EL-O+ <i>Ind</i> validities	63
<b>4.5</b>	<b>EL-O+<i>Ind</i> as a fragment of epistemic logic S5</b>	<b>64</b>
4.5.1	A canonical model for EL-O+ <i>Ind</i>	64
4.5.2	Completeness of the EL-O+ <i>Ind</i> fragment w.r.t. S5 validity	66
<b>4.6</b>	<b>Conclusion</b>	<b>67</b>

The standard axiomatizations of the logic of common knowledge, given in Chapter 2 and recalled here in Table 4.1, contain either the induction axiom schema, alias greatest fixed-point axiom *GFP* [87, 83, 65], or alternatively the induction rule *RGFP* [66, 52]:

$$\begin{array}{l} \textit{GFP} \quad CK(\varphi \rightarrow EK\varphi) \rightarrow (\varphi \rightarrow CK\varphi); \\ \textit{RGFP} \quad \text{from } \varphi \rightarrow EK(\psi \wedge \varphi), \text{ infer } \varphi \rightarrow CK\psi. \end{array}$$

In the proof theory literature there exist sequent system counterparts of these principles, e.g. in [2, 75]. Similar axioms and rules were used to axiomatize common belief [31, 91].

Such inductive principles are common in temporal logics, where they mirror induction on the natural numbers. There, the reading is obvious and the intuitive meaning is clear. More generally, we can make sense of such principles when interpreted on well-founded orderings. However, the meaning of the induction axiom schema is less obvious when the modal operator is that of common knowledge, and one might even wonder whether it is a reasonable principle at all. To witness the difficulty to find an intuitive reading to the above principles, consider the reading of *RGFP* that is given in the introductory chapter of the Handbook of Epistemic Logic:

“If it is the case that  $\varphi$  is ‘self-evident’, in the sense that if it is true, then everyone knows it, and, in addition, if  $\varphi$  is true, then everyone knows  $\psi$ , we can show by induction that if  $\varphi$  is true, then so is  $EK^k(\psi \wedge \varphi)$  for all  $k$ .” [49]

The explanations in the standard texts resort to concepts such as ‘ $\varphi$  indicates to every agent that  $\psi$  is true’ [89], ‘ $\varphi$  is evident’ [105], ‘it is public that  $\varphi$  is true’ [117], or ‘ $\varphi$  is a common basis implying shared belief in  $\psi$ ’ [55]. With these understandings *RGFP* can be read “if  $\varphi$  is public and indicates  $\psi$  to everybody then truth of  $\varphi$  implies that  $\psi$  is common knowledge”. The formalisation of these supplementary concepts however introduces further complications, see e.g. [44] for a tentative to settle the logic of ‘indicates’.

Can the above inductive principles be replaced by principles with more intuitive appeal? In this chapter we propose a new axiom schema:

$$\textit{GFP0} \quad CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi).$$

Unlike *GFP* and *RGFP*, it can be read straightforwardly: “if it is common knowledge that everybody knows whether  $\varphi$  then it is common knowledge whether  $\varphi$ ”; or alternatively: “common knowledge that the truth value of  $\varphi$  is shared knowledge implies that the truth value of  $\varphi$  is common knowledge”. In this chapter we consider *KT*- and *S5*-based common knowledge. We prove the following results:

1. *GFP0* is a theorem if the logic of individual knowledge is at least *KT*;
2. *GFP0* is equivalent to *GFP* if the logic of individual knowledge is *S5*;
3. *GFP0* leads to a simple and intuitive axiomatization of *S5*-based ‘common knowledge whether’, based on the axiom *GFP2*:

$$\textit{GFP2} \quad CKif\varphi \leftrightarrow (EKif\varphi \wedge CKifEKif\varphi);$$

<i>GFP</i> -based axiomatics	
At least $\text{KT}(K_i)$	see Table 2.3
$RN(CK)$	from $\varphi$ , infer $CK\varphi$
$K(CK)$	$CK(\varphi \rightarrow \psi) \rightarrow (CK\varphi \rightarrow CK\psi)$
$T(CK)$	$CK\varphi \rightarrow \varphi$
$*5(CK)$	$\neg CK\varphi \rightarrow CK\neg CK\varphi$
$FP'$	$CK\varphi \rightarrow EKCK\varphi$
$GFP$	$CK(\varphi \rightarrow EK\varphi) \rightarrow (\varphi \rightarrow CK\varphi)$
<i>RGFP</i> -based axiomatics	
At least $\text{KT}(K_i)$	see Table 2.3
$FP$	$CK\varphi \rightarrow EK(\varphi \wedge CK\varphi)$
$RGFP$	from $\varphi \rightarrow EK(\psi \wedge \varphi)$ , infer $\varphi \rightarrow CK\psi$

Table 4.1: Two axiomatizations of common knowledge for logics that are at least  $\text{KT}$ : the *GFP* axiomatics with an induction axiom of [65] (top) and the *RGFP* axiomatics with an induction rule of [66, 52] (bottom). For  $\text{S5}$  common knowledge, add the starred axiom  $*5(CK)$  to the *GFP*-based axiomatics.

$\text{KT}(K_i)$	
$RN(CK)$	from $\varphi$ , infer $CK\varphi$
$K(CK)$	$CK(\varphi \rightarrow \psi) \rightarrow (CK\varphi \rightarrow CK\psi)$
$4(CK)$	$CK\varphi \rightarrow CKCK\varphi$
$FP0$	$CK\varphi \rightarrow EK\varphi$
$GFP0$	$CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)$

Table 4.2: Alternative axiomatization of  $\text{S5}$  common knowledge: the *GFP0* axiomatics.

#### 4. *GFP0* is specific to knowledge and fails for belief.

For the sake of simplicity we here only consider shared and common knowledge of the set of all agents. However, everything in sections 4.1 to 4.3 straightforwardly generalises to common knowledge of arbitrary sets of agents.

The third point listed above brings us back to  $\text{EL-O}$ : can this new axiomatization point us towards a version of  $\text{EL-O}$  in which the induction axiom for common knowledge whether is valid? In the second half of this chapter we present a such version of  $\text{EL-O}$ , which we name  $\text{EL-O}+Ind$ . We study the properties of  $\text{EL-O}+Ind$  and show that it is a fragment of the logic  $\text{S5}$ .



## 4.1 An alternative axiomatization of S5 common knowledge

Table 4.2 contains a new axiomatics of common knowledge. The main difference w.r.t. the *GFP* axiomatics recalled in Table 4.1 is that the induction axiom *GFP* is replaced by *GFP0*. A further difference is that our axiomatics explicits  $\downarrow(CK)$ , which is a theorem of the *GFP* and *RGFP* axiomatics. Finally and thanks to  $\downarrow(CK)$ , our version of the fixed-point axiom *FP0* is weaker than *FP'* (and a fortiori weaker than *FP*). It is however strong enough to entail  $T(CK)$ :  $CK\varphi \rightarrow \varphi$  (together with  $T(K_i)$ ).

Observe that it follows from Proposition 2.2 and the fact that *CK* is a normal modal operator that the two axioms

$$\begin{aligned} \text{GFP0} \quad & CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi) \\ \text{GFP1} \quad & CK \bigwedge_{i \in \text{Agt}} (K_i\varphi \vee K_i\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi) \end{aligned}$$

are equivalent. The second axiom says that if it is common knowledge that each agent has an epistemic position w.r.t.  $\varphi$  then either  $\varphi$  or  $\neg\varphi$  are common knowledge.

### 4.1.1 Soundness of the *GFP0* axiomatics

We prove soundness w.r.t. the S5-based *GFP* axiomatics of Table 4.1. The result holds both for the KT-based and the S5-based versions.

The inference rules are the same: *RN(CK)* and modus ponens. It remains to show that our axioms of Table 4.2 are theorems of the S5-based *GFP* axiomatics. The only ones that are missing there are  $\downarrow(CK)$ , *FP0*, and *GFP0*. First,  $\downarrow(CK)$  is, by Proposition 2.4, a theorem of the KT-based *GFP* axiomatics and a fortiori of the S5-based *GFP* axiomatics. Second, *FP0* can be proved from *FP'* and  $T(CK)$ . Third, here is a proof of *GFP0* that relies on  $T(K_i)$ , or rather, its consequence  $T(EK)$ :

**Proposition 4.1.** *GFP0 is a theorem of the KT-based GFP axiomatics (and a fortiori of the S5-based GFP axiomatics).*

*Proof.* We distinguish the two cases  $\varphi$  and  $\neg\varphi$  and prove that  $CK(EK\varphi \vee EK\neg\varphi)$  implies both  $\varphi \rightarrow CK\varphi$  and  $\neg\varphi \rightarrow CK\neg\varphi$ ; from that *GFP0* follows by propositional logic reasoning.

1.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow CK(\varphi \rightarrow EK\varphi)$       by  $T(EK)$ , *RN(CK)*,  $K(CK)$
2.  $CK(\varphi \rightarrow EK\varphi) \rightarrow (\varphi \rightarrow CK\varphi)$       *GFP*
3.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow (\varphi \rightarrow CK\varphi)$       from (1), (2)
4.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow (\neg\varphi \rightarrow CK\neg\varphi)$   
from (3) by uniform subst. of  $\varphi$  by  $\neg\varphi$
5.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)$       from (3), (4)

□

Therefore all theorems of our new *GFP0* axiomatics are also theorems of the *GFP* axiomatics and, by Proposition 2.7, of the *RGFP* axiomatics.

### 4.1.2 Completeness of the $GFP0$ axiomatics for S5 knowledge

We prove completeness w.r.t. the S5-based  $GFP$  axiomatics. We have already seen in Section 4.1.1 that the inference rules are the same; it remains to show that the axioms of the S5-based  $GFP$  axiomatics of Table 4.1 that are not in our  $GFP0$  axiomatics are theorems of the latter. These axioms are  $5(CK)$ ,  $FP'$ , and  $GFP$ . Proposition 2.5 tells us that  $5(CK)$  can be proved from the rest of the S5-based  $GFP$  axiomatics and is therefore redundant: it could be dropped from the  $GFP$  axiomatics. Axiom  $FP'$  can be proved from our  $FP0$ ,  $4(CK)$ ,  $K(CK)$ , and  $RN(CK)$ . It remains to show that  $GFP$  is a theorem of our new axiomatics. The next lemma will be instrumental; its proof uses  $B(EK)$  (via Proposition 2.3) and  $4(CK)$ . (Recall that  $B(EK)$  is a theorem of  $S5(K_i)$ , but not of  $KT(K_i)$ .)

**Lemma 4.1.** *The schema  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  is provable from the axiom schemas  $K(CK)$ ,  $4(CK)$ ,  $RN(CK)$ ,  $FP$ , and the S5 axioms for  $K_i$ .*

*Proof.* The proof is as follows:

1.  $CK(\varphi \rightarrow EK\varphi) \rightarrow EK(\varphi \rightarrow EK\varphi)$  by  $FP$ ,  $EK$  normal
2.  $EK(\varphi \rightarrow EK\varphi) \rightarrow (EK\neg EK\varphi \rightarrow EK\neg\varphi)$   $EK$  normal
3.  $\neg\varphi \rightarrow EK\neg EK\varphi$  Proposition 2.3
4.  $CK(\varphi \rightarrow EK\varphi) \rightarrow (\neg\varphi \rightarrow EK\neg\varphi)$  from (1), (2), (3)
5.  $CKCK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  from (4) by  $RN(CK)$  and  $K(CK)$
6.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CKCK(\varphi \rightarrow EK\varphi)$   $4(CK)$
7.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  from (5), (6)

□

**Proposition 4.2.**  *$GFP$  is provable in the  $GFP0$  axiomatics.*

*Proof.* The proof is as follows:

1.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)$   $GFP0$
2.  $(CK(\varphi \rightarrow EK\varphi) \wedge CK(\neg\varphi \rightarrow EK\neg\varphi)) \rightarrow CK(EK\varphi \vee EK\neg\varphi)$   
by  $RN(CK)$  and  $K(CK)$
3.  $(CK(\varphi \rightarrow EK\varphi) \wedge CK(\neg\varphi \rightarrow EK\neg\varphi)) \rightarrow (CK\varphi \vee CK\neg\varphi)$   
from (1) and (2)
4.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  Lemma 4.1
5.  $CK(\varphi \rightarrow EK\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)$  from (3), (4)
6.  $CK(\varphi \rightarrow EK\varphi) \rightarrow (CK\varphi \vee \neg\varphi)$  from (5) by  $T(CK)$

□

<i>CPC</i>	axiomatics of classical propositional calculus
<i>Sym(Kif<sub>i</sub>)</i>	$Kif_i\varphi \leftrightarrow Kif_i\neg\varphi$
<i>RE(Kif<sub>i</sub>)</i>	from $\varphi \leftrightarrow \psi$ , infer $Kif_i\varphi \leftrightarrow Kif_i\psi$
<i>RN(Kif<sub>i</sub>)</i>	from $\varphi$ , infer $Kif_i\varphi$
<i>Conj(Kif<sub>i</sub>)</i>	$(\varphi \wedge \psi) \rightarrow (Kif_i(\varphi \wedge \psi) \leftrightarrow (Kif_i\varphi \wedge Kif_i\psi))$
<i>*45<sub>1</sub>(Kif<sub>i</sub>)</i>	$Kif_iKif_i\varphi$
<i>*45<sub>2</sub>(Kif<sub>i</sub>)</i>	$Kif_i(\varphi \wedge Kif_i\varphi)$
<i>Def2(EKif)</i>	$EKif\varphi \leftrightarrow \bigwedge_{i \in Agt} Kif_i\varphi$
<i>Sym(CKif)</i>	$CKif\varphi \leftrightarrow CKif\neg\varphi$
<i>RE(CKif)</i>	from $\varphi \leftrightarrow \psi$ , infer $CKif\varphi \leftrightarrow CKif\psi$
<i>RN(CKif)</i>	from $\varphi$ , infer $CK\varphi$
<i>Conj(CKif)</i>	$(\varphi \wedge \psi) \rightarrow (CKif(\varphi \wedge \psi) \leftrightarrow (CKif\varphi \wedge CKif\psi))$
<i>*45<sub>1</sub>(CKif)</i>	$CKifCKif\varphi$
<i>*45<sub>2</sub>(CKif)</i>	$CKif(\varphi \wedge CKif\varphi)$
<i>GFP2</i>	$CKif\varphi \leftrightarrow (EKif\varphi \wedge CKifEKif\varphi)$
<i>Def2(K<sub>i</sub>)</i>	$K_i\varphi \leftrightarrow (\varphi \wedge Kif_i\varphi)$
<i>Def2(EK)</i>	$EK\varphi \leftrightarrow (\varphi \wedge EKif\varphi)$
<i>Def2(CK)</i>	$CK\varphi \leftrightarrow (\varphi \wedge CKif\varphi)$

Table 4.3: Axiomatization of S5 common knowledge whether: the *GFP2* axiomatics.

## 4.2 Commonly knowing whether

In this section we show that our axiomatics of Table 4.2 leads to a simple axiomatization of the S5-based ‘common knowledge whether’ operator.

A straightforward possibility for axiomatizing the ‘knowing whether’ operators is to add to the axiomatics of Table 4.2 the following axioms:

$$\begin{aligned}
 Def1(Kif_i) & Kif_i\varphi \leftrightarrow (K_i\varphi \vee K_i\neg\varphi) \\
 Def1(EKif) & EKif\varphi \leftrightarrow (EK\varphi \vee EK\neg\varphi) \\
 Def1(CKif) & CKif\varphi \leftrightarrow (CK\varphi \vee CK\neg\varphi)
 \end{aligned}$$

However, we are going to take another road here, in view of axiomatizing the fragment without ‘knowing that’ operators. Our axiomatics in Table 4.3 takes the ‘knowing whether’ operators as primitive and defines the ‘knowing that’ operators. The first part is proper to  $Kif_i$  and  $EKif$ . We might have taken over as well the axiomatics of [54]; the principles  $Sym(Kif_i)$ ,  $RE(Kif_i)$ , and  $RN(Kif_i)$  can also be found there, but we find the rest of our axioms a bit simpler than theirs. Axiom  $45_1(Kif_i)$  can be found in [106]. The second part of our axiomatics parallels the first part and moreover has a single greatest fixed-point axiom relating  $EKif$  and  $CKif$  (that is perhaps better called a fixed-point axiom *tout court*: its syntactical form is very close to that of a possible fixed-point axiom for common belief  $CB\varphi \leftrightarrow (EB\varphi \wedge EB CB\varphi)$ ). The third part contains the definitions of the ‘knowing that’ operators.

We are going to prove soundness and completeness of the axiomatics of Table 4.3 w.r.t. the S5-based *GFP0* axiomatics (more precisely: w.r.t. the extension of the latter by  $Def1(Kif_i)$ ,  $Def1(EKif)$ , and  $Def1(CKif)$ ).

**Proposition 4.3.** *For the S5-based GFP2 axiomatics of Table 4.3, all inference rules are derivable and all axioms are theorems in the S5-based GFP0 axiomatics.*

To show this, we prove each principle of Table 4.3. We start with the last three definitions so that we can use them in the rest of the proofs.

$$\boxed{\text{Def2}(K_i) \quad K_i\varphi \leftrightarrow (\varphi \wedge Kif_i\varphi)}$$

*Proof.* 1.  $K_i\varphi \leftrightarrow (\varphi \wedge (K_i\varphi \vee K_i\neg\varphi))$  from  $T(K_i)$

2.  $K_i\varphi \leftrightarrow (\varphi \wedge Kif_i\varphi)$  from (1) and  $\text{Def1}(Kif_i)$   $\square$

$$\boxed{\text{Def2}(EK) \quad EK\varphi \leftrightarrow (\varphi \wedge EKif\varphi)}$$

*Proof.* The proof follows the lines of that of  $\text{Def2}(K_i)$ , using  $\text{Def1}(EKif)$  instead of  $\text{Def1}(Kif_i)$  and use that  $T(EK)$  is a theorem.  $\square$

$$\boxed{\text{Def2}(CK) \quad CK\varphi \leftrightarrow (\varphi \wedge CKif\varphi)}$$

*Proof.* The proof follows the lines of that of  $\text{Def2}(K_i)$ , using use  $\text{Def1}(CKif)$  instead of  $\text{Def1}(Kif_i)$  and  $T(CK)$  instead of  $T(K_i)$ .  $\square$

$$\boxed{\text{Sym}(Kif_i): Kif_i\varphi \leftrightarrow Kif_i\neg\varphi}$$

*Proof.* 1.  $(K_i\varphi \vee K_i\neg\varphi) \leftrightarrow (K_i\neg\varphi \vee K_i\neg\neg\varphi)$   $K_i$  normal

2.  $Kif_i\varphi \leftrightarrow Kif_i\neg\varphi$  from (1) by  $\text{Def1}(Kif_i)$   $\square$

$$\boxed{\text{RE}(Kif_i): \text{from } \varphi \leftrightarrow \psi, \text{ infer } Kif_i\varphi \leftrightarrow Kif_i\psi}$$

*Proof.* 1.  $\varphi \leftrightarrow \psi$  hypothesis

2.  $K_i\varphi \leftrightarrow K_i\psi$  from (1),  $K_i$  normal

3.  $K_i\neg\varphi \leftrightarrow K_i\neg\psi$  from (1),  $K_i$  normal

4.  $(K_i\varphi \vee K_i\neg\varphi) \leftrightarrow (K_i\psi \vee K_i\neg\psi)$  from (2), (3)

5.  $Kif_i\varphi \leftrightarrow Kif_i\psi$  from (4) by  $\text{Def1}(Kif_i)$   $\square$

$$\boxed{\text{RN}(Kif_i): \text{from } \varphi, \text{ infer } Kif_i\varphi}$$

*Proof.* 1.  $\varphi$  hypothesis

2.  $K_i\varphi$  from (1),  $K_i$  normal

3.  $K_i\varphi \vee K_i\neg\varphi$  from (2)

4.  $Kif_i\varphi$  from (3) by  $Def1(Kif_i)$   $\square$

$$\boxed{Conj(Kif_i): (\varphi \wedge \psi) \rightarrow (Kif_i(\varphi \wedge \psi) \leftrightarrow (Kif_i\varphi \wedge Kif_i\psi))}$$

*Proof.* We prove the two implications  $(\varphi \wedge \psi \wedge Kif_i(\varphi \wedge \psi)) \rightarrow Kif_i\varphi$  and  $(\varphi \wedge \psi \wedge Kif_i\varphi \wedge Kif_i\psi) \rightarrow Kif_i(\varphi \wedge \psi)$ , each time using that we have already proved  $Def2(K_i)$  to be a theorem. For the former:

1.  $K_i(\varphi \wedge \psi) \rightarrow (K_i\varphi \vee K_i\neg\varphi)$   $K_i$  normal
2.  $(\varphi \wedge \psi \wedge Kif_i(\varphi \wedge \psi)) \rightarrow Kif_i\varphi$  from (1), theorem  $Def2(K_i)$

For the latter:

1.  $(K_i\varphi \wedge K_i\psi) \rightarrow K_i(\varphi \wedge \psi)$   $K_i$  normal
2.  $(\varphi \wedge Kif_i\varphi \wedge \psi \wedge Kif_i\psi) \rightarrow (\varphi \wedge \psi \wedge Kif_i(\varphi \wedge \psi))$  from (1), thm.  $Def2(K_i)$
3.  $(\varphi \wedge \psi \wedge Kif_i\varphi \wedge Kif_i\psi) \rightarrow Kif_i(\varphi \wedge \psi)$  from (2)

$\square$

$$\boxed{45_1(Kif_i): Kif_i Kif_i\varphi}$$

*Proof.* Similar to the next proof of  $45_2(Kif_i)$ .  $\square$

$$\boxed{45_2(Kif_i): Kif_i(\varphi \wedge Kif_i\varphi)}$$

- Proof.*
1.  $K_i\varphi \vee K_i\neg\varphi \vee (\neg K_i\varphi \wedge \neg K_i\neg\varphi)$
  2.  $K_i\varphi \rightarrow K_i(\varphi \wedge Kif_i\varphi)$  from 4( $K_i$ ) and thm.  $Def2(K_i)$ ,  $K_i$  normal
  3.  $K_i\neg\varphi \rightarrow K_i\neg(\varphi \wedge Kif_i\varphi)$  from  $K_i$  normal
  4.  $(\neg K_i\varphi \wedge \neg K_i\neg\varphi) \rightarrow (K_i\neg K_i\varphi \wedge K_i\neg K_i\neg\varphi)$  from thm. \*5( $K_i$ )
  5.  $(K_i\neg K_i\varphi \wedge K_i\neg K_i\neg\varphi) \rightarrow K_i\neg Kif_i\varphi$  from  $Def1(Kif_i)$ ,  $K_i$  normal
  6.  $(\neg K_i\varphi \wedge \neg K_i\neg\varphi) \rightarrow K_i\neg(\varphi \wedge Kif_i\varphi)$  from (4), (5),  $K_i$  normal
  7.  $K_i(\varphi \wedge Kif_i\varphi) \vee K_i\neg(\varphi \wedge Kif_i\varphi)$  from (1), (2), (3), (6)
  8.  $Kif_i(\varphi \wedge Kif_i\varphi)$  from (7),  $Def1(Kif_i)$

$\square$

$$\boxed{Def2(EKif): EKif\varphi \leftrightarrow \bigwedge_{i \in Agt} Kif_i\varphi}$$

*Proof.* This is Proposition 2.2.  $\square$

$$\boxed{Sym(CKif): CKif\varphi \leftrightarrow CKif\neg\varphi}$$

*Proof.* Follow the lines of that of  $Sym(Kif_i)$ .  $\square$

$$\boxed{RE(CKif): \text{from } \varphi \leftrightarrow \psi, \text{ infer } CKif\varphi \leftrightarrow CKif\psi}$$

*Proof.* Follow the lines of that of  $RE(Kif_i)$ .  $\square$

$$\boxed{RN(CKif) \quad \text{from } \varphi, \text{ infer } CK\varphi}$$

*Proof.* Follow the lines of that of  $RN(Kif_i)$ .  $\square$

$$\boxed{Conj(CKif) \quad (\varphi \wedge \psi) \rightarrow (CKif(\varphi \wedge \psi) \leftrightarrow (CKif\varphi \wedge CKif\psi))}$$

*Proof.* Follow the lines of that of  $Conj(Kif_i)$ .  $\square$

$$\boxed{*45_1(CKif) \quad CKif\ CKif\varphi}$$

*Proof.* Follow the lines of that of  $45_1(Kif_i)$ .  $\square$

$$\boxed{45_2(CKif) \quad CKif(\varphi \wedge CKif\varphi)}$$

*Proof.* Follow the lines of that of  $45_2(Kif_i)$ .  $\square$

$$\boxed{GFP2 \quad CKif\varphi \leftrightarrow (EKif\varphi \wedge CKifEKif\varphi)}$$

*Proof.* We prove the three implications  $CKif\varphi \rightarrow EKif\varphi$ ,  $CKif\varphi \rightarrow CKifEKif\varphi$ , and  $(EKif\varphi \wedge CKifEKif\varphi) \rightarrow CKif\varphi$ . For the first:

1.  $(CK\varphi \vee CK\neg\varphi) \rightarrow (EK\varphi \vee EK\neg\varphi)$  from  $FP0$
2.  $CKif\varphi \rightarrow EKif\varphi$  from (1),  $Def1(EKif)$ ,  $Def1(CKif)$

For the second:

1.  $CK\varphi \rightarrow CKEK\varphi$  from  $4(CK)$ ,  $FP0$
2.  $CK\varphi \rightarrow CKEKif\varphi$  from (1),  $Def(EKif)$ , normal  $CK$
3.  $CK\neg\varphi \rightarrow CKEKif\neg\varphi$  from (2) by uniform substitution
4.  $CK\neg\varphi \rightarrow CKEKif\varphi$  from (3) by  $Sym(K_i)$
5.  $CKif\varphi \rightarrow CKEKif\varphi$  from (2), (4),  $Def1(CKif)$
6.  $CKif\varphi \rightarrow CKifEKif\varphi$  from (5),  $Def1(CKif)$

For the third:

1.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)$   $GFP0$
2.  $CKEKif\varphi \rightarrow CKif\varphi$  from (1),  $Def1(EKif)$ ,  $Def1(CKif)$
3.  $(EKif\varphi \wedge CKifEKif\varphi) \rightarrow CKif\varphi$  from (2), thm.  $Def2(CK)$

$\square$

**Proposition 4.4.** *For the S5-based GFP0 axiomatics of Table 4.2, all inference rules are derivable and all axioms are theorems in the S5-based GFP2 axiomatics. Moreover, the equivalences  $Def1(K_i)$ ,  $Def1(EK)$ , and  $Def1(CK)$  are theorems in the S5-based GFP2 axiomatics.*

To show this we prove all principles of Table 4.2, starting with the last three definitions.

$$\boxed{\text{Def1}(Kif_i) \quad Kif_i\varphi \leftrightarrow (K_i\varphi \vee K_i\neg\varphi)}$$

- Proof.*
1.  $(K_i\varphi \vee K_i\neg\varphi) \leftrightarrow ((\varphi \wedge Kif_i\varphi) \vee (\neg\varphi \wedge Kif_i\neg\varphi))$  from *Def2*( $K_i$ )
  2.  $Kif_i\neg\varphi \leftrightarrow Kif_i\varphi$  *Sym*( $Kif_i$ )
  3.  $Kif_i\varphi \leftrightarrow (K_i\varphi \vee K_i\neg\varphi)$  from (1), (2)
- 

$$\boxed{\text{Def1}(EKif) \quad EKif\varphi \leftrightarrow (EK\varphi \vee EK\neg\varphi)}$$

*Proof.* Follow the lines of that of *Def1*( $Kif_i$ ). □

$$\boxed{\text{Def1}(CKif) \quad CKif\varphi \leftrightarrow (CK\varphi \vee CK\neg\varphi)}$$

*Proof.* Follow the lines of that of *Def1*( $Kif_i$ ). □

$$\boxed{RN(K_i) \quad \text{from } \varphi, \text{ infer } K_i\varphi}$$

- Proof.*
1.  $\varphi$  hypothesis
  2.  $Kif_i\varphi$  from (1) by *RN*( $Kif_i$ )
  3.  $\varphi \wedge K_i\varphi$  from *Def2*( $K_i$ )
  4.  $K_i\varphi$  from (3)
- 

$$\boxed{K(K_i) \quad K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)}$$

- Proof.*
1.  $(\varphi \wedge Kif_i\varphi \wedge (\varphi \rightarrow \psi) \wedge Kif_i(\varphi \rightarrow \psi)) \rightarrow Kif_i(\varphi \wedge (\varphi \rightarrow \psi))$   
from *Conj*( $Kif_i$ )
  2.  $(\varphi \wedge Kif_i\varphi \wedge (\varphi \rightarrow \psi) \wedge Kif_i(\varphi \rightarrow \psi)) \rightarrow Kif_i(\varphi \wedge \psi)$   
from (1) by *RE*( $Kif_i$ )
  3.  $(\varphi \wedge \psi \wedge Kif_i(\varphi \wedge \psi)) \rightarrow Kif_i\psi$  from *Conj*( $Kif_i$ )
  4.  $(\varphi \wedge Kif_i\varphi \wedge (\varphi \rightarrow \psi) \wedge Kif_i(\varphi \rightarrow \psi)) \rightarrow (\psi \wedge Kif_i\psi)$  from (2), (3)
  5.  $K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$  from (4) by *Def2*( $K_i$ )
- 

$$\boxed{T(K_i) \quad K_i\varphi \rightarrow \varphi}$$

- Proof.*
1.  $(\varphi \wedge Kif_i\varphi) \rightarrow \varphi$
  2.  $K_i\varphi \rightarrow \varphi$  from (1) by *Def2*( $K_i$ )
- 

$$\boxed{*5(K_i) \quad \neg K_i\varphi \rightarrow K_i\neg K_i\varphi}$$

- Proof.*
1.  $Kif_i(\varphi \wedge Kif_i\varphi)$   $45_2(Kif_i)$
  2.  $Kif_iK_i\varphi$  from (1) by  $Def2(K_i)$
  3.  $Kif_i\neg K_i\varphi$  from (2) by  $Sym(Kif_i)$
  4.  $\neg K_i\varphi \rightarrow (\neg K_i\varphi \wedge Kif_i\neg K_i\varphi)$  from (3)
  5.  $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$  from (4) by  $Def2(K_i)$
- 

$$\boxed{RN(CK) \quad \text{from } \varphi, \text{ infer } CK\varphi}$$

*Proof.* Follow the lines of that of  $RN(K_i)$ . □

$$\boxed{K(CK) \quad CK(\varphi \rightarrow \psi) \rightarrow (CK\varphi \rightarrow CK\psi)}$$

*Proof.* Follow the lines of that of  $K(K_i)$ . □

$$\boxed{T(CK) \quad CK\varphi \rightarrow \varphi}$$

*Proof.* Follow the lines of that of  $T(K_i)$ . □

$$\boxed{FP0 \quad CK\varphi \rightarrow EK\varphi}$$

- Proof.*
1.  $(\varphi \wedge CKif\varphi) \rightarrow (\varphi \wedge EKif\varphi)$  from  $GFP2$
  2.  $CK\varphi \rightarrow EK\varphi$  from (1) by  $Def2(CK)$ ,  $Def2(EK)$
- 

$$\boxed{GFP0 \quad CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)}$$

- Proof.*
1.  $(EKif\varphi \wedge CKifEKif\varphi) \rightarrow CKif\varphi$  from  $GFP2$
  2.  $((EK\varphi \vee EK\neg\varphi) \wedge CKif(EK\varphi \vee EK\neg\varphi)) \rightarrow CKif\varphi$   
from (1) by thm.  $Def1(EKif)$  and  $RE(CKif)$
  3.  $CK(EK\varphi \vee EK\neg\varphi) \rightarrow (CK\varphi \vee CK\neg\varphi)$   
from (2) by  $Def2(CK)$ , thm.  $Def1(CKif)$
- 

It follows from propositions 4.3 and 4.4 that the first two parts of Table 4.3 provide a sound and complete axiomatization for the fragment of the language with only ‘knowing whether’ operators.

**Proposition 4.5.** *If formula  $\varphi$  has no  $K_i$ ,  $EK$ ,  $CK$  operators then  $\varphi$  is a theorem of the S5-based  $GFP2$  axiomatics of Table 4.3 if and only if it is provable without the axioms  $Def2(K_i)$ ,  $Def2(EK)$ , and  $Def2(CK)$ .*

*Proof.* Suppose no  $K_i$ ,  $EK$ ,  $CK$  occur in  $\varphi$  and suppose  $\varphi$  is a theorem of the S5-based  $GFP2$  axiomatics. Whenever the proof of  $\varphi$  uses axiom  $Def2(K_i)$ ,  $Def2(EK)$ , or  $Def2(CK)$ , we can eliminate that axiom by replacing the *definiendum* by the *definiens* everywhere in the proof. □



### 4.3 Discussion: epistemic logics weaker than S5

#### 4.3.1 Epistemic logics between KT and S5

We have seen that our new axiom  $GFP0$  is sound for logics of knowledge, understood as logics where the logic of individual knowledge is at least KT, and that it is complete when the logic of individual knowledge is S5.

We conjecture that the KT-based  $GFP0$  axiomatics is incomplete. We however do not have a formal proof for the time being. Such a proof would have to delve into semantics: it typically consists in designing a non-standard semantics for which the axiomatics with  $GFP0$  is complete. We leave this aside for the time being.

Under the hypothesis that the KT-based  $GFP0$  axiomatics is incomplete, one may wonder which axiom is missing to obtain completeness. A tempting avenue is to add the formula  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  of Lemma 4.1 as an axiom schema to the axiomatics of Table 4.2. The proof of Proposition 4.2 then gives us completeness because it uses none of the S5 axioms but  $T(K_i)$ . However it can be shown that this amounts to adding  $5(CK)$ : it can be shown that the formula is equivalent to  $5(CK)$  in the presence of  $T(CK)$ .

**Proposition 4.6.** *In the GFP-based axiomatics for KT,  $5(CK)$  and the formula  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  are interderivable.*

*Proof.* From the GFP-based axiomatics for KT and  $5(CK)$  (recall that  $4(CK)$  is derivable from  $FP'$ ,  $RN(CK)$  and  $GFP$ ):

1.  $CKCK(\varphi \rightarrow EK\varphi) \rightarrow CK(\varphi \rightarrow CK\varphi)$  from  $GFP$ ,  $RN(CK)$ ,  $K(CK)$
2.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CKCK(\varphi \rightarrow EK\varphi)$   $4(CK)$
3.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CKCK(\neg CK\varphi \rightarrow \neg\varphi)$  from (1), (2) and  $4(CK)$
4.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(CK\neg CK\varphi \rightarrow CK\neg\varphi)$  from (3) and  $K(CK)$
5.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg CK\varphi \rightarrow CK\neg\varphi)$  from (4) and  $5(CK)$
6.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  from (5),  $FP'$  and  $T(CK)$

From the GFP-based axiomatics for KT and  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$ :

1.  $CK(\varphi \rightarrow EK\varphi) \rightarrow CK(\neg\varphi \rightarrow EK\neg\varphi)$  hypothesis
2.  $CK(CK\varphi \rightarrow EKCK\varphi)$  from  $FP'$  and  $RN(CK)$
3.  $CK(\neg CK\varphi \rightarrow EK\neg CK\varphi)$  from (2) and (1)
4.  $\neg CK\varphi \rightarrow CK\neg CK\varphi$  from (3) and  $GFP$

□

Just as common knowledge is necessarily positively introspective even when individual knowledge isn't, it can still be argued that S5 common knowledge can make sense even when individual knowledge is not S5: one can imagine, e.g., that common knowledge is “written on a blackboard”, or otherwise easily available to agents such that they are able to immediately verify what is and is not commonly known. We leave further explorations to future work.

### 4.3.2 *GFP0* is not appropriate for belief

Up to now we have only discussed common knowledge; we now briefly discuss common belief.

Let us write  $B_i\varphi$  for “ $i$  believes that  $\varphi$ ”,  $EB\varphi$  for “it is shared belief that  $\varphi$ ”, and  $CB\varphi$  for “it is common belief that  $\varphi$ ”, and let us suppose the logic of the  $B_i$  operators is KD (or, alternatively, any logic without the  $T$  axiom).

It is intuitively clear that the belief-version of *GFP1*,

$$CB \bigwedge_{i \in \text{Agt}} (B_i\varphi \vee B_i\neg\varphi) \rightarrow (CB\varphi \vee CB\neg\varphi),$$

should not hold: if there is common belief—and even common knowledge—that everybody has an opinion about  $\varphi$  then it by no means follows that there is common belief about  $\varphi$ .

What about *GFP0*? The fact that *GFP1* is unintuitive need not disqualify *GFP0*. Indeed, while these two axioms are equivalent in epistemic contexts, they fail to be so in doxastic contexts: in KD45,  $\bigwedge_{i \in \text{Agt}} (B_i\varphi \vee B_i\neg\varphi)$  does not imply  $EB\varphi \vee EB\neg\varphi$ , and does not do so a fortiori in KD; and therefore the belief-counterpart of Proposition 2.2 does not hold.

As it turns out, *GFP0* is not a reasonable principle of common belief either. This can be highlighted by the following example. Suppose that the set of agents under concern is  $\text{Agt} = \{1, 2\}$  and that there is a misunderstanding between 1 and 2 about an inform act of a third agent. That third agent is not relevant here, and we suppose that  $\text{Agt} = \{1, 2\}$ . Let us suppose that 1 believes the third agent said  $p$  and therefore believes that  $p$  is in the common ground ( $B_1CBp$ ), while 2 believes that  $\neg p$  is in the common ground ( $B_2CB\neg p$ ). It follows by  $\wp(CB)$  and by the (intuitively still valid) belief-counterpart of *FP0* that

$$B_1CBEBp \wedge B_2CBEB\neg p.$$

As both  $CB$  and  $EB$  are normal operators, it follows that

$$B_1CB(EBp \vee EB\neg p) \wedge B_2CB(EBp \vee EB\neg p),$$

i.e., that  $EBCB(EBp \vee EB\neg p)$ . The latter is equivalent to  $CB(EBp \vee EB\neg p)$  thanks to the belief-version of the fixed-point axiom, which is  $CB\varphi \leftrightarrow EBCB\varphi$ . From that the counter-intuitive consequence  $CBp \vee CB\neg p$  would follow by the belief-counterpart of *GFP0*.

To sum up, unlike the standard induction principles the status of our new versions of the induction axiom differs between knowledge and belief: they are specific to common knowledge and fail for common belief.

## 4.4 A second version of EL-O

We introduce in this section a new version of EL-O, called EL-O+*Ind*, which we will show to be a fragment of the logic S5. This is achieved on the semantic side by restricting states to particular sets of atoms, and on the syntactic side by adding an axiom to the logic inspired by *GFP2*.

### 4.4.1 Language and semantics of EL-O+Ind

The language of EL-O+Ind is exactly the language of EL-O. In Chapter 3, EL-O states are allowed to be arbitrary sets of atoms. We will for EL-O+Ind be considering only states that are members of the following set:

$$\begin{aligned} IND\text{-STATES} = \{s \subseteq ATM : & \text{for all } \alpha \in ATM, \text{ if for all } i \in Agt \\ & S_i \alpha \in s \cup I\text{-ATM and } JS S_i \alpha \in s \cup I\text{-ATM} \\ & \text{and } JS \alpha \notin I\text{-ATM then } JS \alpha \in s\} \end{aligned}$$

This guarantees the validity of the following principle: if for some atom  $\alpha$ ,  $JS S_i \alpha$  and  $S_i \alpha$  are true for all  $i \in Agt$ , then  $JS \alpha$  should also be true.

In the rest of this chapter we will only be using the term “states” to refer to members of *IND-STATES*, and will be referring to arbitrary subsets of *ATM* simply as sets of atoms.

**Example 4.1** (Letter, cont.). Suppose that both agents read the letter separately, then tell each other that they have read the letter. Then they both know the contents of the letter, and they jointly see that each of them knows whether  $p$  is true:  $S_1 p$ ,  $S_2 p$ ,  $JS S_1 p$  and  $JS S_2 p$  are all true. But then  $JS p$  should also be true: it is the case that both agents know whether  $p$  is true, both know that they both know it, and so on. For this reason,  $\{S_1 p, S_2 p, JS S_1 p, JS S_2 p\}$  is not a state, but  $\{S_1 p, S_2 p, JS S_1 p, JS S_2 p, JS p\}$  is.

We first state a few basic properties of *IND-STATES*.

**Proposition 4.7.** *IND-STATES* is closed under intersection: if two states  $s$  and  $s'$  are in *IND-STATES* then  $s \cap s'$  is also in *IND-STATES*.

**Proposition 4.8.** If  $s \in IND\text{-STATES}$  then  $s^{\Rightarrow} \cup I\text{-ATM} \in IND\text{-STATES}$ .

We call *IND-STATES<sub>F</sub>* the set of all finite states. The set of all introspectively closed states is

$$I\text{-IND-STATES} = \{s^{\Rightarrow} \cup I\text{-ATM} : s \in IND\text{-STATES}\}.$$

The truth conditions for formulas in EL-O+Ind are then just as in EL-O:

$$\begin{aligned} s \models_{EL\text{-O+Ind}} \alpha & \quad \text{iff} \quad \alpha \in s^{\Rightarrow} \cup I\text{-ATM} \\ s \models_{EL\text{-O+Ind}} \neg\varphi & \quad \text{iff} \quad s \not\models_{EL\text{-O+Ind}} \varphi \\ s \models_{EL\text{-O+Ind}} \varphi \wedge \psi & \quad \text{iff} \quad s \models_{EL\text{-O+Ind}} \varphi \text{ and } s \models_{EL\text{-O+Ind}} \psi \end{aligned}$$

where  $s \in IND\text{-STATES}$ . In the rest of this chapter we will use  $\models$  as a shorthand for  $\models_{EL\text{-O+Ind}}$  (rather than as a shorthand for  $\models_{EL\text{-O}}$  as we do in the other chapters).

In the rest of this section we establish a finite model property for EL-O+Ind, prove NP-completeness of EL-O+Ind satisfiability, and give an axiomatization.

### 4.4.2 Completion of a finite state and finite model property

The finite model property for EL-O was obtained by considering, given an EL-O state  $s$  and a formula  $\varphi$ , the state  $(s^{\Rightarrow} \cup I\text{-ATM}) \cap \text{Voc}(\varphi)$ . However, given an

$A_0 = A$	$A_1$	$A_2 = C(A)$
$\overline{JS S_1 S_2 p}, (\overline{JS S_2 S_2 p}),$ $\overline{S_1 S_2 p}, (\overline{S_2 S_2 p}),$ $\overline{JS S_1 p},$ $S_1 p, S_2 p$	$\overline{JS S_1 S_2 p},$ $\overline{S_1 S_2 p},$ $\overline{JS S_1 p}, \mathbf{JS S_2 p},$ $\overline{S_1 p}, \overline{S_2 p}$	$\overline{JS S_1 S_2 p},$ $\overline{S_1 S_2 p},$ $\overline{JS S_1 p}, \overline{JS S_2 p},$ $\overline{S_1 p}, \overline{S_2 p},$ $\mathbf{JS p}$

Table 4.4: Construction of the completion of  $A$  in Example 4.2. The atoms added at each step are in bold, the atoms that cause these additions are underlined. The atoms that are parenthesized are the introspective atoms not explicitly in  $A$ .

EL-O+*Ind* state  $s$  and a formula  $\varphi$ , it is not always the case that  $s \cap \text{Voc}(\varphi)$  is still in *IND-STATES*. A bit more work is necessary to obtain a finite model property for EL-O+*Ind*. We first show how to obtain for any set of atoms  $A$  the smallest state containing  $A$ .

**Definition 4.1.** Let  $A \subseteq \text{ATM}$  be an arbitrary set of atoms. We define the completion of  $A$ , noted  $C(A)$ , as follows:

$$\begin{aligned}
A_0 &= A \\
A_{k+1} &= A_k \cup \{JS \alpha : S_i \alpha, JS S_i \alpha \in A_k \cup I\text{-ATM} \text{ for all } i \in \text{Agt} \\
&\quad \text{and } JS \alpha \notin A_k \cup I\text{-ATM}\} \\
C(A) &= \bigcup_{k \in \mathbb{N}} A_k
\end{aligned}$$

**Example 4.2.** Let  $\text{Agt} = \{1, 2\}$  and  $A = \{S_1 p, S_2 p, JS S_1 p, S_1 S_2 p, JS S_1 S_2 p\}$ . Then  $S_1 S_2 p, JS S_1 S_2 p \in A$ ,  $S_2 S_2 p, JS S_2 S_2 p \in I\text{-ATM}$ , and  $JS S_2 p \notin I\text{-ATM}$ . Therefore  $A_1 = A \cup \{JS S_2 p\}$ . But then  $S_1 p, S_2 p, JS S_1 p, JS S_2 p \in A_1$ , and  $JS p \notin I\text{-ATM}$ , hence  $A_2 = A_1 \cup \{JS p\}$ . No more atoms can be added after this step:  $C(A) = A_2$ . See Table 4.4 for a more visual breakdown of the operation.

We now show that this completion process ends in a finite number of steps for any finite set of atoms  $A$ .

**Proposition 4.9.** Let  $A$  be a finite set of atoms and  $\ell$  be the length of the longest atom in  $A$ . If  $\ell < 3$  then for all  $k \in \mathbb{N}$ ,  $A_k = A$ . Otherwise for any  $k \geq \ell - 2$ ,  $A_k = A_{k+1}$ . That is,  $C(A) = A$  if  $\ell < 3$  and  $C(A) = A_{\ell-2}$  if  $\ell \geq 3$ .

*Proof.* If  $\ell < 3$  then there can be no atoms in  $A$  of the form  $JS S_i \alpha$ , hence the first result.

When  $\ell \geq 3$ , we show by induction on  $k$  that the longest atom in  $A_k \setminus A_{k-1}$  is of length no more than  $\ell - k$  for any  $k \geq 1$ . Any atom  $JS \alpha$  in  $A_1 \setminus A$  must be such that  $JS S_i \alpha \in A$  for some  $i$ , and therefore the length of  $JS \alpha$  is at most  $\ell - 1$ . From there on, for any atom  $JS \alpha$  in  $A_{k+1} \setminus A_k$ , it must be the case that  $JS S_i \alpha \in A_k \setminus A_{k-1}$  for some  $i \in \text{Agt}$  (otherwise  $JS \alpha$  would be in  $A_k$ ). Therefore the length of  $JS S_i \alpha$  must be no more than  $\ell - k$  by the induction hypothesis, and the length of  $JS \alpha$  is therefore at most  $\ell - k - 1$ .

Hence atoms in  $A_{\ell-2} \setminus A_{\ell-3}$  are of length at most 2, that is, the only atoms that can be in  $A_{\ell-2} \setminus A_{\ell-3}$  are of the form  $JS p$  with  $p \in \text{Prop}$ . As these are not of the form  $JS S_i \alpha$ , no more atoms can be added from  $A_{\ell-1}$  on.  $\square$

It follows that if  $A$  is finite then  $C(A)$  is also finite. Moreover, clearly  $C(A) \in \text{IND-STATES}$  for any  $A$ .

**Proposition 4.10.** *Let  $A \subseteq \text{ATM}$  be an arbitrary set of atoms. If  $A$  is finite then  $C(A) \in \text{IND-STATES}_F$ .*

As before, we say that two states  $s$  and  $s'$  agree on the set of atoms  $A \subseteq \text{ATM}$  when they assign the same truth value to every element of  $A$ , i.e., when for every  $\alpha \in A$ ,  $s \models \alpha$  iff  $s' \models \alpha$ . By a straightforward induction on the structure of formulas we get:

**Proposition 4.11.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$  be a formula and  $s, s' \in \text{IND-STATES}$  two states agreeing on  $\text{Voc}(\varphi)$ . Then  $s \models \varphi$  iff  $s' \models \varphi$ .*

With this we obtain the finite model property for  $\text{EL-O}+\text{Ind}$ :

**Proposition 4.12** (Finite model property). *For any formula  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$ , if  $s \models \varphi$  for some  $s \in \text{IND-STATES}$  then there exists a finite state  $s' \in \text{IND-STATES}_F$  such that  $s' \models \varphi$ .*

*Proof.* Let  $s \in \text{IND-STATES}$  be such that  $s \models \varphi$ . We first remark that as  $\text{Voc}(\varphi)$  is finite, so is  $C(\text{Voc}(\varphi))$  by Proposition 4.10. Using Propositions 4.7 and 4.8, we obtain that  $(s^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi)) \in \text{IND-STATES}_F$ . It is clear that  $s$  and  $s^{\Rightarrow} \cup I\text{-ATM}$  agree on all atoms. Then  $s$  and  $(s^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi))$  agree on  $\text{Voc}(\varphi)$ . Therefore by Proposition 4.11,  $s$  and  $(s^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi))$  agree on  $\varphi$ .  $\square$

### 4.4.3 Complexity of $\text{EL-O}+\text{Ind}$ satisfiability

The  $\text{EL-O}+\text{Ind}$  satisfiability problem consists in deciding, given a formula  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$ , whether there exists a  $s \in \text{IND-STATES}$  such that  $s \models \varphi$ . In this section we analyze the complexity of this problem. We start by reducing  $\text{EL-O}+\text{Ind}$  model checking to classical model checking.

**Lemma 4.2.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$  be a formula and  $s \in \text{IND-STATES}$  be a state. Then  $s \models \varphi$  iff  $s^{\Rightarrow} \cup I\text{-ATM} \models^{\text{CPC}} \varphi$ .*

**Proposition 4.13.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$  be a formula. Then  $\varphi$  is  $\text{EL-O}$ -satisfiable iff there exists a set of atoms  $s \subseteq C(\text{Voc}(\varphi))$  such that  $s$  is in  $\text{IND-STATES}_F$ ,  $s = (s^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi))$  and  $s \models^{\text{CPC}} \varphi$ .*

*Proof.* The left-to-right direction follows from Proposition 4.12 and Lemma 4.2: if there exists  $s \in \text{IND-STATES}$  such that  $s \models \varphi$ , consider  $s' = (s^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi))$ . Then  $s' \in \text{IND-STATES}_F$ ,  $s' = (s'^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi))$ , and  $s' \models \varphi$  by the proof of Proposition 4.12. By Lemma 4.2,  $s'^{\Rightarrow} \cup I\text{-ATM} \models^{\text{CPC}} \varphi$ , hence  $(s'^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi)) \models^{\text{CPC}} \varphi$ , that is,  $s' \models^{\text{CPC}} \varphi$ .

Suppose now that there exists  $s \subseteq C(\text{Voc}(\varphi))$  such that  $s \in \text{IND-STATES}_F$ ,  $s = (s^{\Rightarrow} \cup I\text{-ATM}) \cap C(\text{Voc}(\varphi))$  and  $s \models^{\text{CPC}} \varphi$ . Then  $s^{\Rightarrow} \cup I\text{-ATM} \models^{\text{CPC}} \varphi$  and by Lemma 4.2,  $s \models \varphi$ .  $\square$

In order to determine complexity of satisfiability in  $\text{EL-O}+\text{Ind}$ , we must analyze the complexity of calculating  $C(\text{Voc}(\varphi))$  for any formula  $\varphi$ .

**Lemma 4.3.** *Let  $A \subseteq ATM$  be a finite set of atoms. Let  $\ell$  be the length of the longest atom in  $A$ , and  $|A|$  be the number of atoms of  $A$ . Then the completion  $C(A)$  of  $A$  can be constructed in time  $O(|A|^2 \ell^3)$ .*

*Proof.* The algorithm follows the construction of  $A_k$  given in Definition 4.1, and stops when a fixpoint is reached, after at most  $\ell - 2$  steps by Proposition 4.9. Finding which atoms must be added to construct  $A_{k+1}$  can be done in  $O(|A_k|^2)$  operations. Moreover, the number of atoms added in  $A_1$  is no more than  $\frac{|A|}{2}$  (more precisely, it is no more than  $\frac{|A|}{2|A_{gt}|}$ ), and as any atom added from there on must be on the basis of an atom added at the previous step, no more than  $\frac{|A|}{2}$  atoms can be added at each step. Therefore for any  $k \leq \ell - 2$ ,  $|A_k| \leq |A|(1 + \frac{\ell-2}{2})$ , that is,  $|A_k| \leq \frac{1}{2}|A|\ell$ . Therefore each step is done in time  $O(|A|^2 \ell^2)$ . As there are at most  $\ell - 2$  such steps, the entire construction can be done in time  $O(|A|^2 \ell^3)$ .  $\square$

**Proposition 4.14.** *The satisfiability problem for EL-O+Ind is NP-complete.*

*Proof.* Hardness is the case because CPC can be embedded into EL-O+Ind, just as with EL-O. For membership, let  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  be a formula. The algorithm is as follows:

1. Construct  $C(\text{Voc}(\varphi))$  (in  $O(\ell(\varphi)^5)$ ). Note that  $|\text{Voc}(\varphi)| \leq \ell(\varphi)$  and the longest atom of  $\text{Voc}(\varphi)$  is of length at most  $\ell(\varphi)$ , hence  $|C(\text{Voc}(\varphi))| \leq \ell(\varphi)^2$ .
2. Guess a set  $s \subseteq C(\text{Voc}(\varphi))$ .
3. Check that  $s \in \text{IND-STATES}$  (in  $O(|s|^2)$ , i.e.  $O(\ell(\varphi)^4)$  because  $|s| \leq |C(\text{Voc}(\varphi))| \leq \ell(\varphi)^2$ ).
4. Check that for any  $\alpha \in s$  and  $\beta \in C(\text{Voc}(\varphi))$ , if  $\alpha \Rightarrow \beta$  then  $\beta \in s$  and if  $\beta \in I\text{-ATM}$  then  $\beta \in s$  (in  $O(|s||C(\text{Voc}(\varphi))|)$ , i.e.  $O(\ell(\varphi)^4)$ ).
5. Check that  $s \models^{\text{CPC}} \varphi$ .

$\square$

Hence the satisfiability problem in EL-O+Ind has the same complexity as its EL-O and classical counterparts.

#### 4.4.4 Axiomatization of EL-O+Ind validities

The EL-O+Ind validities of our language  $\mathcal{L}_{\text{bool}}(ATM)$  are axiomatized by the schemas of Table 4.5 together with CPC. Notice the new axiom *Ind* compared to the axiomatization of EL-O. This axiom is a direct translation of the right-to-left direction of the induction axiom for common knowledge whether *GFP2* discussed in Section 4.2.

**Proposition 4.15.** *Let  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$  be a formula. Then  $\varphi$  is EL-O+Ind valid iff  $\varphi$  is provable in CPC from the axiom schemas  $\text{Vis}_1 - \text{Vis}_5 + \text{Ind}$  of Table 4.5.*

Table 4.5: Axioms for EL-O

$Vis_1$	$S_i S_i \alpha$
$Vis_2$	$JS JS \alpha$
$Vis_3$	$JS S_i S_i \alpha$
$Vis_4$	$JS \alpha \rightarrow S_i \alpha$
$Vis_5$	$JS \alpha \rightarrow JS S_i \alpha$
$Ind$	$\left( \bigwedge_{i \in Agt} (JS S_i \alpha \wedge S_i \alpha) \right) \rightarrow JS \alpha$

*Proof.* We again take advantage of Lemma 4.2 and show that  $Vis_1 - Vis_5 + Ind$  characterise the set of introspectively closed states  $I-IND-STATES = \{s^{\Rightarrow} \cup I-ATM : s \in IND-STATES\}$ . We first remark that as  $I-IND-STATES = I-STATES \cap IND-STATES$ , the axiom schemas  $Vis_1 - Vis_5$  must be valid in all introspective states. Moreover, we know that any set of atoms satisfying  $Vis_1 - Vis_5$  must be in  $I-STATES$ . Only two elements are left to prove: first, the axiom schema  $Ind$  is valid in all introspectively closed states. Second, any set of atoms satisfying  $Ind$  must be a member of  $IND-STATES$ . Both of these elements are straightforward.  $\square$

## 4.5 EL-O+ $Ind$ as a fragment of epistemic logic S5

As in Chapter 3, we identify  $S_i$  with  $Kif_i$  and  $JS$  with  $CKif$ , and consider the language  $\mathcal{L}_{bool}(ATM)$  as the EL-O fragment of  $Fml_{EL}^{if}$ . In this section we show that standard S5 epistemic logic is a conservative extension of EL-O+ $Ind$ , that is, that Table 4.5 axiomatizes the validities of S5 over the EL-O fragment of the full language  $Fml_{EL}$ . It follows by Proposition 4.14 that our lightweight epistemic logic is a fragment of standard epistemic logic with an NP-complete satisfiability problem.

### 4.5.1 A canonical model for EL-O+ $Ind$

As in Chapter 3, we define a particular Kripke model  $M^{EL-O+Ind}$  in which worlds are EL-O+ $Ind$  valuations, starting with defining equivalence relations between states. For every  $s, s' \in IND-STATES$  we define:

$$sR_i^{EL-O+Ind}s' \text{ iff } s \text{ and } s' \text{ agree on every } \alpha \text{ such that } s \models S_i \alpha;$$

$$sR_{Agt}^{EL-O+Ind}s' \text{ iff } s \text{ and } s' \text{ agree on every } \alpha \text{ such that } s \models JS \alpha.$$

As before, all relations  $R_i^{EL-O+Ind}$  and  $R_{Agt}^{EL-O+Ind}$  are equivalence relations, and all  $R_i^{EL-O+Ind}$  are included in  $R_{Agt}^{EL-O+Ind}$ .

We then define the *canonical Kripke model for EL-O+ $Ind$*  as the tuple

$M^{\text{EL-O}+Ind} = \langle W^{\text{EL-O}+Ind}, \sim_{Agt}^{\text{EL-O}+Ind}, \{\sim_i^{\text{EL-O}+Ind}\}_{i \in Agt}, V^{\text{EL-O}+Ind} \rangle$  with

$$\begin{aligned} W^{\text{EL-O}+Ind} &= \{s \Rightarrow \cup I-ATM : s \in IND-STATES_F\}, \\ \sim_{Agt}^{\text{EL-O}+Ind} &= R_{Agt}^{\text{EL-O}+Ind} \cap (W^{\text{EL-O}+Ind} \times W^{\text{EL-O}+Ind}), \\ \sim_i^{\text{EL-O}+Ind} &= R_i^{\text{EL-O}+Ind} \cap (W^{\text{EL-O}+Ind} \times W^{\text{EL-O}+Ind}), \\ V^{\text{EL-O}+Ind}(w) &= w \cap Prop \text{ for every } w \in W^{\text{EL-O}+Ind}. \end{aligned}$$

Hence the possible worlds of  $M^{\text{EL-O}}$  are the introspective closures of finite states. We know it suffices to consider finite models due to the finite model property proved in Section 4.4.2; the fact that we are considering closures of finite states rather than arbitrary states will be instrumental in our proofs.

**Proposition 4.16.** *For any  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$ , for any  $s \in W^{\text{EL-O}+Ind}$ ,  $s \models \varphi$  iff  $M^{\text{EL-O}+Ind}, s \models \varphi$ .*

*Proof.* This is shown by induction on the form of  $\varphi$ . It suffices to show by induction on  $\alpha$  that for all  $s \in W^{\text{EL-O}+Ind}$ , for all  $\alpha \in ATM$ ,  $s \models \alpha$  iff  $M^{\text{EL-O}+Ind}, s \models \alpha$ .

By definition of  $M^{\text{EL-O}+Ind}$ , for any  $p \in Prop$  and any  $s$ ,  $M^{\text{EL-O}+Ind} \models p$  iff  $s \models p$ , and for any atom  $\alpha$  and agent  $i$ , if  $s \models S_i \alpha$  then  $M^{\text{EL-O}+Ind}, s \models S_i \alpha$  and if  $s \models JS \alpha$  then  $M^{\text{EL-O}+Ind}, s \models JS \alpha$ .

If  $s \not\models S_i \alpha$ , consider  $s' = s \cup \{\alpha\}$  if  $\alpha \notin s$  and  $s' = s \setminus \{\alpha\}$  if  $\alpha \in s$ . We show that  $s' \in W^{\text{EL-O}+Ind}$ . Because  $s \in W^{\text{EL-O}+Ind}$ , there exists  $s_0 \in IND-STATES_F$  such that  $s = s_0 \Rightarrow \cup I-ATM$ . Suppose that  $s \models \alpha$  and  $\alpha \notin s_0$ . Then either  $\alpha \in I-ATM$  or there exists  $\beta \in s_0$  such that  $\beta \Rightarrow \alpha$ . In the first case,  $S_i \alpha$  must also be introspective, and in the second,  $\beta \Rightarrow S_i \alpha$ , hence in both cases  $s \models S_i \alpha$ . As this is not possible, we conclude that  $s \models \alpha$  iff  $\alpha \in s_0$ . Consider then  $s'_0 = s_0 \cup \{\alpha\}$  if  $\alpha \notin s_0$  and  $s'_0 = s_0 \setminus \{\alpha\}$  if  $\alpha \in s_0$ . As  $s \not\models S_i \alpha$ ,  $\alpha$  can not be of the form  $JS \alpha'$ , and therefore  $s' = s'_0 \Rightarrow \cup I-ATM$ . We now show that  $s'_0 \in IND-STATES_F$ . Suppose that there exists  $\beta$  such that  $S_j \beta \in s'_0$  and  $JS S_j \beta \in s'_0$  for all  $j \in Agt$ , and  $JS \beta \notin s'_0$ . Necessarily  $\alpha = S_j \beta$  for some  $j$ . But then  $JS \alpha \in s_0$ , and therefore  $S_i \alpha \in s$ , which is impossible. Therefore  $s'_0 \in IND-STATES_F$ . We have constructed  $s' \in W^{\text{EL-O}+Ind}$  such that  $s$  and  $s'$  disagree on  $\alpha$  (and therefore by the induction hypothesis,  $M^{\text{EL-O}+Ind}, s$  and  $M^{\text{EL-O}+Ind}, s'$  also disagree on  $\alpha$ ), and clearly  $s \sim_i^{\text{EL-O}+Ind} s'$ . Therefore  $M^{\text{EL-O}+Ind}, s \not\models S_i \alpha$ .

If  $s \not\models JS \alpha$ , we take the same  $s'$  and show in the same way that  $s' \in W^{\text{EL-O}+Ind}$ . Then  $s$  and  $s'$  disagree on  $\alpha$  and  $s \sim_{Agt}^{\text{EL-O}+Ind} s'$ , and therefore  $M^{\text{EL-O}+Ind}, s \not\models JS \alpha$ .  $\square$

**Corollary 4.1.** *For any formula  $\varphi \in \mathcal{L}_{\text{bool}}(ATM)$ ,  $\models_{\text{EL-O}+Ind} \varphi$  iff  $\models_{M^{\text{EL-O}+Ind}} \varphi$ .*

*Proof.* This is shown by combining the above result with the finite model property of Proposition 4.12: if  $M^{\text{EL-O}+Ind}, s \models \varphi$  for some  $s \in W^{\text{EL-O}+Ind}$ , then  $s \models \varphi$ ; and if  $s \models \varphi$  for some  $s \in IND-STATES$ , then by Proposition 4.12 there exists  $s' \in IND-STATES_F$  such that  $s' \models \varphi$ , therefore  $s' \Rightarrow \cup I-ATM \models \varphi$ , and we conclude by Proposition 4.16 that  $M^{\text{EL-O}+Ind}, s' \Rightarrow \cup I-ATM \models \varphi$ .  $\square$



### 4.5.2 Completeness of the EL-O+Ind fragment w.r.t. S5 validity

We now show that the EL-O+Ind semantics in terms of observability and the standard S5 Kripke semantics have the same validities as far as the EL-O+Ind fragment is concerned. It is clear that the EL-O+Ind axioms are valid in S5. For the other direction, we could attempt to show that  $M^{\text{EL-O+Ind}}$  is an S5 Kripke model. We here take a different approach and instead show that it is at least bisimilar to an S5 model.

We define the model  $M_0$  as follows:  $M_0 = \langle W^{\text{EL-O+Ind}}, \sim_{\text{Agt}}, \{\sim_i^{\text{EL-O+Ind}}\}_{i \in \text{Agt}}, V^{\text{EL-O+Ind}} \rangle$  where  $\sim_{\text{Agt}} = (\bigcup_{i \in \text{Agt}} \sim_i^{\text{EL-O+Ind}})^+$ . That is,  $\sim_{\text{Agt}}$  is the transitive closure of the union of all  $\sim_i^{\text{EL-O+Ind}}$ . Clearly  $M_0$  is an S5 Kripke model. To show that validities of  $\mathcal{L}_{\text{bool}}(\text{ATM})$  are preserved, we will need the following lemma:

**Lemma 4.4.** *Let  $M = \langle W, R_{\text{Agt}}, \{R_i\}_{i \in \text{Agt}}, V \rangle$  be a Kripke model such that  $R_{\text{Agt}}$  and all  $R_i$  are equivalence relations, and let  $s$  be a world in that model. If there exists  $\alpha$  and a finite sequence  $i_1, \dots, i_n$  of agents such that  $(M, s) \not\models S_{i_n} \dots S_{i_1} \alpha$ , then there exists a subsequence  $i_{l_1}, \dots, i_{l_p}$  of  $i_1, \dots, i_n$  and a world  $s'$  in  $M$  such that  $s R_{i_{l_p}} \dots R_{i_{l_1}} s'$  and  $(M, s)$  and  $(M, s')$  disagree on  $\alpha$ .*

*Proof.* This is shown by induction on  $n$ . If  $n = 1$  and  $s \not\models S_{i_1} \alpha$ , then by definition of Kripke models there exists  $s'$  such that  $s R_{i_1} s'$  and  $s$  and  $s'$  disagree on  $\alpha$ .

Suppose that the property is true for all  $k \leq n$  for some  $n$ . Suppose  $s \not\models S_{i_n} \dots S_{i_1} \alpha$ . Let  $m = \min\{k \leq n : \forall l \geq k, s \not\models S_{i_l} \dots S_{i_1} \alpha\}$ . If  $m = 1$  then  $s \not\models S_{i_1} \alpha$  and we are back at the first case. Otherwise we have  $s \not\models S_{i_m} \dots S_{i_1} \alpha$  and  $s \models S_{i_{m-1}} \dots S_{i_1} \alpha$ . Therefore there exists a world  $s_m$  such that  $s R_{i_m} s_m$  and  $s$  and  $s_m$  disagree on  $S_{i_{m-1}} \dots S_{i_1} \alpha$ , that is,  $s_m \not\models S_{i_{m-1}} \dots S_{i_1} \alpha$ . If  $s$  and  $s_m$  disagree on  $\alpha$  we are done. Otherwise the induction hypothesis tells us that there exists a subsequence  $i_{l_1}, \dots, i_{l_p}$  of  $i_1, \dots, i_{m-1}$  and a state  $s'$  such that  $s_m R_{i_{l_p}} \dots R_{i_{l_1}} s'$  and  $s_m$  and  $s'$  disagree on  $\alpha$ . Therefore  $s R_{i_m} s_m R_{i_{l_p}} \dots R_{i_{l_1}} s'$  and  $s$  and  $s'$  disagree on  $\alpha$ .  $\square$

We are now ready to prove that  $M^{\text{EL-O+Ind}}$  and  $M_0$  are bisimilar for formulas of  $\mathcal{L}_{\text{bool}}(\text{ATM})$ .

**Proposition 4.17.** *For any formula  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$ , for any  $s \in W^{\text{EL-O+Ind}}$ ,  $M^{\text{EL-O+Ind}}, s \models \varphi$  iff  $M_0, s \models \varphi$ .*

*Proof.* This is shown by induction on  $\varphi$ . The only interesting case is when  $\varphi = JS \alpha$  and  $M^{\text{EL-O+Ind}}, s \not\models JS \alpha$ . Let us show that in this case there is necessarily a sequence of agents  $i_1, \dots, i_n$  such that  $s \sim_{i_1}^{\text{EL-O+Ind}} \dots \sim_{i_n}^{\text{EL-O+Ind}} s'$  and  $s$  and  $s'$  disagree on  $\alpha$ , and therefore  $M_0, s \not\models JS \alpha$ .

If  $M^{\text{EL-O+Ind}}, s \not\models JS \alpha$  then  $s \not\models JS \alpha$  by Proposition 4.16. As  $s$  is in IND-STATES, it must be the case that either  $s \not\models S_{i_1} \alpha$  for some agent  $i_1$ , or  $s \models S_i \alpha$  for all  $i \in \text{Agt}$  and  $s \not\models JS S_{i_1} \alpha$  for some agent  $i_1$ . In the second case, either  $s \not\models S_{i_2} S_{i_1} \alpha$  for some agent  $i_2$ , or  $s \models S_i S_{i_1} \alpha$  for all  $i \in \text{Agt}$  and  $s \not\models JS S_{i_2} S_{i_1} \alpha$  for some  $i_1$ . Iterating this, we get that either  $s \not\models S_{i_n} \dots S_{i_1} \alpha$  for some sequence  $i_n, \dots, i_1$  of agents, or there exists an infinite sequence  $(i_k)_{k \in \mathbb{N}}$  of agents such that for all  $k$ ,  $s \models S_i S_{i_k} \dots S_{i_1} \alpha$  for all  $i \in \text{Agt}$  and  $s \not\models JS S_{i_k} \dots S_{i_1} \alpha$ . In the

first case the previous lemma gives us the desired result. It remains to show that the second case cannot happen.

As  $s \in W^{\text{EL-O}+Ind}$ , we know that there exists a finite state  $s_0$  such that  $s = s_0^{\Rightarrow} \cup I\text{-ATM}$ . Let  $M = \max\{k \mid S_{i_k} \dots S_{i_1} \alpha \in s_0\}$ . Then  $S_{i_{M+1}} \dots S_{i_1} \alpha \in s_0^{\Rightarrow} \setminus s_0$ . Therefore there exists  $\beta \in s_0$  such that  $\beta \Rightarrow S_{i_{M+1}} \dots S_{i_1} \alpha$ . But then  $\beta \Rightarrow JS S_{i_{M+1}} \dots S_{i_1} \alpha$ , which contradicts the fact that  $s \not\models JS S_{i_{M+1}} \dots S_{i_1} \alpha$ . Hence this second case is impossible. We conclude that  $M_0, s \not\models JS \alpha$ .  $\square$

With this we conclude that  $M^{\text{EL-O}+Ind}$  is bisimilar to an S5 Kripke model. This means that all S5 validities in the EL-O fragment of  $Fml_{\text{EL}}^{if}$  are valid in  $M^{\text{EL-O}+Ind}$  and therefore in  $\text{EL-O}+Ind$ , giving us our completeness result.  $\text{EL-O}+Ind$  is therefore a fragment of standard S5.

## 4.6 Conclusion

We have studied the axiomatization of the logic of common knowledge, coming up with an alternative  $GFP0$  to the standard induction axiom principles that is intuitively appealing as an axiom for common knowledge. While our proofs are not very difficult, we believe that  $GFP0$  will lead to presentations of epistemic logic that are intuitively more appealing. We have then used the alternative presentation of  $GFP0$  with ‘knowing whether’ operators,  $GFP2$ , to inspire a version of EL-O which follows the induction axiom for common knowledge, and shown that the main properties of EL-O hold in this logic  $\text{EL-O}+Ind$ : we have a finite model property, and the satisfiability problem is NP-complete. We have axiomatized  $\text{EL-O}+Ind$  and shown that it is indeed a fragment of the standard logic S5.

Our investigation in the first half of this chapter may appear somewhat old-fashioned: all our proofs are purely syntactical and we do not use any semantical tools, as was done in ‘the syntactic era (1918-1959)’ [24, Section 1.7] before Kripke semantics was invented. We nevertheless believe that axiomatic systems provide an important toolbox to understand intuitively what a logical system is able to express and what not. To witness, consider the inference rule  $RGFP$ : according to the explanations e.g. in [117], the rule says something about  $\varphi$  indicating to everybody that  $\psi$ ; however and as the equivalence with axiom  $GFP$  demonstrates, this is not the case: axiom  $GFP$  of the equivalent  $GFP$ -based axiomatics has a single schematic variable  $\varphi$ , which shows us that the concept of one proposition indicating another proposition is not accounted for by the Kripke semantics. This is in line with the analysis of [44] where it is argued that this concept cannot be modeled in Kripke semantics and where the authors investigate a different semantical framework.

## Chapter 5

# Planning with EL-O

### La planification EL-O

Dans ce chapitre nous décrivons les actions et tâches de planification EL-O et définissons leur sémantique ainsi que la notion de solvabilité pour les planifications séquentielle et parallèle. Ces définitions suivent de très près celles de la planification classique : les actions ont des préconditions et des effets conditionnels, mais ces conditions et effets sont exprimés dans le langage EL-O, et la sémantique doit tenir compte des interactions entre connaissance individuelle et connaissance commune. Par exemple, si un agent  $i$  oublie la valeur de vérité de  $p$ , non seulement  $S_i p$  mais aussi  $JS p$  doivent être retirés le cas échéant de l'état courant.

Etant donnée une tâche de planification EL-O, nous définissons ensuite une ‘expansion’ de cette tâche qui est solvable en planification classique si et seulement si la première tâche est solvable en planification EL-O. Ceci nous permet d’obtenir notre principal résultat de complexité : le problème de solvabilité en planification EL-O est PSPACE-complet, aussi bien dans le cas séquentiel que dans le cas parallèle. Ainsi le cadre EL-O nous permet de faire de la planification épistémique au prix de la planification classique, et la fonction d’expansion que nous définissons permet d’utiliser des planificateurs classiques à cet effet.

Nous illustrons enfin l’applicabilité de notre cadre à l’aide d’un éventail d’exemples simples, dont certains exemples courants de la littérature tels que le problème des deux généraux et le problème de bavardage.

---

In this chapter we describe EL-O actions and planning tasks and define semantics and solvability for both sequential and parallel planning. We define for any EL-O planning task an expansion of that planning task which is classically solvable if and only if the original planning task is EL-O solvable. This gives us our main complexity result: deciding solvability of an EL-O planning task is PSPACE-complete, both in the sequential and parallel cases. We finish by illustrating the applicability of our framework on a variety of examples.

This chapter is mainly based on the following publications:

- Cooper, M. C., Herzig, A., Maffre, F., Maris, F., Perrotin, E., & Régnier, P. (2020). *A lightweight epistemic logic and its application to planning*.

Artificial Intelligence, 103437.

- Cooper, M., Herzig, A., Maris, F., Perrotin, E., & Vianey, J. (2020). *Lightweight parallel multi-agent epistemic planning*. 17th International Conference on Principles of Knowledge Representation and Reasoning (KR 2020).

## Contents

<b>5.1</b>	<b>Epistemic planning with conditional effects . . . .</b>	<b>69</b>
5.1.1	EL-O consistent action descriptions . . . . .	69
5.1.2	Semantics and solvability by a sequential plan . . . .	71
5.1.3	Consistency of a set of actions at a state . . . . .	72
5.1.4	Semantics and solvability by a parallel plan . . . . .	73
<b>5.2</b>	<b>Planning over finite states and complexity . . . .</b>	<b>74</b>
5.2.1	Planning over finite states . . . . .	74
5.2.2	Translation into classical planning . . . . .	76
5.2.3	Complexity of EL-O planning . . . . .	76
<b>5.3</b>	<b>Applications . . . . .</b>	<b>77</b>
5.3.1	Two Generals' Problem . . . . .	77
5.3.2	Learning a message . . . . .	78
5.3.3	Selective communication . . . . .	79
5.3.4	Pedestrian crossing . . . . .	80
5.3.5	Authorisation via plenary meetings . . . . .	80
5.3.6	The generalised gossip problem . . . . .	81
5.3.7	Generalised gossiping with ignorance goals . . . . .	82
<b>5.4</b>	<b>Conclusion . . . . .</b>	<b>83</b>

## 5.1 Epistemic planning with conditional effects

Just as we instantiated the vocabulary  $\mathbb{V}$  of the presentation of boolean languages with  $ATM$  in order to define the language of EL-O, we define EL-O actions by instantiating  $\mathbb{V}$  in the general definition of action descriptions and planning tasks of Chapter 2 (Section 2.3.1) by  $ATM$  to obtain EL-O actions and planning tasks. That is, an EL-O action description is a pair  $\mathbf{a} = \langle pre(\mathbf{a}), eff(\mathbf{a}) \rangle$  where  $pre(\mathbf{a}) \in \mathcal{L}_{\text{bool}}(ATM)$  and  $eff(\mathbf{a}) \subseteq \mathcal{L}_{\text{bool}}(ATM) \times 2^{ATM} \times 2^{ATM}$ , and an EL-O epistemic planning task, or epistemic planning task over  $ATM$ , is a triple  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  where  $Act$  is a set of EL-O action descriptions,  $s_0 \in 2^{ATM}$  is the initial state, and  $Goal \in \mathcal{L}_{\text{bool}}(ATM)$  is the goal formula.

We now describe consistency conditions and semantics of EL-O actions, both in the contexts of sequential and of parallel planning.

### 5.1.1 EL-O consistent action descriptions

An action description  $\mathbf{a} = \langle pre(\mathbf{a}), eff(\mathbf{a}) \rangle$  over  $ATM$  is EL-O *consistent* if:

1. for every  $ce \in eff(\mathbf{a})$ ,  $ceff^-(ce)$  contains no introspective atoms;

2. for every  $ce_1, ce_2 \in \text{eff}(\mathbf{a})$ , if  $\text{pre}(\mathbf{a}) \wedge \text{cnd}(ce_1) \wedge \text{cnd}(ce_2)$  is satisfiable in EL-O then  $\text{ceff}^+(ce_1)$  and  $(\text{ceff}^-(ce_2))^\Leftarrow$  are disjoint.

The first condition is clear: it makes no sense to delete introspective atoms. The second condition means that two conditional effects of an action cannot conflict when  $\text{pre}(\mathbf{a})$  and their triggering conditions are jointly satisfiable. Hence when an action  $\mathbf{a}$  has conditional effects  $ce_1, ce_2 \in \text{eff}(\mathbf{a})$  and there are  $\alpha_1 \in \text{ceff}^+(ce_1)$  and  $\alpha_2 \in \text{ceff}^-(ce_2)$  such that  $\alpha_1 \Rightarrow \alpha_2$  then  $\text{pre}(\mathbf{a}) \wedge \text{cnd}(ce_1) \wedge \text{cnd}(ce_2)$  must be inconsistent. Note that it follows from the transitivity of atomic consequence  $\Rightarrow$  that  $\text{ceff}^+(ce_1)$  and  $(\text{ceff}^-(ce_2))^\Leftarrow$  are disjoint iff  $\text{ceff}^+(ce_1)^\Rightarrow$  and  $(\text{ceff}^-(ce_2))^\Leftarrow$  are.

**Example 5.1** (Lights on, ctd.). Recall the action of flipping a light switch inside a room given in Chapter 2:

$$\text{flip}_i = \langle \text{In}_i, \{\langle \text{On}, \emptyset, \{\text{On} \}\rangle, \langle \neg \text{On}, \{\text{On}\}, \emptyset \rangle\} \rangle$$

As the  $\mathcal{L}_{\text{bool}}(\text{Prop})$  is included in  $\mathcal{L}_{\text{bool}}(\text{ATM})$ , this is an EL-O action description as well as a classical action description.

We suppose that there are only two agents:  $\text{Agt} = \{1, 2\}$ . Consider now the actions of entering and exiting the room. When entering a room, an agent gains individual knowledge of whether or not the light is on, and if the other agent is also in the room, they gain joint knowledge of the state of the light. When exiting the room, an agent no longer knows whether or not the light is on, as the switch could be flipped while they are outside the room. Hence the two actions can be described in the following manner:

$$\begin{aligned} \text{enter}_i &= \langle \neg \text{In}_i, \{\langle \top, \{\text{In}_i, S_i \text{ On}\}, \emptyset \rangle, \langle \text{In}_j, \{JS \text{ On}\}, \emptyset \rangle\} \rangle \\ \text{exit}_i &= \langle \text{In}_i, \{\langle \top, \emptyset, \{\text{In}_i, S_i \text{ On}\}\rangle\} \rangle \end{aligned}$$

for  $i, j \in \{1, 2\}$  and  $i \neq j$ .

**Example 5.2** (Gossip, ctd.). Let us describe calls in the original gossip problem where the goal is to obtain shared knowledge of depth  $k = 1$ . Suppose  $\text{ATM}$  is built from  $\text{Agt} = \{1, \dots, n\}$  and  $\text{Prop} = \{s_i : i \in \text{Agt}\}$ . During the action  $\text{call}_j^i$ , in which a call is made between agents  $i$  and  $j$  (this action being symmetrical), the two agents tell each other every secret they know among all  $n$  secrets. We have  $\text{call}_j^i = \langle \text{pre}(\text{call}_j^i), \text{eff}(\text{call}_j^i) \rangle$  with  $\text{pre}(\text{call}_j^i) = \top$  and:

$$\begin{aligned} \text{eff}(\text{call}_j^i) &= \{\langle S_i s_1 \vee S_j s_1, \{S_i s_1, S_j s_1\}, \emptyset \rangle, \\ &\quad \dots, \\ &\quad \langle S_i s_n \vee S_j s_n, \{S_i s_n, S_j s_n\}, \emptyset \rangle\}. \end{aligned}$$

Intuitively, a secret becomes observable for both agents if at least one of them observes it. Each  $\text{call}_j^i$  is EL-O consistent because it has no negative effects.

Here is a more complex example of a consistent action.

**Example 5.3.** Consider the action `pour` of pouring a bottle of liquid on a healthy lawn [110]. Factually, its effect is that if the liquid is poisonous the

lawn becomes dead. Epistemically, `pour` has several conditional effects: (1) if  $i$  knows that `Poisonous` then  $i$  will know that the lawn is dead; (2) if  $i$  does not know whether `Poisonous` and if the lawn is known to be healthy<sup>1</sup> then  $i$  will no longer know whether the lawn is healthy or dead. (There is a third possible epistemic situation where  $i$  knows that  $\neg$ `Poisonous`, but we need not say anything here because there is no effect as far as the status of `Dead` is concerned.) Consider the description of `pour` with precondition  $pre(\text{pour}) = \top$  and effect

$$\begin{aligned} eff(\text{pour}) = & \{ \langle \text{Poisonous}, \{ \text{Dead} \}, \emptyset \rangle, \\ & \langle \text{Poisonous} \wedge S_i \text{ Poisonous}, \{ S_i \text{ Dead} \}, \emptyset \rangle, \\ & \langle \neg S_i \text{ Poisonous} \wedge \neg \text{Dead} \wedge S_i \text{ Dead}, \emptyset, \{ S_i \text{ Dead} \} \rangle \}. \end{aligned}$$

The first conditional effect is factual and the last two are epistemic. The second represents (1):  $\text{Poisonous} \wedge S_i \text{ Poisonous}$  expresses that  $i$  knows that the liquid is poisonous; and the third represents (2):  $\neg \text{Dead} \wedge S_i \text{ Dead}$  expresses that  $i$  knows that the lawn is healthy. This action description is EL-O consistent in particular because the conjunction of the two last conditions  $\text{Poisonous} \wedge S_i \text{ Poisonous}$  and  $\neg S_i \text{ Poisonous} \wedge \neg \text{Dead} \wedge S_i \text{ Dead}$  is inconsistent.

### 5.1.2 Semantics and solvability by a sequential plan

We define the EL-O semantics of a consistent action  $\mathbf{a}$  in terms of a partial function  $\tau_{\mathbf{a}}^{\text{EL-O}}$  on introspective states, that is, states of  $I\text{-STATES} = \{s \Rightarrow \cup I\text{-ATM} : s \subseteq \text{ATM}\}$ . The function  $\tau_{\mathbf{a}}^{\text{EL-O}}$  is defined at  $s$  if  $s \models pre(\mathbf{a})$ . In that case we say that  $\mathbf{a}$  is executable at  $s$  and stipulate:

$$\tau_{\mathbf{a}}^{\text{EL-O}}(s) = \left( s \setminus \bigcup_{\substack{ce \in eff^-(\mathbf{a}), \\ s \models cnd(ce)}} (ceff^-(ce)) \Leftarrow \right) \cup \bigcup_{\substack{ce \in eff^+(\mathbf{a}), \\ s \models cnd(ce)}} (ceff^+(ce)) \Rightarrow.$$

That is, if the precondition of  $\mathbf{a}$  is satisfied then  $\mathbf{a}$  removes negative effects of all those conditional effects  $ce$  that ‘fire’, i.e., whose triggering conditions are satisfied, plus their causes; and it adds the positive effects of  $ce$ , plus their consequences. Consistency of  $\mathbf{a}$  ensures that it does not matter in which order we apply negative and positive effects.

Following the formalism introduced in Chapter 2, an EL-O planning task  $\langle Act, s_0, Goal \rangle$  is a triple such that  $Act$  is a set of consistent EL-O action descriptions,  $s_0 \subseteq \text{ATM}$  is the finite initial state and  $Goal \in \mathcal{L}_{\text{bool}}(\text{ATM})$  is the goal formula. Such a task is EL-O *solvable by a sequential plan* iff there is a state  $s$  such that  $s \models Goal$  that is reachable from  $s_0 \Rightarrow \cup I\text{-ATM}$  via the EL-O interpretation of the actions in  $Act$ , i.e., via  $\{\tau_{\mathbf{a}}^{\text{EL-O}} : \mathbf{a} \in Act\}$ .

**Example 5.4** (Lights on, ctd.). Consider the state  $s = \{\text{On}, \text{In}_1, S_1 \text{ On}\}$ . If agent 2 enters the room, the atoms added to  $s$  are  $\text{In}_2$ , and  $JS \text{ On}$  as well as all of its consequences. Hence the current state becomes  $\tau_{\text{enter}_2}^{\text{EL-O}}(s) = \{\text{On}, \text{In}_1, \text{In}_2\} \cup \{\sigma \text{ On} : \sigma \in \text{OBS}^+\}$ . If agent 1 then leaves the room, the atoms  $\text{In}_1$  and  $S_1 \text{ On}$

<sup>1</sup>In [110], the condition of the last effect is only that it is not known *that*  $\neg$ `Poisonous`, and the effect is that it is no longer known *that*  $\neg$ `Dead`. This however produces an unintuitive result in states where the liquid is known to be poisonous.

are removed, as well as the only cause of the latter (other than itself)  $JS\text{On}$ . Hence we arrive at  $\tau_{\text{exit}_1}^{\text{EL-O}}(\tau_{\text{enter}_2}^{\text{EL-O}}(s)) = \{\text{On}, \text{In}_2\} \cup \{\sigma\text{On} : \sigma \in \text{OBS}^+, \ell(\sigma) \geq 2\}$ .

It might seem surprising that atoms such as  $JS\text{S}_2p$  remain true after agent 1 exits the room. However, there are two possibilities: either higher order knowledge is not considered in the planning task at hand, in which case this ‘residual knowledge’ has no impact on planning, or higher order knowledge is considered, in which case the definition of the action  $\text{exit}_i$  should be defined accordingly so that all the necessary knowledge is removed. If the first, the alternative semantics given in Section 5.2.1 will allow us to disregard superfluous atoms.

**Example 5.5** (Gossip, ctd.). The planning task corresponding to the original gossip problem is  $G_1 = \langle \text{Act}^{G_1}, s_0^{G_1}, \text{Goal}^{G_1} \rangle$  with

$$\begin{aligned} \text{Act}^{G_1} &= \{\text{call}_j^i : i, j \in \text{Agt} \text{ and } i \neq j\}, \\ s_0^{G_1} &= \{S_i s_i : i \in \text{Agt}\} \cup S \text{ for some } S \subseteq \{s_i : i \in \text{Agt}\}, \\ \text{Goal}^{G_1} &= \bigwedge_{i, j \in \text{Agt}} S_i s_j. \end{aligned}$$

It can be solved via  $2n - 4$  calls; when  $n \geq 4$  then there is no EL-O solution that takes less calls [11, 121, 64]. For instance, for  $n = 4$  the sequence  $\text{call}_2^1; \text{call}_4^3; \text{call}_3^1; \text{call}_4^2$  is an EL-O solution with 4 calls.

Classical planning tasks as defined in Chapter 2 are particular EL-O epistemic planning tasks: when action descriptions contain no observability operators then classical consistency and EL-O consistency of actions coincide; moreover,  $\text{ceff}^-(ce) = (\text{ceff}^-(ce))^\leftarrow$  and  $\text{ceff}^+(ce) = (\text{ceff}^+(ce))^\rightarrow$  for all conditional effects, and therefore  $\tau_{\mathbf{a}}^{\text{EL-O}} = \tau_{\mathbf{a}}^{\text{CPC}}$  for every action  $\mathbf{a}$ . It follows from PSPACE hardness of classical solvability that solvability of EL-O planning tasks is PSPACE-hard, too.

### 5.1.3 Consistency of a set of actions at a state

We also need to take introspection into consideration when defining consistency of sets of EL-O actions. There are two elements to this definition. First, two actions  $\mathbf{a}_1$  and  $\mathbf{a}_2$  that are both executable at  $s$  *have no contradictory effects at  $s$*  if for every  $ce_1 \in \text{eff}(\mathbf{a}_1)$  and  $ce_2 \in \text{eff}(\mathbf{a}_2)$ , if  $s \models \text{cnd}(ce_1) \wedge \text{cnd}(ce_2)$  then  $\text{ceff}^+(ce_1) \cap \text{ceff}^-(ce_2)^\leftarrow = \emptyset$ . Just as with classical actions, this generalizes the definition of consistency for single actions.

Second, we say that two different actions  $\mathbf{a}_1$  and  $\mathbf{a}_2$  that are executable at  $s$  *have no cross-interaction at  $s$*  if the following hold:

1.  $s$  and  $\tau_{\mathbf{a}_1}^{\text{EL-O}}(s)$  agree on  $\text{pre}(\mathbf{a}_2)$  and on the condition  $\text{cnd}(ce_2)$  of every conditional effect  $ce_2 \in \text{eff}(\mathbf{a}_2)$ ;
2.  $s$  and  $\tau_{\mathbf{a}_2}^{\text{EL-O}}(s)$  agree on  $\text{pre}(\mathbf{a}_1)$  and on the condition  $\text{cnd}(ce_1)$  of every conditional effect  $ce_1 \in \text{eff}(\mathbf{a}_1)$ .

Putting things together, we say that a set of actions  $\mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_m\}$  is *consistent in state  $s$*  if for every  $\mathbf{a}_i, \mathbf{a}_j \in \mathbf{A}$  such that  $\mathbf{a}_i \neq \mathbf{a}_j$ ,

1.  $\mathbf{a}_i$  and  $\mathbf{a}_j$  have no contradictory effects in  $s$ ;

2.  $a_i$  and  $a_j$  have no cross-interaction in  $s$ .

**Example 5.6** (Lights on, ctd.). Let  $s = \{\text{On}, \text{In}_1, S_1 \text{ On}\}$ . The actions  $\text{exit}_1$  and  $\text{enter}_2$  are both executable at  $s$ , but they have contradictory effects at  $s$ :  $s$  and  $\tau_{\text{exit}_1}^{\text{EL-O}}(s)$  disagree on  $\text{In}_1$ , which is the condition of one of the conditional effects of  $\text{enter}_2$ . Hence the set  $\{\text{exit}_1, \text{enter}_2\}$  is inconsistent in  $s$ .

**Example 5.7** (Gossip, ctd.). Any set of gossiping calls  $\text{call}_j^i$  is consistent in any state. Therefore conference calls  $\{\text{call}_j^i, \text{call}_r^i\}$  where  $i$  calls  $j$  and  $r$  at the same time are consistent, making the parallel gossiping task solvable in one step.

One way to exclude conference calls is to replace  $\text{call}_j^i$  by  $\text{startCall}_j^i$  plus a single  $\text{endCalls}$  action as follows:

$$\begin{aligned} \text{pre}(\text{startCall}_j^i) &= \text{free}_i \wedge \text{free}_j, \\ \text{eff}(\text{startCall}_j^i) &= \text{eff}(\text{call}_j^i) \cup \{\langle \top, \emptyset, \{\text{free}_i, \text{free}_j\} \rangle\}, \\ \text{pre}(\text{endCalls}) &= \top, \\ \text{eff}(\text{endCalls}) &= \{\langle \top, \{\text{free}_i : i \in \text{Agt}\}, \emptyset \rangle\}, \end{aligned}$$

and to add all  $\text{free}_i$  to the initial state. Then there is no state where a set of actions with conference calls is consistent:  $\text{startCall}_j^i$  and  $\text{startCall}_r^i$  have cross-interaction at any state satisfying  $\text{free}_i \wedge \text{free}_j \wedge \text{free}_r$ .

While this solution is natural (agents cannot call two agents at a time because they are no longer available once they have begun a call), splitting calls into two separate actions artificially doubles the number of steps in an optimal solution. Another possibility that avoids the  $\text{endCalls}$  action is to replace all  $\text{call}_j^i$  by  $\text{Tcall}_j^i$ , with:

$$\begin{aligned} \text{pre}(\text{Tcall}_j^i) &= \top, \\ \text{eff}(\text{Tcall}_j^i) &= \text{eff}(\text{call}_j^i) \cup \\ &\quad \{\langle \text{tg}_i, \emptyset, \{\text{tg}_i\} \rangle\} \cup \{\langle \neg \text{tg}_i, \{\text{tg}_i\}, \emptyset \rangle\} \cup \\ &\quad \{\langle \text{tg}_j, \emptyset, \{\text{tg}_j\} \rangle\} \cup \{\langle \neg \text{tg}_j, \{\text{tg}_j\}, \emptyset \rangle\}. \end{aligned}$$

Here any two calls involving  $i$  each toggles the value of  $\text{tg}_i$ , which means that these calls have cross-interaction at any state satisfying their preconditions.

#### 5.1.4 Semantics and solvability by a parallel plan

A set of actions  $A = \{a_1, \dots, a_m\}$  determines a partial function  $\tau_A^{\text{EL-O}}$  on introspective states. The function  $\tau_A^{\text{EL-O}}$  is defined at  $s$  if every  $a_i \in A$  is executable at  $s$  and  $A$  is consistent in  $s$ . When  $\tau_A^{\text{EL-O}}$  is defined at  $s$  then:

$$\tau_A^{\text{EL-O}}(s) = \left( s \setminus \bigcup_{\substack{a \in A, ce \in \text{eff}(a), \\ s \models \text{cnd}(ce)}} (\text{ceff}^-(ce)) \leftarrow \right) \cup \bigcup_{\substack{a \in A, ce \in \text{eff}(a), \\ s \models \text{cnd}(ce)}} (\text{ceff}^+(ce)) \Rightarrow.$$

The consistency requirements ensure that actions in a consistent set of actions at  $s$  can be interleaved arbitrarily: we have  $\tau_{\{a_1, a_2\}}^{\text{EL-O}}(s) = \tau_{a_2}^{\text{EL-O}}(\tau_{a_1}^{\text{EL-O}}(s)) = \tau_{a_1}^{\text{EL-O}}(\tau_{a_2}^{\text{EL-O}}(s))$ . More generally:



**Proposition 5.1.** *If  $a \in A$  is consistent in  $s$  with any other action in  $A$  then  $\tau_A^{\text{EL-O}}(s) = \tau_a^{\text{EL-O}}(\tau_{A \setminus \{a\}}^{\text{EL-O}}(s)) = \tau_{A \setminus \{a\}}^{\text{EL-O}}(\tau_a^{\text{EL-O}}(s))$ .*

An EL-O planning task  $\langle Act, s_0, Goal \rangle$  is EL-O solvable by a parallel plan iff there is a state  $s$  such that  $s \models Goal$  that is reachable from  $s_0^{\vec{\sigma}} \cup I-ATM$  via the EL-O-interpretation of sets of actions of  $Act$ , i.e., via  $\{\tau_A^{\text{EL-O}} : A \subseteq Act\}$ . Solvability by a sequential plan is the special case where the parallel plan is a sequence of singletons.

**Example 5.8** (Lights on, ctd.). Let  $s = \{On, ln_1, S_1 On\}$ . The set  $\{\text{enter}_2, \text{flip}_1\}$  is consistent at  $s$ , and  $\tau_{\{\text{enter}_2, \text{flip}_1\}}^{\text{EL-O}}(s) = \{ln_1, ln_2\} \cup \{\sigma On : \sigma \in OBS^+\}$ .

**Example 5.9** (Gossip, ctd.). Let  $G'_1$  be modification of  $G_1$  that is obtained by replacing the actions  $\text{call}_j^i$  by  $\text{Tcall}_j^i$  of Example 5.7. Then  $G'_1$  can be solved in  $\lceil \log_2 n \rceil$  steps of parallel calls if the number of agents  $n$  is even, and in  $\lceil \log_2 n \rceil + 1$  steps if  $n$  is odd [19, 86, 79, 41]. For instance, for  $n = 4$  the parallel plan  $\langle \{\text{Tcall}_2^1, \text{Tcall}_4^3\}, \{\text{Tcall}_3^1, \text{Tcall}_4^2\} \rangle$  is a solution of  $G'_1$  with 2 steps.

## 5.2 Planning over finite states and complexity

### 5.2.1 Planning over finite states

While we have defined the semantics of EL-O actions on introspective states for full generality, it is not very feasible in practice to reason with infinite states. In this section we give a different semantics which allows us to work with finite states. This is done by defining semantics relative to a given vocabulary. Intuitively, if the planner does not wish to consider modal depth higher than, say, 3, then the effects of actions of depth 4 or more can be disregarded entirely with no ill effect. In particular, when adding joint vision to a state, there is no need to also add all of its consequences; it is enough to only consider those that are relevant to the planning task at hand.

Let  $Voc \subseteq ATM$  be a given vocabulary, and  $a$  be a consistent action description such that  $\text{Voc}(a) \subseteq Voc$ . The function  $\tau_a^{\text{Voc}}$  is a partial functions on states of the following set:

$$\begin{aligned} I\text{-STATES}|_{Voc} &= \{s \cap Voc : s \in I\text{-STATES}\} \\ &= \{(s^{\vec{\sigma}} \cup I\text{-ATM}) \cap Voc : s \in 2^{ATM}\} \end{aligned}$$

For any state  $s \in I\text{-STATES}|_{Voc}$ ,  $\tau_a^{\text{Voc}}$  is defined if  $s \models \text{pre}(a)$ . In that case

$$\begin{aligned} \tau_a^{\text{Voc}}(s) &= \left( s \setminus \left( \bigcup_{\substack{ce \in \text{eff}^-(a), \\ s \models \text{cnd}(ce)}} \text{ceff}^-(ce) \cap Voc \right) \right) \cup \\ &\quad \bigcup_{\substack{ce \in \text{eff}^+(a), \\ s \models \text{cnd}(ce)}} \left( (\text{ceff}^+(ce))^{\vec{\sigma}} \cap Voc \right). \end{aligned}$$

It is easily shown that if  $s \in I\text{-STATES}|_{Voc}$  then  $\tau_a^{\text{Voc}}(s) \in I\text{-STATES}|_{Voc}$ .

**Example 5.10** (Lights on, ctd.). Consider the state  $s = \{\text{On}, \text{In}_1, S_1 \text{ On}\}$  of Example 5.4, and suppose that we do not care about higher order observation, but only whether observation of  $\text{On}$  is individual or joint: the set of atoms that we consider is  $Voc = \{\text{In}_1, \text{In}_2, \text{On}, S_1 \text{ On}, S_2 \text{ On}, JS \text{ On}\}$ . Suppose now that agent 2 enters the room: the atoms added to  $s$  are  $\text{In}_2$  and  $JS \text{ On}$  as well as all of its consequences that are in  $Voc$ , that is, only  $S_2 \text{ On}$ . Hence the current state becomes  $\tau_{\text{enter}_2}^{Voc}(s) = \{\text{On}, \text{In}_1, S_1 \text{ On}, \text{In}_2, S_2 \text{ On}, JS \text{ On}\}$ . If agent 1 then leaves the room, the atoms  $\text{In}_1$  and  $S_1 \text{ On}$  are removed, as well as the cause of the latter  $JS \text{ On}$ . Hence we arrive at  $\tau_{\text{exit}_1}^{Voc}(\tau_{\text{enter}_2}^{Voc}(s)) = \{\text{On}, \text{In}_2, S_2 \text{ On}\}$ .

We now show that planning with these semantics is equivalent to planning with the more general semantics given previously.

**Lemma 5.1.** *Let  $Voc$  be a subset of  $ATM$ ,  $s \in I\text{-STATES}$  be an introspective state and  $\mathbf{a}$  be a consistent action such that  $\text{Voc}(\mathbf{a}) \subseteq Voc$ . Then  $\tau_{\mathbf{a}}^{\text{EL-O}}(s)$  is defined iff  $\tau_{\mathbf{a}}^{Voc}(s \cap Voc)$  is defined, and in that case  $\tau_{\mathbf{a}}^{Voc}(s \cap Voc) = \tau_{\mathbf{a}}^{\text{EL-O}}(s) \cap Voc$ .*

*Proof.* As  $\text{Voc}(\text{pre}(\mathbf{a})) \subseteq Voc$ ,  $s$  and  $s \cap Voc$  agree on  $\text{pre}(\mathbf{a})$  (following Proposition 3.5). Hence  $\tau_{\mathbf{a}}^{\text{EL-O}}(s)$  is defined iff  $\tau_{\mathbf{a}}^{Voc}(s \cap Voc)$  is defined.

By the same argument,  $s$  and  $s \cap Voc$  agree on  $\text{cnd}(ce)$  for all conditional effects  $ce$  of  $\mathbf{a}$ . Hence:

$$\begin{aligned} \tau_{\mathbf{a}}^{\text{EL-O}}(s) &= \left( s \setminus \bigcup_{\substack{ce \in \text{eff}(\mathbf{a}), \\ s \models \text{cnd}(ce)}} (\text{ceff}^-(ce)) \leftarrow \right) \cup \bigcup_{\substack{ce \in \text{eff}(\mathbf{a}), \\ s \models \text{cnd}(ce)}} (\text{ceff}^+(ce)) \Rightarrow \\ &= \left( s \setminus \bigcup_{\substack{ce \in \text{eff}(\mathbf{a}), \\ s \cap Voc \models \text{cnd}(ce)}} (\text{ceff}^-(ce)) \leftarrow \right) \cup \bigcup_{\substack{ce \in \text{eff}(\mathbf{a}), \\ s \cap Voc \models \text{cnd}(ce)}} (\text{ceff}^+(ce)) \Rightarrow \end{aligned}$$

By intersecting this with  $Voc$  we conclude that  $\tau_{\mathbf{a}}^{\text{EL-O}}(s) \cap Voc = \tau_{\mathbf{a}}^{Voc}(s \cap Voc)$ .  $\square$

This result can be generalized to any sequence of actions: if  $Act$  is such that for all  $\mathbf{a} \in Act$ ,  $\text{Voc}(\mathbf{a}) \subseteq Voc$ , then for every  $s \in I\text{-STATES}|_{Voc}$ , there exists a state  $s'$  that is reachable from  $s$  via  $\{\tau_{\mathbf{a}}^{\text{EL-O}} : \mathbf{a} \in Act\}$  iff there exists a state  $s''$  that is reachable from  $s \cap Voc$  via  $\{\tau_{\mathbf{a}}^{Voc} : \mathbf{a} \in Act\}$ , and in that case  $s'' = s' \cap Voc$ .

**Proposition 5.2.** *Let  $Voc$  be a subset of  $ATM$  and  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  be a planning task such that  $\text{Voc}(\mathcal{P}) \subseteq Voc$ . Then  $\mathcal{P}$  is EL-O solvable by a sequential plan iff the planning task  $\langle Act, (s_0^{\rightarrow} \cup I\text{-ATM}) \cap Voc, Goal \rangle$  is solvable by a sequential plan via  $\{\tau_{\mathbf{a}}^{Voc} : \mathbf{a} \in Act\}$ .*

*Proof.*  $\mathcal{P}$  is EL-O solvable via  $\{\tau_{\mathbf{a}}^{\text{EL-O}} : \mathbf{a} \in Act\}$  if there exists a state  $s$  such that  $s \models Goal$  that is reachable from  $s_0^{\rightarrow} \cup I\text{-ATM}$  via  $\{\tau_{\mathbf{a}}^{\text{EL-O}} : \mathbf{a} \in Act\}$ . As we have just seen, this is the case iff there exists a state  $s'$  reachable from  $(s_0^{\rightarrow} \cup I\text{-ATM}) \cap Voc$  via  $\{\tau_{\mathbf{a}}^{Voc} : \mathbf{a} \in Act\}$ , and in that case  $s' = s \cap Voc$ . In particular, as  $\text{Voc}(\mathcal{P}) \subseteq Voc$ ,  $s'$  and  $s''$  agree on  $\text{Voc}(Goal)$ , and therefore on  $Goal$ .  $\square$

These results can be extended to parallel planning. Given a set of atoms  $Voc$  and a set of actions  $A$ , the function  $\tau_A^{Voc}$  is once again a partial function on states of  $I-STATES|_{Voc}$ . For any state  $s \in I-STATES|_{Voc}$ ,  $\tau_A^{Voc}$  is defined if all  $a \in A$  are executable at  $s$  and  $A$  is consistent at  $s$ . In that case

$$\tau_a^{Voc}(s) = \left( s \setminus \left( \bigcup_{\substack{a \in A, ce \in \text{eff}(a), \\ s \models \text{cnd}(ce)}} (\text{ceff}^-(ce) \setminus I-ATM)^{\Leftarrow} \cap Voc \right) \right) \cup \bigcup_{\substack{a \in A, ce \in \text{eff}(a), \\ s \models \text{cnd}(ce)}} \left( (\text{ceff}^+(ce))^{\Rightarrow} \cap Voc \right).$$

In the same manner as before we can prove:

**Proposition 5.3.** *Let  $Voc$  be a subset of  $ATM$  and  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  be a planning task such that  $\text{Voc}(\mathcal{P}) \subseteq Voc$ . Then  $\mathcal{P}$  is EL-O solvable by a parallel plan iff the planning task  $\langle Act, (s_0^{\Rightarrow} \cup I-ATM) \cap Voc, Goal \rangle$  is solvable by a parallel plan via  $\{\tau_a^{Voc} : a \in Act\}$ .*

## 5.2.2 Translation into classical planning

We now define the *expansion* of any EL-O planning task  $\mathcal{P}$ , which closely follows the semantics of Section 5.2.1, naturally choosing  $\text{Voc}(\mathcal{P})$  as the relevant vocabulary. If  $\mathcal{P} = \langle Act, s_0, Goal \rangle$ , its expansion is defined as:

$$\text{Exp}(\mathcal{P}) = \langle \{ \langle \text{pre}(a), \text{Exp}_{\mathcal{P}}(\text{eff}(a)) \rangle : a \in Act \}, (s_0^{\Rightarrow} \cup I-ATM) \cap \text{Voc}(\mathcal{P}), Goal \rangle$$

where the expansion of the effects of an action relative to  $\mathcal{P}$  is defined as:

$$\text{Exp}_{\mathcal{P}}(\text{eff}(a)) = \{ \langle \text{cnd}(ce), \text{ceff}^+(ce)^{\Rightarrow} \cap \text{Voc}(\mathcal{P}), \text{ceff}^-(ce)^{\Leftarrow} \cap \text{Voc}(\mathcal{P}) \rangle : ce \in \text{eff}(a) \}.$$

After expansion classical solvability and EL-O solvability coincide:

**Proposition 5.4.** *Let  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  be an EL-O epistemic planning task. Then  $\mathcal{P}$  is EL-O solvable by a sequential plan iff  $\text{Exp}(\mathcal{P})$  is classically solvable by a sequential plan. Moreover,  $\mathcal{P}$  is EL-O solvable by a parallel plan iff  $\text{Exp}(\mathcal{P})$  is classically solvable by a parallel plan.*

*Proof.* This follows directly from Propositions 5.2 and 5.3. We only give details for sequential planning:  $\mathcal{P}$  is EL-O solvable by a sequential plan iff the planning task  $\langle Act, (s_0^{\Rightarrow} \cup I-ATM) \cap \text{Voc}(\mathcal{P}), Goal \rangle$  is solvable by a sequential plan via  $\{\tau_a^{\text{Voc}(\mathcal{P})} : a \in Act\}$ . Clearly this is the case iff  $\text{Exp}(\mathcal{P})$  is classically solvable by a sequential plan.  $\square$

## 5.2.3 Complexity of EL-O planning

Because everything in the expansion of planning tasks is restricted to the vocabulary of that planning task, the length of the expansion of epistemic planning tasks  $\mathcal{P}$  is polynomial in the length of  $\mathcal{P}$ :  $\ell(\text{Exp}(\mathcal{P})) \leq (\ell(\mathcal{P}))^2$ .

**Proposition 5.5.** *Solvability of an EL-O planing task, both by a sequential and a parallel plan, is PSPACE-complete.*

*Proof.* Recall that classical sequential and parallel planning are both PSPACE-complete. As the language of CPC is a subset of the language of EL-O, and the classical and EL-O interpretations of actions and sets of actions coincide on  $\mathcal{L}_{\text{bool}}(\text{Prop})$ , classical planning is a particular case of epistemic planning. The solvability problem is therefore PSPACE-hard just like classical planning [33]. Membership follows from Proposition 5.4 and PSPACE-completeness of classical planning.  $\square$

## 5.3 Applications

In this section, we give some examples of epistemic planning tasks. We start with the Byzantine Generals Problem as well as some toy examples, most of which were introduced in [42]. We then focus on the generalised gossip problem.

### 5.3.1 Two Generals' Problem

The Two Generals' Problem [1, 63] is about coordination by communicating over an unreliable channel, highlighting the importance of common knowledge. The Byzantine Generals' Problem [85] is a generalisation of that problem. In short, two generals need to coordinate an attack; they can win only if they attack at the same time: either both attack in the morning ( $m$ ) or both attack in the afternoon ( $\neg m$ ). To communicate they must send a messenger who can be captured. They must decide on a time to attack, agree on this time, and each general must know that the other has agreed, and so on. To do this, one general sends a time for the attack, but since he cannot be sure the message was received, the other general must send an acknowledgment. Of course, just like the original message, the acknowledgment can be lost, so the first general must send an acknowledgment, and so on. It is impossible to be fully coordinated, i.e., to obtain common knowledge of the fact that the second general knows the time of the attack.

Formally, with  $\text{Agt} = \{1, 2\}$  and  $\text{Prop} = \{m\}$ , we define the atom  $\alpha_n$  as:

- $\alpha_n = \underbrace{S_1 S_2 S_1 \dots S_2}_n m$  if  $n$  is even;
- $\alpha_n = \underbrace{S_2 S_1 S_2 \dots S_2}_n m$  if  $n$  is odd.

For  $n \geq 0$  we define the family of actions  $\text{sndMsg}_n$  by stipulating:

$$\begin{aligned} \text{pre}(\text{sndMsg}_n) &= m \wedge S_1 m \wedge S_2 m \wedge \alpha_2 \wedge \alpha_3 \wedge \dots \wedge \alpha_{n-1} \\ \text{eff}(\text{sndMsg}_n) &= \{\langle \top, \{m, S_1 m, S_2 m, \alpha_2, \alpha_3, \dots, \alpha_n\}, \emptyset \rangle\} \end{aligned}$$

whose effect is that the time of the attack is set, both agents know it, agent 1 knows that agent 2 knows it, and so on until  $n$ .

The planning task to be solved is  $\mathcal{P} = \langle \text{Act}, s_0, \text{Goal} \rangle$  with  $s_0 = \emptyset$ ,  $\text{Goal} = JS S_2 m$ , and  $\text{Act} = \{\text{sndMsg}_n : n \geq 0\}$ . There is no solution for this task: while all sequences  $\langle \text{sndMsg}_0, \dots, \text{sndMsg}_n \rangle$  are executable, none of them reaches a state where  $JS S_2 m$  is true.

action	precondition	conditional effects
$\text{enter}_i$	$\neg \text{In}_i$	$\{\langle \top, \{\text{In}_i\}, \emptyset \rangle\}$
$\text{enter}'_i$	$\neg \text{In}_i \wedge \neg \text{In}_j, i \neq j$	$\{\langle \top, \{\text{In}_i\}, \emptyset \rangle\}$
$\text{leave}_i$	$\text{In}_i$	$\{\langle \top, \emptyset, \{\text{In}_i\} \rangle\}$
$\text{reveal}_i$	$\text{In}_i$	$\{\langle \top, \{S_i m\}, \emptyset \rangle\} \cup$ $\{\langle \text{In}_j, \{JS m\}, \emptyset \rangle, j \neq i\}$
$\text{informs}_{1,2} m$	$(\text{In}_1 \leftrightarrow \text{In}_2) \wedge S_1 m$ $\wedge \neg (S_2 m \wedge S_1 S_2 m)$	$\{\langle \top, \{JS m\}, \emptyset \rangle\}$
$\text{informs}'_{1,2} m$	$(\text{In}_1 \leftrightarrow \text{In}_2) \wedge S_1 m$ $\wedge \neg S_2 m \wedge S_1 S_2 m$	$\{\langle \top, \{JS m\}, \emptyset \rangle\}$

Table 5.1: Action descriptions for the message task

*Remark 5.1.* The initial state  $s_0 = \emptyset$  of the Two Generals' Problem illustrates a case of maximal ignorance as discussed in Remark 3.4. The infiniteness of the maximal ignorance Kripke model may explain why the DEL literature pays only little attention to this and other Byzantine coordination problems: most presentations are in semantical terms and use model checking in finite Kripke models. (A finite modeling was however proposed recently in [27].)

### 5.3.2 Learning a message

Suppose two agents 1 and 2 are outside a room ( $\neg \text{In}_1 \wedge \neg \text{In}_2$ ). The room contains a message that we suppose for simplicity to be the value of the propositional variable  $m$ . Each of the agents can enter and leave the room and can (temporarily) reveal the message when she is in the room. In case the other agent is also in the room the message is jointly seen. (So this is a conditional effect.) We define a series of planning tasks where the initial state is  $s_0 = \{m\}$  and where we vary the action descriptions and the goal. For convenience we list all the action descriptions in Table 5.1.

For a start, consider  $Act = \{\text{enter}_i, \text{leave}_i, \text{reveal}_i : i \in \{1, 2\}\}$  where  $\text{enter}_i$  and  $\text{leave}_i$  respectively have preconditions  $\neg \text{In}_i$  and  $\text{In}_i$  and the obvious effects and where  $\text{reveal}_i$  requires to be in the room and has one unconditional effect ( $i$  knows the message) and one conditional effect: if the other agent is in the room then the message becomes common knowledge. We suppose, for simplicity, that there are no epistemic consequences when an agent enters or leaves the room because we are not interested in the epistemic status of  $\text{In}_1$  and  $\text{In}_2$ . Consider  $Goal = \neg \text{In}_1 \wedge \neg \text{In}_2 \wedge S_1 m \wedge S_2 m$ . The sequences

$$\begin{aligned} sol_1 &= \langle \text{enter}_1, \text{reveal}_1, \text{leave}_1, \text{enter}_2, \text{reveal}_2, \text{leave}_2 \rangle, \\ sol_2 &= \langle \text{enter}_1, \text{enter}_2, \text{reveal}_1, \text{leave}_1, \text{leave}_2 \rangle \end{aligned}$$

are both solutions of  $\langle Act, s_0, Goal \rangle$ .

If we replace  $Goal$  by  $Goal' = \neg \text{In}_1 \wedge \neg \text{In}_2 \wedge JS m$  then  $sol_2$  still solves  $\langle Act, s_0, Goal' \rangle$ , but  $sol_1$  no longer does.

Next, consider a variant of  $Act$  in which the room is so small that it does not fit two people: we have to replace  $\text{enter}_i$  by  $\text{enter}'_i$  with precondition  $pre(\text{enter}'_i) = \neg \text{In}_i \wedge \neg \text{In}_j$ , for  $i \neq j$ . Let the resulting action description be  $Act'$ . While there is still a solution of  $\langle Act', s_0, Goal \rangle$ , there is no solution of  $\langle Act', s_0, Goal' \rangle$ .

action	precondition	conditional effects
goright <sub>a</sub>	$\top$	$\{\langle p_{a,k}, \{p_{a,k+1}\}, \{p_{a,k}\} \rangle : 1 \leq k \leq 3\}$
goleft <sub>a</sub>	$\top$	$\{\langle p_{a,k}, \{p_{a,k-1}\}, \{p_{a,k}\} \rangle : 2 \leq k \leq 4\}$
sense <sub>a,q</sub>	$p_{a,2}$	$\{\langle \top, \{S_a q\}, \emptyset \rangle\}$
tell <sub>a,q</sub>	$S_a q$	$\{\langle \varphi_b, \{S_b q\}, \emptyset \rangle, \langle \varphi_c, \{S_c q\}, \emptyset \rangle\}$

Table 5.2: Action descriptions for the selective communication task

Let us augment the set of actions by adding a further action  $\text{informs}_{1,2}m$  whose precondition is that both agents are in the same place, that the speaker 1 knows whether  $m$  and that it is possible for her that  $m$  is informative for the hearer 2. Let the resulting set of actions be  $Act''$ . Then the sequence

$$sol_3 = \langle \text{enter}_1, \text{reveal}_1, \text{leave}_1, \text{informs}_{1,2}m \rangle$$

becomes a further solution of  $\langle Act'', s_0, Goal \rangle$  as well as of  $\langle Act'', s_0, Goal' \rangle$ . Observe that  $sol_3$  is shorter than the other solutions.

Let us finally modify  $Act''$  into  $Act'''$  by a more demanding precondition of informing, namely by requiring that the speaker knows that her utterance is relevant for the hearer. Then  $sol_3$  is no longer a solution of  $\langle Act'', s_0, Goal \rangle$  nor of  $\langle Act'', s_0, Goal' \rangle$ : the speaker lacks knowledge about the hearer. It becomes a solution again if we make the hypothesis that there is initial common knowledge of ignorance about  $m$ , i.e., if we set  $s'_0 = \{m, JS S_1 m, JS S_2 m\}$ .

### 5.3.3 Selective communication

The following example is from Kominis and Geffner:

Let  $a$ ,  $b$ , and  $c$  be three agents in a corridor of four rooms ( $p_1$ ,  $p_2$ ,  $p_3$  and  $p_4$  from left to right). The agents can move from a room to a contiguous room, and when agent  $i$  communicates (tells) some information, all the agents that are in the same room or in a contiguous room, will hear what was communicated. For example, if agent  $i$  expresses in  $p_3$  his knowledge about  $q$ , all agents in rooms  $p_2$ ,  $p_3$  and  $p_4$  will come to know it. We consider the problem where agent  $a$  is initially in room  $p_1$ ,  $b$  in  $p_2$ ,  $c$  in  $p_3$ , and  $a$  has to find out the truth value of a proposition  $q$  and let  $c$  know without agent  $b$  learning it [80].

As in [80], we suppose that  $a$  can only learn the value of  $q$  by a sensing action in room 2. Let  $p_{i,k}$  mean that agent  $i$  is in room  $p_k$ , for  $i \in \{a, b, c\}$  and  $1 \leq k \leq 4$ . Then the formula  $\varphi_b = \bigvee_{1 \leq k, \ell \leq 4, |k-\ell| \leq 1} (p_{a,k} \wedge p_{b,\ell})$  expresses that  $b$  is close to  $a$ , and likewise for  $\varphi_c$ . We model this as an EL-O planning task  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  with  $Act = \{\text{goleft}_a, \text{goright}_a, \text{sense}_{a,q}, \text{tell}_{a,q}\}$ ,  $s_0 = \{p_{a,1}, p_{b,2}, p_{c,3}\}$ ,  $Goal = S_c q \wedge \neg S_b q$ . The descriptions of the actions in  $Act$  are listed in Table 5.2. Then the sequence

$$\langle \text{goleft}_a, \text{sense}_{a,q}, \text{goright}_a, \text{goright}_a, \text{goright}_a, \text{tell}_{a,q} \rangle$$

is a solution of  $\mathcal{P}$ .

action	precondition	conditional effects
$\text{drivesOn}_c$	$d_c \wedge \neg r_c$	$\{\langle \top, \{r_c\}, \{d_c, S_c wr_p\} \rangle\}$
$\text{stopsAndSignals}_c$	$d_c \wedge wr_p \wedge S_c wr_p$	$\{\langle \top, \{JS wr_p\}, \{d_c\} \rangle\}$
$\text{restarts}_c$	$\neg d_c \wedge r_p$	$\{\langle \top, \{d_c\}, \emptyset \rangle\}$
$\text{crosses}_p$	$wr_p \wedge ((\neg d_c \wedge JS wr_p) \vee r_c)$	$\{\langle \top, \{r_p\}, \{wr_p\} \rangle\}$

Table 5.3: Action descriptions for the pedestrian crossing task

### 5.3.4 Pedestrian crossing

In this 2-agent coordination problem that is derived from Lewis's [89], a pedestrian wants to cross a road at the same time as a car wants to reach its destination. The driver can either continue or stop to let the pedestrian cross. In the latter case the driver signals to the pedestrian, for example by flashing the car's headlights, that it is safe for the pedestrian to cross. We assume that the intentions of the pedestrian are obvious to the driver while the car is approaching (but not after it has driven away since the driver will not be looking in the rear-view mirror), but the pedestrian does not know this. Thus, before the pedestrian crosses the road, common knowledge must be established between the two agents (pedestrian and driver) of the pedestrian's intention to cross.

We can model this problem using the following propositional variables:  $d_c$  is true if the car is driving forward,  $r_c$  is true if the car has reached its destination;  $r_p$  is true if the pedestrian has reached the other side of the road,  $wr_p$  is true if the pedestrian wants to reach the other side of the road.

Then the planning task can be described as follows. The agents are  $p$  (the pedestrian) and  $c$  (the car). The initial state is  $s_0 = \{d_c, wr_p, S_c wr_p\}$ , the goal is  $Goal = r_p \wedge r_c$ , and the set of actions  $Act$  has four elements, three concerning the car and one the pedestrian, whose preconditions and effects are listed in Table 5.3. The action  $\text{drivesOn}_c$  has a precondition  $d_c \wedge \neg r_c$  and (unique) conditional effect that the car reaches its destination, stops driving and no longer sees the value of  $wr_p$  (whether the pedestrian wants to cross or not). Note that the fact that  $S_c wr_p$  becomes false implies that  $JS wr_p$  becomes false because the latter is an atomic cause of the former, i.e., because  $JS wr_p \in (S_c wr_p)^\leftarrow$ . The action  $\text{stopsAndSignals}_c$  has a precondition  $d_c \wedge wr_p \wedge S_c wr_p$  and (unique) conditional effect that the car stops and signals (by flashing its lights) that the driver knows that the pedestrian wants to cross which leads immediately to common knowledge of  $wr_p$ . The third possible action of the car is to restart (after a stop). Finally, the only action of the pedestrian is  $\text{crosses}_p$ .

Then the two sequences of actions

$$\begin{aligned} &\langle \text{drivesOn}_c, \text{crosses}_p \rangle, \\ &\langle \text{stopsAndSignals}_c, \text{crosses}_p, \text{restarts}_c, \text{drivesOn}_c \rangle \end{aligned}$$

both solve our planning task  $\langle Act, s_0, Goal \rangle$ .

### 5.3.5 Authorisation via plenary meetings

Here is a further example of actions with common knowledge preconditions. Consider a planning task involving cooperation between different agents that

action	precondition	conditional effects
doTask <sub>ik</sub>	$a_{ik}$	$\{\langle \top, \{t_{ik}, S_i t_{ik}\}, \emptyset \rangle\}$
authorise <sub>k</sub>	$\bigwedge_{i \in Agt} JS t_{i,k-1}$	$\{\langle \top, \{a_{ik} : i \in Agt\}, \{p_i : i \in Agt\} \rangle\}$
goMeeting <sub>i</sub>	$\top$	$\{\langle \top, \{p_i\}, \emptyset \rangle\}$
announce <sub>i</sub>	$\bigwedge_{k \in Agt} p_k$	$\{\langle t_{ik} \wedge S_i t_{ik}, \{JS t_{ik}\}, \emptyset \rangle\}$

Table 5.4: Action descriptions for the authorisation via plenary meetings task

can be divided into  $m$  different stages with tasks to be performed at each stage by each agent. Agents are only authorised to start stage  $k + 1$  if *all* tasks of stage  $k$  have been completed and all agents have common knowledge of this. The only way this can be achieved is by having a plenary meeting at the end of each stage during which each agent announces that their stage- $k$  task has been completed.<sup>2</sup>

Each agent  $i$  requires an authorisation  $a_{ik}$  to start their stage- $k$  task. Initially,  $a_{i1}$  is true for all agents  $i$ , i.e.,  $s_0 = \{a_{i1} : i \in Agt\}$ . Let  $t_{ik}$  represent the fact that agent  $i \in Agt$  has completed their stage- $k$  task. The propositional variable  $p_i$  is true if agent  $i$  is present at the meeting.

The preconditions and effects of the four action types doTask<sub>ik</sub>, authorise<sub>k</sub>, goMeeting<sub>i</sub>, and announce<sub>i</sub> are described in Table 5.4. For the action doTask<sub>ik</sub>, each agent sees when they have accomplished their task, but not when other agents accomplish theirs. The action authorise<sub>k</sub> ( $k = 2, \dots, m$ ) authorises the start of stage  $k$ . Authorisations for stage  $k > 1$  are only issued if there is common knowledge that all tasks of stage  $k - 1$  have been completed. They are thus issued at the end of each plenary meeting. A side effect of this is that all agents leave the meeting. The action goMeeting<sub>i</sub> allows each agent to go to the meeting room whenever they wants. Once everybody is present, each agent  $i$  can announce that their task has been completed via the action announce<sub>i</sub>.

The goal is common knowledge that all tasks have been completed:

$$Goal = \bigwedge_{i \in Agt} \bigwedge_{k=1, \dots, m} (t_{ik} \wedge JS t_{ik}).$$

With 2 agents and 2 stages, the following sequence is a solution:

$$\langle \text{doTask}_{11}, \text{doTask}_{21}, \text{goMeeting}_1, \text{goMeeting}_2, \text{announce}_1, \text{announce}_2, \\ \text{authorise}_2, \text{doTask}_{12}, \text{doTask}_{22}, \text{goMeeting}_1, \text{goMeeting}_2, \text{announce}_1, \text{announce}_2 \rangle.$$

### 5.3.6 The generalised gossip problem

Let the set of agents be  $Agt = \{1, \dots, n\}$  and let the set of propositional variables be  $Prop = \{s_i : i \in Agt\}$ . Let us describe the planning task for the generalised gossip problem of depth  $k$  as an EL-O planning task  $G_k = \langle Act^{G_k}, s_0^{G_k}, Goal^{G_k} \rangle$ . The initial state is  $s_0^{G_k} = \{S_i s_i : i \in Agt\} \cup S$  for some  $S \subseteq \{s_i : i \in Agt\}$  (cf. Example 3.1). The goal of obtaining shared knowledge of depth  $k$  can be

<sup>2</sup>The need for common knowledge can be interpreted as agents not only needing to know that they have communicated about the completion of their own task, but also needing to know that that communication was successful, similarly to the Byzantine Generals' Problem.



described in EL-O by

$$Goal^{G_k} = \bigwedge_{i \in Agt} \bigwedge_{\sigma \in OBS^+, \ell(\sigma) \leq k} \sigma s_i.$$

The set of actions is  $Act^{G_k} = \{\text{call}_j^i : i, j \in Agt, i \neq j\}$ , where  $pre(\text{call}_j^i) = \top$  and where for every  $0 \leq m < k$ , every  $\sigma_m \in OBS^{\leq m}$  of length at most  $m$  and every  $r \in Agt$  there are conditional effects  $ce \in \text{eff}(\text{call}_j^i)$  of the form:

$$\begin{aligned} \text{end}(ce) &= S_i \sigma_m s_r \vee S_j \sigma_m s_r; \\ \text{ceff}^+(ce) &= \{\sigma S_i \sigma_m s_r : \sigma \in \{S_i, S_j\}^{\leq k-m-1}\} \\ &\quad \cup \{\sigma S_j \sigma_m s_r : \sigma \in \{S_i, S_j\}^{\leq k-m-1}\} \\ &= \{\sigma \sigma_m s_r : \sigma \in \{S_i, S_j\}^{\leq k-m}\}; \\ \text{ceff}^-(ce) &= \emptyset; \end{aligned}$$

where  $\{S_i, S_j\}^{\leq k-m}$  denotes the set all sequences of observability operators  $S_i$  and  $S_j$  of length at most  $k-m$ . For  $k=1$  we obtain  $\text{end}(ce) = S_i s_r \vee S_j s_r$ ,  $\text{ceff}^+(ce) = \{S_i s_r, S_j s_r\}$ , and  $\text{ceff}^-(ce) = \emptyset$ , matching Example 5.2.

We recall that the original gossip problem with  $n \geq 4$  agents can be solved in  $2(n-2)$  calls [11, 121, 64] and that there is no shorter solution. It is known that the generalised gossip problem  $G_k$  can be solved in at most  $(k+1)(n-2)$  calls [73] and that there is no shorter solution [41]. For instance, for  $k=2$  and  $n=5$  the sequence

$$\text{call}_3^1, \text{call}_4^1, \text{call}_5^2, \text{call}_5^1, \text{call}_3^1, \text{call}_4^2, \text{call}_4^1, \text{call}_5^1, \text{call}_3^2$$

is a solution with  $3 \times 3 = 9$  calls, which is therefore optimal.

### 5.3.7 Generalised gossiping with ignorance goals

We can also easily model ignorance goals. Given a set of atoms  $A$ , let  $G\text{-neg}_{k,A} = \langle Act^{G\text{-neg}_{k,A}}, s_0^{G\text{-neg}_{k,A}}, Goal^{G\text{-neg}_{k,A}} \rangle$  be the planning task for the generalised gossip problem of depth  $k$  with the atoms of  $A$  as the only negative goals. The initial state and the actions remain the same:  $s_0^{G\text{-neg}_{k,A}} = \{S_i s_i : i \in Agt\} \cup S$  for some  $S \subseteq \{s_i : i \in Agt\}$  and  $Act^{G\text{-neg}_{k,A}} = \{\text{call}_j^i : i, j \in Agt, i \neq j\}$ ; but the goal changes:

$$Goal^{G\text{-neg}_{k,A}} = \left( \bigwedge_{\sigma \in \{S_i : i \in Agt\}^{\leq k}} \bigwedge_{j \in Agt, \sigma s_j \notin A} \sigma s_j \right) \wedge \left( \bigwedge_{\alpha \in A} \neg \alpha \right).$$

Here are some examples:

- $G\text{-neg}_{1, \{S_1 s_2\}}$  corresponds to the case where we want everyone to know all secrets, except that agent 1 should not know the secret of agent 2;
- $G\text{-neg}_{1, \{S_1 s_3, S_2 s_4\}}$ , to the case where agent 1 should not get to know the secret of agent 3 and agent 2, the secret of agent 4;
- $G\text{-neg}_{2, \{S_1 s_2 s_3\}}$ , to the case of epistemic depth 2 where only agent 1 should not know whether agent 2 knows the secret of agent 3, while agents 1 and 2 should know the secret of agent 3;

- $G\text{-neg}_{3,\{s_1 s_2 s_3\}}$  is the same as  $G\text{-neg}_{2,\{s_1 s_2 s_3\}}$  but with depth 3.

It was shown in [43] that the existence of a solution for gossiping with ignorance goals is an NP-complete problem.

## 5.4 Conclusion

We have defined EL-O sequential and parallel planning through conditional actions with conditional effects very similar to actions in classical planning. We have given two semantics for EL-O actions: one in the general case, working with infinite introspective states, and one which allows us to work with finite states when restricting the considered vocabulary. In general, this vocabulary will be that of the planning task at hand. We have shown the equivalence of these two semantics, and used the second to show our main result for this section: deciding the existence of a plan given an EL-O planning task is PSPACE-complete, the same complexity class as classical planning. This is the case both for sequential and parallel planning. We have then illustrated our framework on a number of application examples, both original and from the literature.

An encoding of EL-O action descriptions into the standard Planning Domain Definition Language (PDDL) as well as some experimental results can be found in [39] and [42].

## Chapter 6

# Around EL-O planning

### Autour de la planification EL-O

Dans le chapitre précédent nous avons défini les bases de la planification EL-O. Nous étudions maintenant ce cadre plus en profondeur, de ses possibles ramifications à ses liens avec d'autres modèles de planification épistémique. Tout d'abord, nous étudions la possibilité d'étendre les définitions et résultats du Chapitre 5 aux autres versions de EL-O présentées dans cette thèse : la version avec constantes, EL-OC, et la version qui suit l'axiome d'induction pour la connaissance commune, EL-O+*Ind*. Nous définissons ensuite une extension dynamique de EL-OC en y ajoutant les opérateurs dynamiques de la logique DL-PPA [74], et montrons que dans cette logique les actions EL-O peuvent être réduites à des programmes et la solvabilité d'une tâche de planification devient un problème de vérification de modèle. Enfin, nous développons une comparaison entre les actions EL-O et les modèles à événements de DEL, ainsi qu'entre EL-O et d'autres modélisations simplifiées de planification épistémique.

---

We have defined the basics of EL-O planning in Chapter 5. In this chapter we further investigate this framework, both in its possible ramifications and in its relation to other epistemic planning frameworks. We first discuss possibilities of extending the definitions and results of Chapter 5 to the other versions of EL-O that we have presented throughout this thesis: the version with constants EL-OC and the version which follows the induction axiom for common knowledge EL-O+*Ind*. We then take the EL-OC framework and define a dynamic extension of it based on the logic DL-PPA of [74]. Finally, we attempt to situate our framework with regard to DEL as well as other simple epistemic planning frameworks of the literature.

This chapter is mainly based on the following publications:

- Cooper, M. C., Herzig, A., Maffre, F., Maris, F., Perrotin, E., & Régnier, P. (2020). *A lightweight epistemic logic and its application to planning*. Artificial Intelligence, 103437.
- Herzig, A., Maris, F., & Perrotin, E. (2021). *A dynamic epistemic logic with finite iteration and parallel composition*. Proceedings of the Inter-

national Conference on Principles of Knowledge Representation and Reasoning (KR 2021) (Vol. 18, No. 1, pp. 676-680).

## Contents

<b>6.1 Discussion: planning with other versions of EL-O</b>	<b>85</b>
6.1.1 EL-OC . . . . .	85
6.1.2 EL-O+ <i>Ind</i> . . . . .	86
<b>6.2 A dynamic version of EL-OC: DEL-PPAOC . . . . .</b>	<b>87</b>
6.2.1 Language of DEL-PPAOC . . . . .	87
6.2.2 Semantics of DEL-PPAOC . . . . .	87
6.2.3 Action descriptions . . . . .	88
6.2.4 DEL-PPAOC planning tasks . . . . .	90
<b>6.3 Relation to planning with with DEL event models</b>	<b>91</b>
6.3.1 From DEL event models to EL-O action descriptions	91
6.3.2 From EL-O action descriptions to DEL event models	94
6.3.3 Discussion: DEL vs. EL-O . . . . .	96
<b>6.4 Other related works . . . . .</b>	<b>98</b>
6.4.1 Muise et al. . . . .	98
6.4.2 Kominis and Geffner . . . . .	100
6.4.3 Petrick and Bacchus . . . . .	101
<b>6.5 Conclusion . . . . .</b>	<b>101</b>

## 6.1 Discussion: planning with other versions of EL-O

### 6.1.1 EL-OC

Going from EL-O planning to EL-OC planning is quite simple. Actions and planning tasks are simply defined over  $ATM-C$  instead of  $ATM$ . That is, an EL-OC action description is a pair  $\mathbf{a} = \langle pre(\mathbf{a}), eff(\mathbf{a}) \rangle$  where  $pre(\mathbf{a}) \in \mathcal{L}_{\text{bool}}(ATM-C)$  and  $eff(\mathbf{a}) \subseteq \mathcal{L}_{\text{bool}}(ATM-C) \times 2^{ATM-C} \times 2^{ATM-C}$ , and an EL-O epistemic planning task, or epistemic planning task over  $ATM-C$ , is a triple  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  where  $Act$  is a set of EL-O action descriptions,  $s_0 \in 2^{ATM-C}$  is the initial state, and  $Goal \in \mathcal{L}_{\text{bool}}(ATM-C)$  is the goal formula. Consistency conditions and semantics are the same as those of EL-O, simply replacing all instances of  $\models_{\text{EL-O}}$  with  $\models_{\text{EL-OC}}$ .

**Example 6.1** (Lights on, ctd.). The door is now locked with a code  $c$ . An agent must know  $c$  in order to enter the room.

$$\text{enter}_i = \langle \neg \text{In}_i \wedge S_i c, \{ \langle \top, \{ \text{In}_i, S_i \text{ On} \}, \emptyset \rangle \\ \langle \text{In}_j, \{ JS \text{ On} \}, \emptyset \rangle \} \rangle$$

We can also imagine that seeing agent  $i$  enter the room will lead agent  $j$  to know that agent  $i$  knows the code for the door, and that agents might communicate about the code, all of which is easily expressed in EL-OC.

**Example 6.2** (Gossip, ctd.). We consider the variant of the gossip problem with  $n$  agents in which agents must know another agent's telephone number in order to call them, and agents can communicate about telephone numbers as well as secrets. The set of constants is  $Cst = \{s_i : i \in Agt\} \cup \{t_i : i \in Agt\}$ . We only consider knowledge of depth 1 as in the original gossip problem. The action  $callNumber_j^i$  of agent  $i$  calling agent  $j$  is described as follows:

$$\begin{aligned} callNumber_j^i = & \langle S_i t_j, \{ \langle S_i s_1 \vee S_j s_1, \{ S_i s_1, S_j s_1 \}, \emptyset \rangle \\ & \langle S_i t_1 \vee S_j t_1, \{ S_i t_1, S_j t_1 \}, \emptyset \rangle \\ & \dots \\ & \langle S_i s_n \vee S_j s_n, \{ S_i s_n, S_j s_n \}, \emptyset \rangle \\ & \langle S_i t_n \vee S_j t_n, \{ S_i t_n, S_j t_n \}, \emptyset \rangle \} \rangle \end{aligned}$$

The expansion of EL-OC planning tasks is defined exactly as the one for EL-O planning tasks. It can therefore be proved in the same way as before that:

**Proposition 6.1.** *Solvability of an EL-O planing task, both by a sequential and a parallel plan, is PSPACE-complete.*

### 6.1.2 EL-O+Ind

Working with EL-O+Ind planning is a little bit trickier as one must be careful to stay within EL-O+Ind states, and more precisely members of  $I-IND-STATES$ . Here the language is that of EL-O, meaning that EL-O+Ind actions and planning tasks are exactly EL-O actions and planning tasks. The difference is in the semantics. An EL-O+Ind action  $a$  induces a partial function  $\tau_a^{EL-O+Ind}$  on  $I-IND-STATES$ . If  $s \in I-IND-STATES$ , the function  $\tau_a^{EL-O+Ind}$  is defined at  $s$  if  $s \models pre(a)$ . In that case we say that  $a$  is executable at  $s$  and stipulate that  $\tau_a^{EL-O+Ind}(s) = C(\tau_a(s)) \Rightarrow$ , where  $C$  is the completion function defined in Chapter 4 (Section 4.4.2).

**Example 6.3** (Lights on, ctd.). It is natural to suppose that both agents know who is inside and outside the room at all times. This means that they always know exactly who knows whether the light in the room is turned on. Supposing that this is common knowledge, we can consider the state  $s = \{JS S_1 On, JS S_2 On, S_1 On, ln_1\}$  in which only agent 1 is in the room and the light is off. We can then define the action of an agent entering the room as follows:

$$enter'_i = \langle \neg ln_i, \{ \langle \top, \{ ln_i, S_i On \}, \emptyset \rangle \} \rangle$$

If agent 2 enters the room, the current state becomes:

$$\begin{aligned} \tau_{enter'_i}^{EL-O+Ind}(s) &= C(\tau_{enter'_i}(s)) \Rightarrow \\ &= C(\{JS S_1 On, JS S_2 On, S_1 On, ln_1, S_2 On, ln_2\}) \Rightarrow \\ &= \{JS S_1 On, JS S_2 On, S_1 On, ln_1, S_2 On, ln_2, JS On\} \Rightarrow \\ &= \{ln_1, ln_2\} \cup \{\sigma On : \sigma \in OBS^+\} \end{aligned}$$

This change in semantics makes translating actions to classical actions more difficult. It is still an open problem whether EL-O+Ind planning can be polynomially translated into classical planning.

## 6.2 A dynamic version of EL-OC: DEL-PPAOC

A dynamic extension of EL-O, DEL-PAO, was proposed in [70], and a modeling of planning tasks in DEL-PAO with no common knowledge and no parallel plans was given in [40]. We here give a fuller and more succinct dynamic logic, adding in particular an operator of parallel composition as well as an operator of inclusive nondeterministic composition. Both are imported from Dynamic Logic of Parallel Propositional Assignments DL-PPA [74]. In the resulting logic, which we call DEL-PPAOC, the solvability of various planning problems can be captured: the existence of sequential and parallel plans, both in the unbounded and in the finite horizon version. The operator of inclusive nondeterministic composition turns out to be instrumental for the succinct modeling of parallel planning.

### 6.2.1 Language of DEL-PPAOC

The language of DEL-PPAOC extends the language of EL-OC with the dynamic operator  $\langle \pi \rangle$ , where  $\pi$  is a program. Programs  $\pi$  and formulas  $\varphi$  are defined by the following grammar:

$$\begin{aligned} \varphi & ::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle \pi \rangle \varphi, \\ \pi & ::= \alpha \leftarrow \varphi \mid \varphi? \mid \pi; \pi \mid \pi \cup \pi \mid \pi \sqcup \pi \mid \pi \sqcap \pi \mid \pi^*, \end{aligned}$$

where  $\alpha$  ranges over the set of atomic formulas  $ATM-C$ . The formula  $\langle \pi \rangle \varphi$  reads “there is a possible execution of  $\pi$  such that  $\varphi$  is true afterwards”. The program  $\alpha \leftarrow \varphi$  assigns the truth value of  $\varphi$  to  $\alpha$ .  $\varphi?$  tests whether  $\varphi$  is true (and fails when  $\varphi$  is false).  $\pi_1; \pi_2$  executes  $\pi_1$  and  $\pi_2$  in sequence.  $\pi_1 \cup \pi_2$  nondeterministically chooses between executing either  $\pi_1$  or  $\pi_2$ ; and  $\pi_1 \sqcup \pi_2$  nondeterministically chooses between executing either  $\pi_1$ , or  $\pi_2$ , or both.  $\pi_1 \sqcap \pi_2$  is the parallel composition of  $\pi_1$  or  $\pi_2$ . The set of all formulas is  $Fm_{DEL-PPAOC}$ . As per usual, the set of atoms occurring in a program  $\pi$  is denoted by  $Voc(\pi)$ . For example,  $Voc(JS p \leftarrow (S_i q \vee S_j S_i r)) = \{JS p, S_i q, S_j S_i r\}$ . The formula  $[\pi]\varphi$  abbreviates  $\neg \langle \pi \rangle \neg \varphi$  and is therefore read “ $\varphi$  is true after every possible execution of  $\pi$ ”. We define  $n$ -times iteration of  $\pi$  by induction on  $n$ :  $\pi^0 = \top?$  and  $\pi^{n+1} = \pi; \pi^n$ . From there, we define  $\pi^{\leq n}$  as  $\bigcup_{0 \leq i \leq n} \pi^i$ .  $\pi^*$  is the unbounded iteration of  $\pi$ .

### 6.2.2 Semantics of DEL-PPAOC

The interpretation of a program  $\pi$  is a ternary relation on the set of valuations:  $\|\pi\| \subseteq 2^{ATM-C} \times 2^{ATM-C} \times 2^{ATM-C}$ . When  $\langle s, U, W \rangle \in \|\pi\|$  then there is an execution of  $\pi$  from state  $s$  to state  $U$  assigning the variables in  $W$ . The interpretation function is defined by mutual recursion, from the interpretation of programs given in Table 6.1 and the following interpretation of formulas:

$$\begin{aligned} s \models \alpha & \quad \text{iff } \alpha \in s^{\Rightarrow} \cup I-ATM-C, \\ s \models \langle \pi \rangle \varphi & \quad \text{iff there are } U, W \text{ such that } \langle s, U, W \rangle \in \|\pi\| \text{ and } U \models \varphi, \end{aligned}$$

and as usual for boolean operators.

The interpretation of the assignment  $\alpha \leftarrow \varphi$  is that either (1) the value of  $\varphi$  is true,  $\alpha$  gets the value true, and the set of assigned variables is the singleton  $\{\alpha\}$ , or (2) the value of  $\varphi$  is false, all causes of  $\alpha$  get the value false, and the set of assigned variables is the set  $\alpha^{\leftarrow}$  of all causes of  $\alpha$ .

$$\begin{aligned}
\|\alpha \leftarrow \varphi\| &= \{\langle s, s \cup \{\alpha\}, \{\alpha\} \rangle : s \models \varphi\} \cup \{\langle s, s \setminus \alpha^{\Leftarrow}, \alpha^{\Leftarrow} \rangle : s \not\models \varphi\}, \\
\|\varphi?\| &= \{\langle s, s, \emptyset \rangle : s \models \varphi\}, \\
\|\pi_1; \pi_2\| &= \left\{ \langle s, U, W \rangle : \begin{array}{l} \text{there are } U_1, W_1, W_2 \text{ such that} \\ \langle s, U_1, W_1 \rangle \in \|\pi_1\|, \langle U_1, U, W_2 \rangle \in \|\pi_2\|, \\ \text{and } W = W_1 \cup W_2 \end{array} \right\}, \\
\|\pi_1 \cup \pi_2\| &= \|\pi_1\| \cup \|\pi_2\|, \\
\|\pi_1 \sqcup \pi_2\| &= \|\pi_1\| \cup \|\pi_2\| \cup \|\pi_1 \sqcap \pi_2\|, \\
\|\pi_1 \sqcap \pi_2\| &= \left\{ \langle s, U, W \rangle : \begin{array}{l} \text{there are } U_1, W_1, U_2, W_2 \text{ such that} \\ \langle s, U_1, W_1 \rangle \in \|\pi_1\|, \langle s, U_2, W_2 \rangle \in \|\pi_2\|, \\ W_1 \cap W_2 \cap U_1 = W_1 \cap W_2 \cap U_2, \\ U = (s \setminus W) \cup (U_1 \cap W_1) \cup (U_2 \cap W_2), \\ \text{and } W = W_1 \cup W_2 \end{array} \right\}, \\
\|\pi^*\| &= \bigcup_{k \in \mathbb{N}_0} \|\pi^k\|.
\end{aligned}$$

Table 6.1: Interpretation of DEL-PPAOC programs

The interpretation of parallel composition  $\pi_1 \sqcap \pi_2$  is that each subprogram  $\pi_k$  is executed locally; then it is checked whether the modifications (in terms of assigned variables) are compatible: this is the case when all variables that are assigned by both subprograms (namely the variables in  $W_1 \cap W_2$ ) get assigned the same truth value. If this is not the case then the parallel composition fails; otherwise the resulting valuation  $U$  is computed by putting together (1) the unchanged part of  $s$  (i.e.,  $s \setminus W$ ), (2) the updates of  $\pi_1$  (i.e.,  $U_1 \cap W_1$ ), (3) the updates of  $\pi_2$  (i.e.,  $U_2 \cap W_2$ ). Moreover, the set of variables  $W$  assigned by a parallel composition is the union of the sets of variables assigned by the subprograms.

The interpretation of inclusive nondeterministic composition  $\pi_1 \sqcup \pi_2$  is the exclusive nondeterministic composition of the three programs  $\pi_1$ ,  $\pi_2$  and  $\pi_1 \sqcap \pi_2$ .

In the rest of this section we show how to model EL-O planning tasks in DEL-PPAOC.

### 6.2.3 Action descriptions

DEL-PPAOC actions are defined just like EL-OC actions, except that preconditions and conditions of conditional effects are formulas of the full dynamic language  $Fml_{\text{DEL-PPAOC}}$ . That is, a DEL-PPAOC action description is a pair  $\mathbf{a} = \langle pre(\mathbf{a}), eff(\mathbf{a}) \rangle$  where  $pre(\mathbf{a})$  is a formula from  $Fml_{\text{DEL-PPAOC}}$  (the precondition of  $\mathbf{a}$ ) and  $eff(\mathbf{a}) \subseteq Fml_{\text{DEL-PPAOC}} \times 2^{ATM-C} \times 2^{ATM-C}$  is the set of conditional effects of  $\mathbf{a}$ . Consistency and semantics of actions and sets of atoms are defined just as for EL-O actions. In particular, we consider semantics over a finite set of atoms  $Voc \in ATM-C$ , following the formalism of Section 5.2.1. We call  $\tau_{\mathbf{A}}^{\bullet Voc}$  the interpretation function of a set of actions  $\mathbf{A}$  in DEL-PPAOC relative to  $Voc$ .

To every such vocabulary and set of conditional actions we can associate a DEL-PPAOC program which behaves exactly like the parallel execution of its elements. First, execution of a single action  $\mathbf{a}$  relative to a finite set of atoms

$Voc$  can be simulated by the DEL-PPAOC program  $\text{exeAct}(Voc, \mathbf{a})$  defined when  $\mathbf{a}$  is consistent. In that case  $\text{exeAct}(Voc, \mathbf{a})$  is defined as follows:

$$\text{exeAct}(Voc, \mathbf{a}) = \text{pre}(\mathbf{a})? \sqcap \prod_{ce \in \text{eff}(\mathbf{a})} \left( \begin{array}{c} \neg \text{cnd}(ce)? \\ \cup \left( \begin{array}{c} \text{cnd}(ce)? \\ \sqcap \prod_{\alpha \in \text{ceff}^+(ce) \Rightarrow \cap Voc} \alpha \leftarrow \top \\ \sqcap \prod_{\alpha \in \text{ceff}^-(ce) \Leftarrow \cap Voc} \alpha \leftarrow \perp \end{array} \right) \end{array} \right).$$

The formula

$$\begin{aligned} \text{NoCrossInt}(Voc, \mathbf{a}, \mathbf{a}') &= \langle \text{exeAct}(Voc, \mathbf{a}') \rangle \text{pre}(\mathbf{a}) \wedge \\ &\quad \bigwedge_{ce \in \text{eff}(\mathbf{a})} (\text{cnd}(ce) \leftrightarrow \langle \text{exeAct}(Voc, \mathbf{a}') \rangle \text{cnd}(ce)) \end{aligned}$$

then expresses (given that  $(\text{Voc}(\mathbf{a}) \cup \text{Voc}(\mathbf{a}')) \subseteq Voc$ ) that neither executability nor effects of  $\mathbf{a}$  are sensitive to the execution of  $\mathbf{a}'$ , i.e., that  $\mathbf{a}$  and  $\mathbf{a}'$  have no cross-interaction in a given state. Then to every finite set of atoms  $Voc$  and set of actions  $\mathbf{A}$  we associate the DEL-PPAOC program

$$\text{exeAct}(Voc, \mathbf{A}) = \bigwedge_{\mathbf{a}, \mathbf{a}' \in \mathbf{A}, \mathbf{a} \neq \mathbf{a}'} \text{NoCrossInt}(Voc, \mathbf{a}, \mathbf{a}')?; \prod_{\mathbf{a} \in \mathbf{A}} \text{exeAct}(Voc, \mathbf{a}).$$

**Proposition 6.2.** *For every finite set of actions  $\mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_m\}$  and finite set of atoms  $Voc \subseteq \text{ATM-C}$  such that  $\bigcup_{1 \leq k \leq m} \text{Voc}(\mathbf{a}_k) \subseteq Voc$ ,*

1.  $\tau_{\mathbf{A}}^{\bullet Voc}$  is defined in  $s$  iff there exist some  $U, W$  such that  $\langle s, U, W \rangle \in \|\text{exeAct}(Voc, \mathbf{A})\|$ ;
2. If  $\tau_{\mathbf{A}}^{\bullet Voc}$  is defined in  $s$  then  $\tau_{\mathbf{A}}^{\bullet Voc}(s) = U$  iff  $\langle s, U, W \rangle \in \|\text{exeAct}(Voc, \mathbf{A})\|$  for some  $W$ .

*Proof.* Consider first a single action:  $\mathbf{A} = \{\mathbf{a}\}$ . Note that in this case we have  $\|\text{exeAct}(Voc, \mathbf{A})\| = \|\text{exeAct}(Voc, \mathbf{a})\|$ . Let  $s$  be an arbitrary state.  $\tau_{\mathbf{a}}^{\bullet Voc}(s)$  is not defined iff either (1)  $\mathbf{a}$  is inconsistent, in which case  $\text{exeAct}(Voc, \mathbf{a})$  is undefined, or (2)  $s \not\models \text{pre}(\mathbf{a})$ , in which case the program fails because  $\|\text{pre}(\mathbf{a})?\| = \emptyset$ . When  $\tau_{\mathbf{a}}^{\bullet Voc}(s)$  is defined then  $s \models \text{pre}(\mathbf{a})$ , so  $\langle s, s, \emptyset \rangle \in \|\text{pre}(\mathbf{a})?\|$ . Moreover, for each  $ce \in \text{eff}(\mathbf{a})$  such that  $s \models \text{cnd}(ce)$  the programs  $(\prod_{\alpha \in \text{ceff}^+(ce) \Rightarrow \cap Voc} \alpha \leftarrow \top)$  and  $(\prod_{\alpha \in \text{ceff}^-(ce) \Leftarrow \cap Voc} \alpha \leftarrow \perp)$  are executed in parallel and all the assignments are consistent (no  $\alpha \leftarrow \perp$  and  $\alpha \leftarrow \top$  are executed in parallel for the same  $\alpha$  by consistency of  $s$ ). Then the parallel composition of all these programs leads, by definition, to the state  $\tau_{\mathbf{a}}^{\bullet Voc}(s) = U$ , with  $\langle s, U, W \rangle \in \|\text{exeAct}(Voc, \mathbf{a})\|$ , where  $U$  is the set of variables assigned to  $\top$  and  $W$  the set of all assigned variables in the program  $\text{exeAct}(Voc, \mathbf{a})$ .

Consider now now  $\mathbf{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_m\}$  with  $m \geq 2$ . Let  $s$  be an arbitrary state. We have already shown that for all  $i \in \{1, \dots, m\}$ ,  $\text{exeAct}(Voc, \mathbf{a}_i)$  behaves correctly and produces the same effects as the single action  $\{\mathbf{a}_i\}$ , and is executable iff  $\tau_{\mathbf{a}_i}^{\bullet Voc}(s)$  is defined. If two actions  $\mathbf{a}, \mathbf{a}' \in \mathbf{A}$  have contradictory effects in state  $s$ , there are  $ce \in \text{eff}(\mathbf{a})$  and  $ce' \in \text{eff}(\mathbf{a}')$  and  $\alpha \in \text{ATM-C}$  such that  $s \models \text{cnd}(ce) \wedge \text{cnd}(ce')$  and  $(\text{ceff}^+(ce) \cap \text{ceff}^-(ce') \Leftarrow) \cup (\text{ceff}^-(ce) \Leftarrow \cap \text{ceff}^+(ce'))$  contains  $\alpha$ , and then the program fails because of the execution of the parallel composition of  $\alpha \leftarrow \perp$  and  $\alpha \leftarrow \top$  in the program  $\text{exeAct}(Voc, \mathbf{a}) \sqcap \text{exeAct}(Voc, \mathbf{a}')$ .



If actions  $\mathbf{a}$  and  $\mathbf{a}'$  are executable at  $s$  and have cross-interaction at  $s$ , then either  $\tau_{\mathbf{a}'}^{\bullet, Voc}(s) \not\models pre(\mathbf{a})$  or there is a conditional effect  $ce \in eff(\mathbf{a})$  such that  $s$  and  $\tau_{\mathbf{a}'}^{\bullet, Voc}(s)$  disagree on  $cond(ce)$ , or the symmetric cases (swapping  $\mathbf{a}$  and  $\mathbf{a}'$ ). In the first case, the execution of  $\langle exeAct(Voc, \mathbf{a}')pre(\mathbf{a})? \rangle$  in  $NoCrossInt(Voc, \mathbf{a}, \mathbf{a}')$  fails, and in the second, the execution of  $\langle cond(ce) \leftrightarrow \langle exeAct(Voc, \mathbf{a}') \rangle cond(ce) \rangle?$  in  $NoCrossInt(Voc, \mathbf{a}, \mathbf{a}')$  fails. Finally, the parallel composition of all these programs check contradictory effects and cross-interaction and leads to the state  $\tau_A^{\bullet, Voc}(s)$  by definition because of the parallel execution of all  $exeAct(Voc, \mathbf{a}_i)$ .  $\square$

### 6.2.4 DEL-PPAOC planning tasks

A DEL-PPAOC planning task is a triple  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  where  $Act$  is a finite set of consistent actions,  $s_0 \in 2^{ATM-C}$  is a finite state (the initial state) and  $Goal \in Fml_{DEL-PPAOC}$  is a DEL-PPAOC formula. Solvability is defined as usual.

**Theorem 6.1.** *A planning task  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  is solvable by a parallel plan with no more than  $k$  steps if and only if:*

$$s_0 \models \left\langle \left( \prod_{\mathbf{a} \in Act} x_{\mathbf{a}} \leftarrow \perp; \bigsqcup_{\mathbf{a} \in Act} x_{\mathbf{a}} \leftarrow \top; \pi_x \right)^{\leq k} \right\rangle Goal$$

where the  $x_{\mathbf{a}}$  are fresh variables ( $x_{\mathbf{a}} \notin Voc(\mathcal{P})$  for all  $\mathbf{a} \in Act$ ) and where

$$\begin{aligned} \pi_x = & \bigwedge_{\mathbf{a}, \mathbf{a}' \in Act, \mathbf{a}' \neq \mathbf{a}} ((x_{\mathbf{a}} \wedge x_{\mathbf{a}'}) \rightarrow NoCrossInt(Voc(\mathcal{P}), \mathbf{a}, \mathbf{a}'))? ; \\ & \prod_{\mathbf{a} \in Act} (\neg x_{\mathbf{a}}? \cup (x_{\mathbf{a}}?; exeAct(Voc(\mathcal{P}), \mathbf{a}))). \end{aligned}$$

*Proof.* The program  $\prod_{\mathbf{a} \in Act} x_{\mathbf{a}} \leftarrow \perp$  initialises a special fresh variable  $x_{\mathbf{a}} \notin Voc(\mathcal{P})$  to  $\perp$ , for each action  $\mathbf{a} \in Act$ . Then the inclusive nondeterministic composition  $\bigsqcup_{\mathbf{a} \in Act} x_{\mathbf{a}} \leftarrow \top$  chooses some non empty subset of actions  $A \subseteq Act$  and executes the program  $\prod_{\mathbf{a} \in A} x_{\mathbf{a}} \leftarrow \top$ . At this point,  $x_{\mathbf{a}} = \top$  iff  $\mathbf{a} \in A$ , and the program  $\pi_x$  is executed. It is easily seen that, for a given chosen set of actions  $A$ ,  $\pi_x$  behaves like the program  $exeAct(Voc(\mathcal{P}), A)$ . We know by Proposition 6.2 that the latter program behaves correctly and produces the same effect as the parallel execution of all actions in  $A$ . The sequence  $\prod_{\mathbf{a} \in Act} x_{\mathbf{a}} \leftarrow \perp; \bigsqcup_{\mathbf{a} \in Act} x_{\mathbf{a}} \leftarrow \top; \pi_x$  is then repeated a number of times lesser or equal than  $k$ . This produces a sequence of at most  $k$  parallel executions of action sets, i.e., a parallel plan bounded by  $k$ . Therefore the formula is satisfied in the initial state if and only if there exists a parallel plan of length bounded by  $k$  after which the goal is satisfied, i.e., if and only if the planning task is solvable with a sequence of at most  $k$  parallel steps.  $\square$

Solvability by a sequential plan is the special case where the parallel plan is a sequence of singletons.

**Theorem 6.2.** *A planning task  $\langle Act, s_0, Goal \rangle$  is solvable by a sequential plan with no more than  $k$  actions if and only if:*

$$s_0 \models \left\langle \left( \bigcup_{\mathbf{a} \in Act} exeAct(Voc(\mathcal{P}), \mathbf{a}) \right)^{\leq k} \right\rangle Goal.$$

*Proof.* Our formula reads “there exists an execution of  $(\bigcup_{a \in Act} \text{exeAct}(a))^{\leq k}$  after which *Goal* is true.” We know by Proposition 6.2 that  $\text{exeAct}(\text{Voc}(\mathcal{P}), a)$  behaves correctly and produces the same effects as the single action  $a$ . The program  $(\bigcup_{a \in Act} \text{exeAct}(\text{Voc}(\mathcal{P}), a))^{\leq k}$  non-deterministically chooses an action  $a$  from *Act* and executes the corresponding program  $\text{exeAct}(\text{Voc}(\mathcal{P}), a)$ , then repeats this a number of times lesser or equal than  $k$ . This produces a sequence of at most  $k$  actions, i.e., a sequential plan bounded by  $k$ . Therefore the formula is satisfied in the initial state if and only if there exists a sequential plan of length bounded by  $k$  after which the goal is satisfied, i.e., if and only if the planning task is solvable with a sequence of at most  $k$  actions.  $\square$

We express solvability in the general case by replacing the  $k$ -bound with unbounded iteration:

**Theorem 6.3.** *A planning task  $\mathcal{P} = \langle Act, s_0, Goal \rangle$  is solvable by a parallel plan with no more than  $k$  steps if and only if:*

$$s_0 \models \left\langle \left( \prod_{a \in Act} x_a \leftarrow \perp; \bigsqcup_{a \in Act} x_a \leftarrow \top; \pi_x \right)^* \right\rangle Goal$$

where the  $x_a$  are fresh variables ( $x_a \notin \text{Voc}(\mathcal{P})$  for all  $a \in Act$ ) and where

$$\begin{aligned} \pi_x = & \bigwedge_{a, a' \in Act, a' \neq a} ((x_a \wedge x_{a'}) \rightarrow \text{NoCrossInt}(\text{Voc}(\mathcal{P}), a, a'))? ; \\ & \prod_{a \in Act} (\neg x_a? \cup (x_a?; \text{exeAct}(\text{Voc}(\mathcal{P}), a))). \end{aligned}$$

*A planning task  $\langle Act, s_0, Goal \rangle$  is solvable by a sequential plan with no more than  $k$  actions if and only if:*

$$s_0 \models \left\langle \left( \bigcup_{a \in Act} \text{exeAct}(\text{Voc}(\mathcal{P}), a) \right)^* \right\rangle Goal.$$

## 6.3 Relation to planning with with DEL event models

We show that EL-O action descriptions can capture several important kinds of DEL event models (precisely, public, private and semi-private announcements) and the other way round, we show for some of our action descriptions how they correspond to DEL event models. One might expect that all such descriptions have a corresponding DEL event model; however, the correspondence is not obvious due to fundamental differences in the way epistemic effects are modeled. We explain this in detail in Section 6.3.3.

### 6.3.1 From DEL event models to EL-O action descriptions

In what follows we show that we can express as EL-O actions several kinds of event models introduced in Chapter 2 (Section 2.3.5) and recalled in Figure 6.1, where we sometimes restrict preconditions to literals and conjunctions thereof. We are going to associate an EL-O action description  $a(\text{Evt}, e)$  to a given pointed event model  $(\text{Evt}, e)$ . In all cases, the precondition of  $a(\text{Evt}, e)$  is  $\text{pre}^{\text{Evt}}(e)$ .

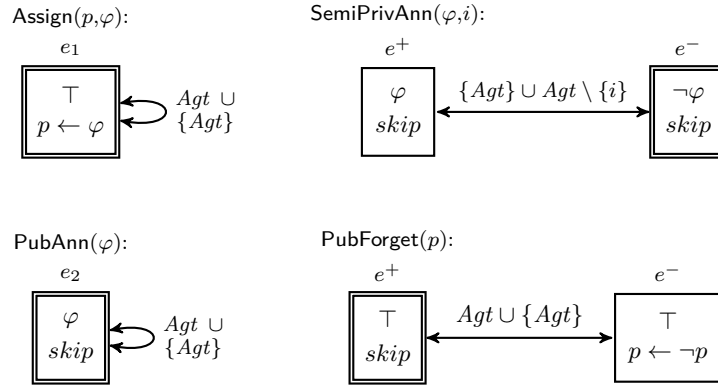


Figure 6.1: The event models  $\text{Assign}(p, \varphi)$ ,  $\text{PubAnn}(\varphi)$ ,  $\text{SemiPrivAnn}(\varphi, i)$ , and  $\text{PubForget}(p)$ .

(More precisely, it is the reduction of  $\text{pre}^{\text{Evt}}(e)$  to an  $\mathcal{L}_{\text{bool}}(\text{ATM})$  formula obtained by exhaustively applying the reduction axioms of Section 3.3.3.) Hence  $(M^{\text{EL-O}}, s) \models \text{pre}^{\text{Evt}}(e)$  exactly when  $\tau_{\mathbf{a}(\text{Evt}, e)}^{\text{EL-O}}$  is defined at  $s$ . For such points  $s$ , it can be shown that the pointed product update  $(M^{\text{EL-O}} \otimes \text{Evt}, (s, e))$  is bisimilar to the pointed Kripke model  $(M^{\text{EL-O}}, \tau_{\mathbf{a}(\text{Evt}, e)}^{\text{EL-O}}(s))$ . Recall that bisimulation is defined in Section 2.2.6.

First, let us look at some kinds of public assignments. The public assignment of  $\top$  to  $p$ ,  $\text{Assign}(p, \top)$ , is captured by the action description  $\mathbf{a}(\text{Assign}(p, \top), e_1) = \langle \top, \{\langle \top, \{JS p, p\}, \emptyset \rangle\} \rangle$ . The public assignment of  $\perp$  to  $p$ ,  $\text{Assign}(p, \perp)$ , is captured by the action  $\mathbf{a}(\text{Assign}(p, \perp), e_1) = \langle \top, \{\langle \top, \{JS p\}, \{p\} \rangle\} \rangle$ . Finally, the public toggling of the truth value of  $p$ ,  $\text{Assign}(p, \neg p)$ , is captured by an action description with two conditional effects

$$\mathbf{a}(\text{Assign}(p, \neg p), e_1) = \langle \top, \{\langle \neg p, \{p\}, \emptyset \rangle, \langle p, \emptyset, \{p\} \rangle\} \rangle.$$

Observe that, contrarily to  $\text{Assign}(p, \top)$  and  $\text{Assign}(p, \perp)$ ,  $\text{Assign}(p, \neg p)$  does not modify  $JS p$ : if an agent does not know whether  $p$  before the public toggling of  $p$  then they also do not know whether  $p$  afterwards.

Second, let us look at some kinds of public announcements. The truthful public announcement of a propositional variable  $p$  is captured by the action description with precondition  $p$ , unconditional positive effect  $JS p$ , and without negative effects:  $\mathbf{a}(\text{PubAnn}(p), e_2) = \langle p, \{\langle \top, \{JS p\}, \emptyset \rangle\} \rangle$ . Symmetrically, the public announcement of  $\neg p$  is captured by the action  $\mathbf{a}(\text{PubAnn}(\neg p), e_2) = \langle \neg p, \{\langle \top, \{JS p\}, \emptyset \rangle\} \rangle$ . More generally, we can capture the public announcement of conjunctions of atoms and negations of atoms:

$$\mathbf{a}(\text{PubAnn}((\bigwedge_{\alpha \in A^+} \alpha) \wedge (\bigwedge_{\alpha \in A^-} \neg \alpha)), e_2) = \langle (\bigwedge_{\alpha \in A^+} \alpha) \wedge (\bigwedge_{\alpha \in A^-} \neg \alpha), \{\langle \top, \{JS \alpha : \alpha \in A^+ \cup A^-\}, \emptyset \rangle\} \rangle.$$

The precondition guarantees truthfulness of the announcement. Here are some examples of public announcements of the above conjunctions. We also add the

equivalent formulation in the standard epistemic language.

$$\begin{aligned}
\mathbf{a}(\text{PubAnn}(\neg K_i p \wedge \neg K_i \neg p), e_2) &= \mathbf{a}(\text{PubAnn}(\neg S_i p), e_2) \\
&= \langle \neg S_i p, \{\langle \top, \{JS S_i p\}, \emptyset \rangle\} \rangle, \\
\mathbf{a}(\text{PubAnn}(K_i p), e_2) &= \mathbf{a}(\text{PubAnn}(p \wedge S_i p), e_2) \\
&= \langle p \wedge S_i p, \{\langle \top, \{JS p\}, \emptyset \rangle\} \rangle, \\
\mathbf{a}(\text{PubAnn}(p \wedge \neg K_i p), e_2) &= \mathbf{a}(\text{PubAnn}(p \wedge \neg S_i p), e_2) \\
&= \langle p \wedge \neg S_i p, \{\langle \top, \{JS p\}, \emptyset \rangle\} \rangle.
\end{aligned}$$

In the announcement of  $p \wedge S_i p$  and of the Moore sentence  $p \wedge \neg S_i p$  we have dropped  $JS S_i p$  from the add-list because the latter already contains  $JS p$  (cf. the EL-O axiom  $Vis_5$ ).

Third, the semi-private announcement to  $i$  that  $p$  is true is captured by the action description

$$\mathbf{a}(\text{SemiPrivAnn}(p, i), e^+) = \langle p, \{\langle \top, \{S_i p, JS S_i p\}, \emptyset \rangle\} \rangle.$$

The effect is that  $i$  sees whether  $p$  and that all agents jointly see that; in other words, that it becomes common knowledge that  $i$  sees whether  $p$ .

Fourth, the event model  $\text{PubForget}(p)$  of publicly forgetting  $p$  corresponds to the action description

$$\mathbf{a}(\text{PubForget}(p), n) = \langle \top, \{\langle \top, \{JS S_i p : i \in \text{Agt}\}, \{S_i p : i \in \text{Agt}\} \rangle\} \rangle.$$

That is, all  $S_i p$  become false and this becomes common knowledge.

**Proposition 6.3.** *For each of the above event models  $\text{Evt}$ :*

1.  $(M^{\text{EL-O}}, s) \models \text{pre}^{\text{Evt}}(e)$  if and only if  $\tau_{\mathbf{a}(\text{Evt}, e)}^{\text{EL-O}}$  is defined at  $s$ ;
2. If  $\tau_{\mathbf{a}(\text{Evt}, e)}^{\text{EL-O}}(s)$  is defined then  $(M^{\text{EL-O}}, \tau_{\mathbf{a}(\text{Evt}, e)}^{\text{EL-O}}(s))$  and  $(M^{\text{EL-O}} \otimes \text{Evt}, (s, e))$  are bisimilar.

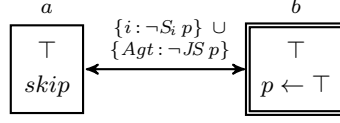
*Proof.* The first item is the case because, as already mentioned, the precondition of the action descriptions  $\mathbf{a}(\text{Evt}, e)$  are all equivalent to  $\text{pre}^{\text{Evt}}(e)$ .

We sketch the proof of the second item for the public announcement of  $p$ . Let  $M^{\text{EL-O}} \otimes \text{PubAnn}(p) = \langle W, R_{\text{Agt}}, \{R_i\}_{i \in \text{Agt}}, V \rangle$ , and consider  $s \in W^{\text{EL-O}}$  such that  $\tau_{\mathbf{a}(\text{PubAnn}(p), e)}^{\text{EL-O}}(s)$  is defined. Then  $\tau_{\mathbf{a}(\text{PubAnn}(p), e)}^{\text{EL-O}}(s) = s \cup \{JS p\}^{\Rightarrow}$ . We show that the pointed models  $(M^{\text{EL-O}}, \tau_{\mathbf{a}(\text{PubAnn}(p), e)}^{\text{EL-O}}(s))$  and  $(M^{\text{EL-O}} \otimes \text{PubAnn}(p), (s, e_2))$  are bisimilar, where  $e_2$  is the single event of the event model  $\text{PubAnn}(p)$ . To that end we define the relation  $Z$  between  $W^{\text{EL-O}}$  and  $W$  as:  $s'Z(s'', e_2)$  iff  $s' = s'' \cup \{JS p\}^{\Rightarrow}$ . It can be checked that the three conditions for  $Z$  being a bisimulation are satisfied: atomic, forth, and back.

The other proofs are similar; we only sketch the case of public forgetting of  $p$ ,  $\text{PubForget}(p)$ . As this action has a trivial precondition,  $\tau_{\mathbf{a}(\text{PubForget}(p))}^{\text{EL-O}}(s)$  is defined for all  $s$ . We show that given a state  $s \in W^{\text{EL-O}}$ , the pointed models  $(M^{\text{EL-O}}, \tau_{\mathbf{a}(\text{PubForget}(p))}^{\text{EL-O}}(s))$  and  $(M^{\text{EL-O}} \otimes \text{PubForget}(p), (s, e^+))$  are bisimilar via the following relation  $Z$  between  $W$  and  $W^{\text{EL-O}}$ :

$$\begin{aligned}
Z &= \{ \langle \tau_{\mathbf{a}(\text{PubForget}(p))}^{\text{EL-O}}(s'), (s', e^+) \rangle : s' \in W^{\text{EL-O}} \} \cup \\
&\quad \{ \langle \tau_{\mathbf{a}(\text{PubForget}(p))}^{\text{EL-O}}(s') \cup \{p\}, (s', e^-) \rangle : s' \in W^{\text{EL-O}} \text{ and } p \notin s' \} \cup \\
&\quad \{ \langle \tau_{\mathbf{a}(\text{PubForget}(p))}^{\text{EL-O}}(s') \setminus \{p\}, (s', e^-) \rangle : s' \in W^{\text{EL-O}} \text{ and } p \in s' \}.
\end{aligned}$$

It can be checked that  $Z$  is indeed a bisimulation.  $\square$

Figure 6.2: An event model equivalent to  $\langle \top, \{\langle \top, \{p\}, \emptyset \rangle\}$ .

Something that cannot be modeled in general in our framework is the public announcement that  $\neg K_i p$ , or, expressed in the EL-O fragment, that  $\neg(p \wedge S_i p)$ . The reason is that it is not clear how the disjunctive effect of that announcement could be described in terms of an add-list.

### 6.3.2 From EL-O action descriptions to DEL event models

We now consider the converse direction: given an EL-O action description  $\mathbf{a}$ , can we find a DEL event model  $(\text{Evt}, e)$  such that for any state  $s \in W^{\text{EL-O}}$ , the product  $(M^{\text{EL-O}} \otimes \text{Evt}, (s, e))$  of  $(M^{\text{EL-O}}, s)$  and  $(\text{Evt}, e)$  is defined iff  $s \models \text{pre}(\mathbf{a})$ , and such that in that case,  $(M^{\text{EL-O}} \otimes \text{Evt}, (s, e))$  and  $(M^{\text{EL-O}}, \tau_{\mathbf{a}}^{\text{EL-O}}(s))$  are bisimilar?

Let us here restrict our attention to a few very basic EL-O action descriptions in which there is a single conditional effect either adding or deleting one atom. The idea behind the construction of an equivalent event model is the following: when adding or deleting an atom from a world, only agents who see this atom know that it has been modified.

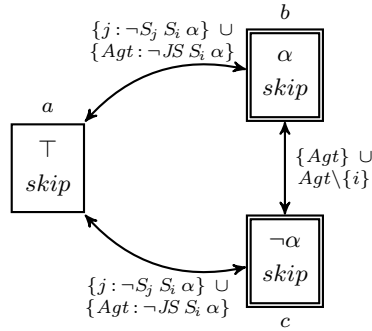
The easiest way to translate this into DEL event models is through the use of Bolander's *edge-conditioned event models* [25]. Such models incorporate agents on accessibility relations, so that a model is a tuple  $\text{Evt} = \langle W^{\text{Evt}}, R_{Agt}^{\text{Evt}}, \{R_i^{\text{Evt}}\}_{i \in Agt}, \text{pre}^{\text{Evt}}, \text{post}^{\text{Evt}} \rangle$  where  $R_{Agt}^{\text{Evt}} : W^{\text{Evt}} \times W^{\text{Evt}} \rightarrow \text{Fml}_{\text{EL}}$ ,  $R_i^{\text{Evt}} : W^{\text{Evt}} \times W^{\text{Evt}} \rightarrow \text{Fml}_{\text{EL}}$  for every  $i \in Agt$ , and the rest is defined as previously. The product of such an edge-conditioned event model with a model  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  is  $M \otimes \text{Evt} = \langle W', R'_{Agt}, \{R'_i\}_{i \in Agt}, V' \rangle$  where

$$\begin{aligned} (s, e)R'_{Agt}(t, f) & \text{ iff } sR_{Agt}t \text{ and } (M, s) \models R_{Agt}^{\text{Evt}}(e, f) \text{ and } (M, t) \models R_{Agt}^{\text{Evt}}(e, f); \\ (s, e)R'_i(t, f) & \text{ iff } sR_it \text{ and } (M, s) \models R_i^{\text{Evt}}(e, f) \text{ and } (M, t) \models R_i^{\text{Evt}}(e, f); \end{aligned}$$

and the rest is defined as previously. As mentioned in [25], edge-conditioned event models can be simulated by standard event models, though the former are more succinct.

We now give a few examples of translations of simple EL-O action descriptions into edge-conditioned event models. We will only be giving the bisimulation relations without the proofs that they are indeed bisimulations, as those proofs are tedious but uncomplicated.

An equivalent event model to the assignment of  $p$  to true,  $\langle \top, \{\langle \top, \{p\}, \emptyset \rangle\}$ , is given in Figure 6.2. The bisimulation relation  $Z$  is as follows: for every state  $s \in W^{\text{EL-O}}$ ,  $sZ(s, a)$  and  $(s \cup \{p\})Z(s, b)$ . If we wish to assign  $p$  to false rather than true (action  $\langle \top, \{\langle \top, \emptyset, \{p\} \rangle\}$ ), we simply replace the assignment  $p \leftarrow \top$  in the actual event by  $p \leftarrow \perp$ , and change the bisimulation relation so that  $(s \setminus \{p\})Z(s, b)$  for any  $s \in W^{\text{EL-O}}$ . This action can be interpreted as the value

Figure 6.3: An event model equivalent to  $\langle \top, \{\langle \top, \{S_i \alpha\}, \emptyset \rangle\} \rangle$ .

of  $p$  (the property of an object) changing, while only agents who are looking at this object can see the change.

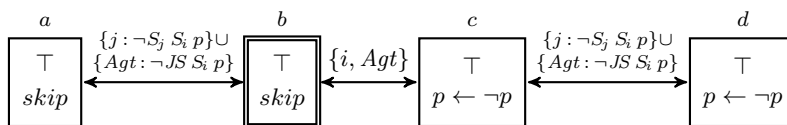
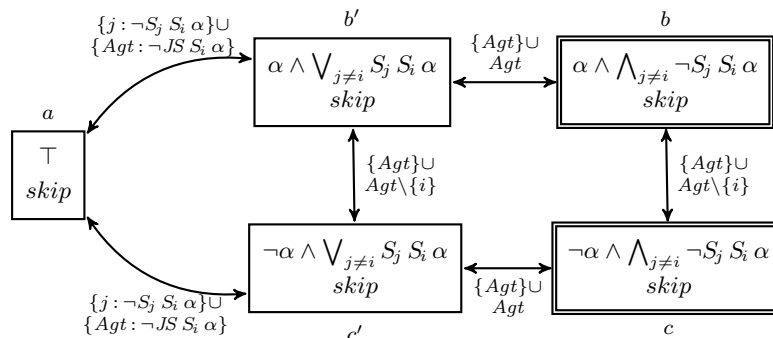
For the action  $\langle \top, \{\langle \top, \{S_i \alpha\}, \emptyset \rangle\} \rangle$  making  $S_i \alpha$  true, we need in the general case two actual events, one for when  $\alpha$  is true and one for when it is false. An event model equivalent to this action description is given in Figure 6.3. The corresponding bisimulation  $Z$  is the following: for every state  $s \in W^{\text{EL-O}}$ ,

- $sZ(s, a)$ ;
- $(s \cup \{S_i \alpha\})Z(s, b)$  if  $s \models \alpha$ ;
- $(s \cup \{S_i \alpha\})Z(s, c)$  if  $s \models \neg \alpha$ .

A more complex example, given in Figure 6.4, is that of an event model equivalent to an action making the atom  $S_i p$  false. This can be interpreted as agent  $i$  looking away from  $p$ . Agents who saw agent  $i$  looking at  $p$  will see them look away, and the other agents will not be aware of this change. While this is a fairly natural action to consider, it is difficult to generalize to assigning  $S_i \alpha$  to false for any given  $\alpha$ , as only propositional variables can receive assignments in DEL event models. The bisimulation  $Z$  in the case of Figure 6.4 is as follows: for every state  $s \in W^{\text{EL-O}}$ ,

- $sZ(s, a)$ ;
- $(s \setminus \{JS p, S_i p\})Z(s, b)$ ;
- $(s \setminus \{JS p, S_i p, p\})Z(s, c)$  if  $s \models p$ ;
- $((s \cup \{p\}) \setminus \{JS p, S_i p\})Z(s, c)$  if  $s \models \neg p$ ;
- $(s \setminus \{p\})Z(s, d)$  if  $s \models p$ ;
- $(s \cup \{p\})Z(s, d)$  if  $s \models \neg p$ .

Let us give one more example, this time involving a precondition: private announcements. In an **S5** setting, an announcement can only be private if no one is watching (as we cannot represent agents “believing that nothing has happened”). That is, we can only privately announce whether  $\alpha$  is true to an

Figure 6.4: An event model equivalent to  $\langle \top, \{\langle \top, \emptyset, \{S_i p\}\rangle\} \rangle$ .Figure 6.5: An event model equivalent to the private announcement whether  $\alpha$  to  $i$ .

agent  $i$  if for all  $j \neq i$ ,  $\neg S_j S_i \alpha$  holds. The corresponding action description has condition  $\bigwedge_{j \neq i} \neg S_j S_i \alpha$  and a single conditional effect  $\langle \top, \{S_i \alpha\}, \emptyset \rangle$ . The corresponding event model is given in Figure 6.5. The bisimulation  $Z$  is: for every state  $s \in W^{\text{EL-O}}$ ,  $sZ(s, a)$ , and  $(s \cup S_i \alpha)Z(s, e)$  if  $s \models \text{pre}^{\text{Evt}}(e)$  for  $e \in \{b, b', c, c'\}$ .

We stop here without delving into the territory of deleting longer atoms or combining several effects. It is, we hope, fairly obvious that given an action description, the construction of an equivalent DEL event model is neither straightforward nor systematic. This, we believe, is an argument in favour of our framework, in which many actions can be defined very succinctly and in a natural manner.

### 6.3.3 Discussion: DEL vs. EL-O

The comparison with DEL leads us to a fundamental question: how should an action be described? All the approaches in the literature advocating DEL for epistemic planning presuppose that event models provide an appropriate tool for the description of actions. This, however, seems questionable to us. For example, how should we describe the action of pushing a light button? It is useful for our discussion to recall the fundamental distinction in philosophy of action between *action types* and *action tokens* [10]. Pushing a light button is an example of the former, while an example of the latter is the pushing of a particular button at a particular point in time by a particular agent. The elements of the set of action descriptions  $Act$  are clearly action types. In contrast, the elements of the execution of a solution to a planning task are typically action

tokens.<sup>1</sup>

Let us suppose that the action type of pushing a light button can be described by means of preconditions and effects. Reasonably, the precondition is that the agent is close to the switch<sup>2</sup> and the effect is that the light is on. However, a DEL event model has to contain more information than that: it also has to say how the button-pushing is perceived (or not) by each of the agents. Now remember that we have to describe an action type: we have to account for all possible circumstances of button pushing. For each agent there are at least two cases: the one where the agent is present and observes the light bulb and the one where they are absent. Altogether, this requires an exponential numbers of points in an event model. Worse, we also have to take care of the evolution of the agents' higher-order knowledge: for each couple of observing agents  $i$  and  $j$  we have to distinguish whether  $i$  sees that  $j$  observes the bulb, and so on, the limit case being when there is joint observation of the bulb. Hence exhaustive action descriptions by means of DEL event models systematically have to have an infinite number of points in order to take all the possible effects into account. This can actually be seen as an epistemic version of the ramification problem [101].

Contrasting with DEL event models, our EL-O-based action descriptions do not require to fully describe all possible ramifications: the direct effect of pushing the light button is simply that the light is on, and the indirect epistemic effects follow from the description of the state, namely who observes the light bulb and who doesn't. This is a fundamentally different answer to the question of how actions should be described. It makes the modeller's task much simpler and more natural. (A similar argument was first put forward in [30].)

At the core of this distinction between EL-O and DEL is the observation that DEL models get bigger the more uncertainty there is, while the opposite is true for EL-O; and that EL-O models give a straightforward account of the knowledge in a situation, while DEL reasons about *possibilities* (which worlds are possible for each agent, which event may have happened according to each agent, etc.). One or the other of the approaches may then be better suited depending on the problem at hand, with EL-O being stronger when reasoning with more uncertainty and straightforward actions of learning and forgetting as well as more general action types, while DEL is better suited to situations with more common knowledge and more convoluted events with specific action tokens.

Of course, DEL remains more expressive than EL-O, and not all actions can be modeled in EL-O. For example, the public announcement of the disjunction  $p \vee q$  (the event model  $\text{PubAnn}(p \vee q)$ ) or the assignment of  $q$  to  $p$  (the event model  $\text{Assign}(p, q)$ ) cannot be captured. This is because all propositional variables are independent in our approach, as witnessed by the fact that the knowledge operator distributes over disjunctions of literals: for different  $p$  and  $q$ ,  $K_i(p \vee q)$  is equivalent to  $K_i p \vee K_i q$ , and  $K_i(p \leftrightarrow q)$  is equivalent to  $(K_i p \wedge K_i q) \vee (K_i \neg p \wedge K_i \neg q)$ . This is clearly a limitation of our approach. In particular, we cannot capture the muddy children problem, in which the children learn that one of them has a muddy forehead without knowing who. Note that while DEL

<sup>1</sup>In DEL-based planning this distinction is clarified for the first time in [90].

<sup>2</sup>There might be more, such as that the wires leading to the light bulb are working, that the light bulb is not broken etc.: we here neglect the qualification problem [101] which is that the precondition is typically an infinite conjunction.



is more general as far as preconditions are concerned (any complex formula can be announced), its postconditions are restricted: they assign propositional variables and therefore cannot model actions with disjunctive ontic effects such as  $p \vee q$ .

For the same reason we cannot capture the poisonous liquid example from [110]. In this problem there is a bottle of liquid, a healthy lawn, and two actions: `pour` and `senseLawn`. The first is the action of pouring some liquid on the lawn that we have already described in Example 5.3 (Section 5.1.1). The second senses whether or not the lawn is dead: we have  $pre(\text{senseLawn}) = \top$  and  $eff(\text{senseLawn}) = \{\top, \{S_i \text{ Dead}\}, \emptyset\}$ . Then for the set of actions  $Act = \{\text{pour}, \text{senseLawn}\}$ , there is no solution of the planning task

$$\langle ATM, Act, \{\neg \text{Dead}, S_i \text{ Dead}\}, S_i \text{ Poisonous} \rangle.$$

In particular, pouring some liquid on the lawn and checking whether the lawn is dead does not inform the agent whether `Poisonous` is the case. This is because after pouring the liquid the agent should know that the state of the lawn is tied to the toxicity of the liquid. Such knowledge cannot be captured in our framework.

## 6.4 Other related works

### 6.4.1 Muise et al.

Muise et al. [107] follow a strategy that is very similar to ours, applying the logic of proper epistemic knowledge bases (PEKBs) of [84] to simple epistemic planning. In [84], formulas are restricted to boolean combinations of so-called restricted modal literals (RMLs): sequences of belief operators and negations that are followed by a propositional variable; in other terms, formulas of  $Fml_{EL}$  without conjunctions and disjunctions. Formally, the grammar for RMLs is:

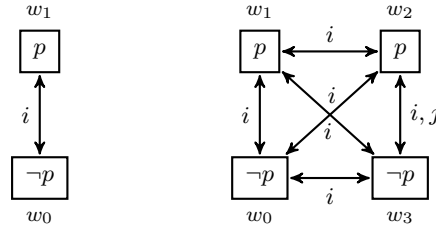
$$\lambda ::= p \mid \neg \lambda \mid K_i \lambda$$

where  $p$  ranges over  $Prop$  and  $i$  over  $Agt$ . PEKBs are conjunctions of such RMLs. Muise et al. have also started to consider the integration of ‘knowing whether’ operators in [103], but have not integrated this into the planning formalism of [107].

The major advantage of our EL-O-based approach over the RMLs-based approach is that boolean combinations of RMLs cannot express  $K_i(K_j p \vee K_j \neg p) \wedge \neg K_i p \wedge \neg K_i \neg p$  is true, that is,  $i$  knows that  $j$  knows whether  $p$  while  $i$  does not know about  $p$  herself. As we have argued in Chapter 1, such situations are important in interaction and more specifically in communication. They can be expressed in our framework by  $\mathcal{L}_{\text{bool}}(ATM)$  formulas of the form  $S_i S_j p \wedge \neg S_i p$ . To prove this formally it suffices to show that  $K_i(K_j p \vee K_j \neg p)$  cannot be expressed by boolean combinations of RMLs.

**Proposition 6.4.** *Let  $i$  and  $j$  be different. Then there is no boolean combination of RMLs  $\varphi$  that is equivalent to  $K_i(K_j p \vee K_j \neg p)$  in S5 Kripke models.*

*Proof.* Consider the two S5 Kripke models  $M = \langle W, R_{Agt}, \{R_i\}_{i \in Agt}, V \rangle$  and

Figure 6.6: Models  $M$  (left) and  $M'$  (right); reflexive arrows omitted.

$M' = \langle W', R'_{Agt}, \{R'_i\}_{i \in Agt}, V' \rangle$  with

$$\begin{aligned}
 W &= \{w_0, w_1\}, & W' &= \{w_0, w_1, w_2, w_3\}, \\
 R_{Agt} &= W \times W = R_i, & R'_{Agt} &= W' \times W' = R'_i, \\
 R_j &= \delta_W, & R'_j &= \delta_{W'} \cup \{\langle w_2, w_3 \rangle, \langle w_3, w_2 \rangle\}, \\
 V(w_0) &= \emptyset, & V'(w_0) &= V'(w_3) = \emptyset, \\
 V(w_1) &= \{p\}; & V'(w_1) &= V'(w_2) = \{p\};
 \end{aligned}$$

where  $\delta_W = \{\langle w_0, w_0 \rangle, \langle w_1, w_1 \rangle\}$  is the identity relation of  $W$  and  $\delta_{W'}$  is the identity relation of  $W'$ . They are depicted in Figure 6.6. Clearly,  $(M, w_0) \models K_i(K_j p \vee K_j \neg p)$  and  $(M', w_0) \not\models K_i(K_j p \vee K_j \neg p)$ . We prove that no boolean combinations of RMLs can tell the pointed models  $(M, w_0)$  and  $(M', w_0)$  apart. From that it immediately follows that there can be no such boolean combination that is equivalent to  $K_i(K_j p \vee K_j \neg p)$ .

To establish that of  $(M, w_0)$  and  $(M', w_0)$  satisfy the same formulas as far as the language of boolean combinations of RMLs is concerned it suffices to prove that  $(M, w_0) \models \lambda$  iff  $(M', w_0) \models \lambda$  for every RML  $\lambda$ . We use induction on the length of  $\lambda$ . We have to check the following exhaustive list of cases; each of them is straightforward.

**Case 1.**  $\lambda = q$  for some  $q \in Prop$ . Obvious. ( $(M, w_0) \models \lambda$  and  $(M', w_0) \models \lambda$ , regardless of whether  $q$  equals  $p$  or not.)

**Case 2.**  $\lambda = \neg \mu$ , for some RML  $\mu$ . Straightforward application of the induction hypothesis.

**Case 3.**  $\lambda = K_i \mu$ . We have the following subcases.

**Case 3.1.**  $\lambda = K_i q$  for some  $q \in Prop$ . Obvious. ( $(M, w_0) \models \lambda$  and  $(M', w_0) \models \lambda$ .)

**Case 3.2.**  $\lambda = K_i \neg \mu$ , for some RML  $\mu$ . We have to dive deeper into the subcases here.

**Case 3.2.1.**  $\lambda = K_i \neg q$ . Obvious. ( $(M, w_0) \models \lambda$  and  $(M', w_0) \models \lambda$ .)

**Case 3.2.2.**  $\lambda = K_i \neg \neg \mu$ . We use that  $\lambda \leftrightarrow K_i \mu$  is valid and apply the induction hypothesis.

**Case 3.2.3.**  $\lambda = K_i \neg K_i \mu$ . We use that  $K_i \neg K_i \mu$  and  $\neg K_i \mu$  are equivalent in S5 and apply the induction hypothesis.

**Case 3.2.4.**  $\lambda = K_i \neg K_j \mu$ . We do yet another case analysis.

**Case 3.2.4.1.**  $\lambda = K_i \neg K_j q$ . Obvious. ( $(M, w_0) \models \lambda$  and  $(M', w_0) \models \lambda$ .)

**Case 3.2.4.2.**  $\lambda = K_i \neg K_j \neg \mu$ . This is the last case analysis we have to do:

**Case 3.2.4.2.1.**  $\lambda = K_i \neg K_j \neg q$ . Obvious. ( $(M, w_0) \models \lambda$  and  $(M', w_0) \models \lambda$ .)

**Case 3.2.4.2.2.**  $\lambda = K_i \neg K_j \neg \mu$ . We use that  $\lambda \leftrightarrow K_i \neg K_j \mu$  is valid and apply the induction hypothesis.

**Case 3.2.4.2.3.**  $\lambda = K_i \neg K_j \neg K_i \mu$ . We have  $(M, w_0) \models \lambda \leftrightarrow K_i \mu$  and  $(M', w_0) \models \lambda \leftrightarrow K_i \mu$ , allowing us to apply the induction hypothesis.

**Case 3.2.4.2.4.**  $\lambda = K_i \neg K_j \neg K_j \mu$ . We use that  $K_i \neg K_j \neg K_j \mu$  and  $K_i K_j \mu$  are equivalent in S5 and apply the induction hypothesis.

**Case 3.2.4.3.**  $\lambda = K_i \neg K_j K_i \mu$ . We have  $(M, w_0) \models \lambda \leftrightarrow K_i \mu$  and  $(M', w_0) \models \lambda \leftrightarrow K_i \mu$ , allowing us to apply the induction hypothesis.

**Case 3.2.4.4.**  $\lambda = K_i \neg K_j K_j \mu$ . We use that  $K_i \neg K_j K_j \mu$  and  $K_i \neg K_j \mu$  are equivalent in S5 and apply the induction hypothesis.

**Case 3.3.**  $\lambda = K_i K_i \mu$ . We use that  $K_i K_i \mu$  and  $K_i \mu$  are equivalent in S5 and apply the induction hypothesis.

**Case 3.4.**  $\lambda = K_i K_j \mu$ . We have  $(M, w_0) \models \lambda \leftrightarrow K_i \mu$  and  $(M', w_0) \models \lambda \leftrightarrow K_i \mu$ , allowing us to apply the induction hypothesis.

**Case 4.**  $\lambda = K_j \mu$ . We have  $(M, w_0) \models \lambda \leftrightarrow \mu$  and  $(M', w_0) \models \lambda \leftrightarrow \mu$ , allowing us to apply the induction hypothesis.  $\square$

## 6.4.2 Kominis and Geffner

Kominis and Geffner's approach [80, 81] distinguishes three kinds of actions: physical actions modifying the world, public updates (that are nothing but DEL-like public announcements), and sensing actions by means of which an agent learns whether a formula is true or not.

On the one hand, their approach is more general than ours because agents can sense arbitrary formulas. This allows them to model the muddy children problem, which we cannot express as an EL-O planning task (see Section 6.3.3). However, if we restrict the formulas describing the set of initially possible states, the formulas that are publicly announced, and the formulas that are sensed to observability atoms (more precisely: to the equivalent  $Fml_{EL}$  formulas) then everything that can be modeled in Kominis and Geffner's approach can also be modeled in ours.

On the other hand, Kominis and Geffner's approach imposes three severe restrictions:

- all actions can be placed into one of the three categories;
- the set of initially possible states is common knowledge among all agents;
- all physical actions and all public updates are public;
- all sensing actions are semi-private.

The first hypothesis is clearly too strong for many natural everyday situations, as we have already said in Section 3.3.1 (Remark 3.4). The second hypothesis is also very strong: it forbids the modeling of private actions modifying the world as required in several of our examples of Section 5.3. The third hypothesis means that when an agent  $i$  senses the truth value of  $\varphi$  then all other agents see this: they learn that  $i$  knows whether  $\varphi$  is true or not but ignore whether what  $i$  has learned is  $\varphi$  or  $\neg\varphi$ . This means that there can be no private communication; in particular, one cannot model what we take to be a paradigmatic epistemic planning task, viz. the gossip problem.

### 6.4.3 Petrick and Bacchus

There is also an older approach due to Petrick and Bacchus about planning under incomplete information for a single agent [109, 110] which bears some similarities to ours. Their language is first-order but does not have epistemic operators. Instead, there are several kinds of knowledge bases by means of which the epistemic status of pieces of information is represented. In particular, there is a ‘knowing that’ database whose elements are literals and a ‘knowing whether’ database whose elements are atoms. The point of view is subjective, while ours is objective, i.e., the agent’s knowledge is represented, but not what is true in the world. Therefore the identity of  $K_i p$  and  $p \wedge Kif_i p$  cannot be expressed in their language. Just as the original proposal in [109], our approach does not account for postdiction (deducing the past from the present), as illustrated by Example 5.3. This is because in our approach the knowledge operator distributes over disjunctions of literals. This limitation of our approach is also the reason why we cannot account for the muddy children problem: the latter requires actions where the children learn a disjunction, namely that one of them is dirty. A solution to the muddy children problem where public announcements were integrated into DEL-PAO was presented in [36]; however, in that paper only the semantics was designed, without an axiomatization or complexity result. The study of such mathematical properties will be subject of future work.

## 6.5 Conclusion

We have discussed the extension of the results of Chapter 5 to EL-OC and EL-O+*Ind*, defined the logic DEL-PPAOC as a dynamic extension of EL-OC and discussed the relation between our framework and other related epistemic frameworks. A few open problems remain: first, it is not clear whether it is possible to polynomially translate EL-O+*Ind* planning tasks into classical planning tasks as we have done for standard EL-O planning tasks. This means that the complexity of the solvability problem for EL-O planning is not precisely known. Second, it would be interesting to investigate a systematic translation from EL-O action descriptions to DEL event models, though the few examples we have given here show that this is not at all straightforward.

An interesting avenue for future research are implicitly coordinated plans as introduced by Bolander et al. [28]. For example, suppose agent 2 would like to borrow the apartment of his friend 1 while 1 is away on vacation. Agent 1 has an action *putMat* of putting the key under the door mat that is described by  $pre(\text{putMat}) = \top$  and  $eff(\text{putMat}) = \{\langle \top, \{\text{Mat}\}, \emptyset \rangle\}$ ; agent 2 has an action *tryTake* of trying to take the key with  $pre(\text{tryTake}) = \top$  and  $eff(\text{tryTake}) = \{\langle \top, \{S_2 \text{ Mat}\}, \emptyset \rangle, \langle \text{Mat}, \{\text{hasKey}_2\}, \emptyset \rangle\}$ . The initial state is  $s_0 = \emptyset$ . Then 1 putting the key under the mat and 2 taking the key solves the planning task

$$\langle ATM, Act, s_0, \text{hasKey}_2 \rangle.$$

However, when 2 arrives at the apartment he will not know that the key is under the mat, unless 1 has told him. A better plan, baptised ‘implicitly coordinated’ in [28], involves 1’s action *checkMat* which can be described in EL-O terms by  $pre(\text{checkMat}) = \top$  and  $eff(\text{checkMat}) = \{\langle \top, \{S_2 \text{ Mat}\}, \emptyset \rangle\}$ . The formal characterisation of such plans requires a language in which it can be expressed that agents know whether a given sequence of actions will lead to a desired

state of the world: it is not enough that the action leads to the goal, the acting agent must also know that. This fails to be the case for 2's action `tryTake`: as 2 does not know whether the key is under the mat, he does not know whether the conditional effect  $\langle \text{Mat}, \{\text{hasKey}_2\}, \emptyset \rangle$  will be triggered. A dynamic extension of EL-O such as the one given in Section 6.2, DEL-PPAOC, should enable reasoning about implicitly coordinated plans. The details however remain to be worked out.

## Chapter 7

# True and mere belief about a proposition

La simplicité de EL-O repose sur deux facteurs : l'opérateur 'savoir si' et les capacités de raisonnement limitées des agents. Dans ce chapitre nous considérons une manière d'adapter ces propriétés à des logiques épistémico-doxastiques, dans lesquelles les agents n'ont pas seulement des croyances mais aussi des connaissances.

Le langage naturel n'admet pas de notion de 'croire si' ; nous traiterons donc dans ce chapitre de propositions de la forme "*i* a des connaissances à propos de  $\varphi$ " (*i* sait soit que  $\varphi$  est vrai, soit que  $\varphi$  est faux) et "*i* a une croyance à propos de  $\varphi$ " (c'est à dire que *i* croit soit que  $\varphi$  est vrai, soit que  $\varphi$  est faux). Formellement, nous écrivons  $Kif_i\varphi$  pour la première de ces notions et  $BA_i\varphi$  pour la seconde. Nous poussons l'analyse un peu plus loin et distinguons les cas "*i* a une croyance correcte à propos de  $\varphi$ " et "*i* a une simple croyance à propos de  $\varphi$ ", notées respectivement  $TBA_i\varphi$  et  $MBA_i\varphi$ . Autrement dit,  $MBA_i\varphi$  indique que *i* a une croyance à propos de  $\varphi$ , mais pas de connaissance.

Les modalités  $TBA_i\varphi$  et  $MBA_i\varphi$  sont non-standard, mais toutefois naturelles. Prenons comme exemple l'évolution des connaissances et des croyances dans un variant du test Sally-Ann [127, 18]. Supposons que Sally et Ann se trouvent dans une pièce, Ann tenant une bille. Un panier et une boîte sont posés dans cette pièce. Notons  $S$  pour Sally,  $A$  pour Ann, et  $b$  la proposition "la bille est dans le panier". Décrivons maintenant les états épistémiques de Sally et Ann au cours des événements qui suivent.

1. Ann met la bille dans le panier. Sally et Ann savent alors que la bille est dans le panier, elles ont donc une croyance à propos de  $b$  qui est correcte et n'est pas une simple croyance :

$$b \wedge TBA_A b \wedge \neg MBA_A b \wedge TBA_S b \wedge \neg MBA_S b.$$

2. Sally quitte la pièce. Sally continue de croire, à raison, que la bille est dans le panier, mais elle n'en est plus sûre :

$$b \wedge TBA_A b \wedge \neg MBA_A b \wedge TBA_S b \wedge MBA_S b.$$

3. Ann prend la bille et la met dans la boîte. En supposant que Sally croie toujours que la bille est dans le panier, sa croyance est maintenant incorrecte :

$$\neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge \neg TBA_S b \wedge MBA_S b.$$

4. Sally rentre de nouveau dans la pièce et regarde dans le panier. Elle sait maintenant que la bille n'est pas dans le panier :

$$\neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge TBA_S b \wedge \neg MBA_S b.$$

Nous pourrions étendre l'histoire de sorte à ce que Sally passe par tous les états épistémiques possibles. Par exemple, Sally pourrait initialement être en dehors de la pièce, n'ayant aucune idée de la position de la bille : alors  $TBA_S b$  et  $MBA_S b$  seraient tous les deux faux.

Au-delà de l'étude des situations épistémiques, nous montrons ici que les modalités  $TBA_i$  et  $MBA_i$  sont pleinement expressives, en ce sens que  $K_i \varphi$  et  $B_i \varphi$  peuvent être défini comme abréviations à partir de  $TBA_i \varphi$  et  $MBA_i \varphi$ . De plus, nous montrons que les huit situations épistémiques possibles par rapport à une proposition contingente  $\varphi$  peuvent être caractérisées par les huit combinaisons possibles des trois formules  $\varphi$ ,  $TBA_i \varphi$  et  $MBA_i \varphi$ . Enfin, nous étudions les réductions de modalités consécutives portant sur le même agent. Nous prenons les trois logiques standard de connaissances et croyances définies dans le Chapitre 2 et montrons que dans ces trois logiques,  $MBA_i TBA_i \varphi$  est équivalent à  $MBA_i \varphi$  et  $TBA_i TBA_i \varphi$  est équivalent à  $TBA_i \varphi \vee \neg MBA_i \varphi$ . De plus, nous montrons que pour deux de ces logiques les quatre combinaisons possibles peuvent être réduites à une profondeur modale de 1.

The simplicity of EL-O relies on two factors: the 'knowing whether' operator and the agents' limited reasoning power. In this chapter we consider a way to bring these aspects over to epistemic-doxastic logics, that is, logics that reason about agents' beliefs as well as their knowledge.

This chapter is mainly based on the following publication:

- Herzig, A. & Perrotin, E. (2021). *True belief and mere belief about a proposition and the classification of epistemic-doxastic situations*. *Filosofiska Notiser* (1-Special issue on modal logic) (pp.103-117).

## Contents

<b>7.1 'Belief-about' and epistemic-doxastic situations . . . . .</b>	<b>106</b>
7.1.1 From 'belief-that' to 'belief-about' . . . . .	106
7.1.2 A 'belief-about' fragment of the language $Fml_{EDL}$ . . . . .	107
7.1.3 First-order epistemic-doxastic situations . . . . .	107
7.1.4 Higher-order epistemic-doxastic situations . . . . .	108
7.1.5 From 'belief-about' to 'belief-that' . . . . .	109
<b>7.2 Reduction of 'about' modalities . . . . .</b>	<b>109</b>
7.2.1 Properties of EDL . . . . .	110
7.2.2 Properties of $EDL+BiBK$ . . . . .	111

7.2.3 Properties of $EDL+5(K_i)$ . . . . .	112
<b>7.3 Conclusion</b> . . . . .	<b>113</b>

In natural language one cannot say “I believe whether  $\varphi$ ”; we are therefore going to talk in this chapter about propositions of the form “ $i$  has knowledge about  $\varphi$ ” and “ $i$  has a belief about  $\varphi$ ”. Formally, we write  $Kif_i\varphi$  for the former and  $BA_i\varphi$  for the latter. We further analyse this and distinguish “ $i$  has a true belief about  $\varphi$ ” and “ $i$  has a mere belief about  $\varphi$ ”, respectively written  $TBA_i\varphi$  and  $MBA_i\varphi$ . We understand  $MBA_i\varphi$  as “ $i$  has a belief about  $\varphi$  but does not know whether  $\varphi$ ”.

The modalities  $TBA_i$  and  $MBA_i$  are non-standard, but nevertheless natural. To witness, consider the evolution of knowledge and belief in a variant of the famous Sally-Ann Test [127, 18]. Suppose Sally and Ann are in a room. Ann is holding a marble. There is a basket and a box in the room. Let  $S$  stand for Sally,  $A$  for Ann, and  $b$  for “the marble is in the basket”. We describe Ann’s and Sally’s epistemic situations after the following events took place.

1. Ann puts the marble in the basket. Sally then knows that the marble is in the basket, i.e., she has a belief about  $b$  that is both true and not a mere belief:

$$b \wedge TBA_A b \wedge \neg MBA_A b \wedge TBA_S b \wedge \neg MBA_S b.$$

2. Sally leaves the room. Sally continues to believe that the marble is in the basket, but she no longer knows that:

$$b \wedge TBA_A b \wedge \neg MBA_A b \wedge TBA_S b \wedge MBA_S b.$$

3. Ann transfers the marble to the box. Supposing that Sally still believes that the marble is in the basket she now has a false belief:

$$\neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge \neg TBA_S b \wedge MBA_S b.$$

4. Sally re-enters and looks inside the basket. Now she knows that the marble is not in the basket:

$$\neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge TBA_S b \wedge \neg MBA_S b.$$

We could extend the story in such a way that Sally goes through all possible epistemic situations. For example, Sally could initially be outside the room, not having any idea whether or not the marble is in the basket: then both  $TBA_S b$  and  $MBA_S b$  would be false.

The formal background of our investigation are the three epistemic-doxastic logics having both ‘knowledge that’ and ‘belief that’ modal operators described in Chapter 2 (Section 2.1.4):  $EDL$ ,  $EDL+BiBK$ , and  $EDL+5(K_i)$ . Beyond the study of epistemic situations we show that the modalities  $TBA_i$  and  $MBA_i$  are fully expressive, in the sense that both  $K_i\varphi$  and  $B_i\varphi$  can be defined as abbreviations from them. Furthermore, we show that the eight possible epistemic situations w.r.t. a contingent proposition  $\varphi$  can be characterised in terms of the eight possible combinations of the three formulas  $\varphi$ ,  $TBA_i\varphi$ , and  $MBA_i\varphi$ . Finally, we study reductions of consecutive modalities with the same agent. We show that



$KD5(B_i)$	the principles of modal logic KD5 for $B_i$
$S4(K_i)$	the principles of modal logic S4 for $K_i$
$KiB$	$K_i\varphi \rightarrow B_i\varphi$
$BiKB$	$B_i\varphi \rightarrow K_iB_i\varphi$
$BiBK$	$B_i\varphi \rightarrow B_iK_i\varphi$
$5(K_i)$	$\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$

Table 7.1: Principles of the three logics EDL (first four lines), EDL+ $BiBK$ , and EDL+ $5(K_i)$ .

for all three logics,  $MBA_iTBA_i\varphi$  is equivalent to  $MBA_i\varphi$  and  $TBA_iMBA_i\varphi$  is equivalent to  $TBA_i\varphi \vee \neg MBA_i\varphi$ . Moreover, we show that for the two strongest of the three logics all four possible combinations can be reduced to modal depth one.

The structure of this chapter is as follows. We first define several ‘knowledge about’ and ‘belief about’ modal operators and show that they have the same expressivity as standard ‘knowing that’ and ‘believing that’ operators in the considered logics (Section 7.1). We then give the reductions of consecutive modalities featuring the same agent (Section 7.2) and conclude (Section 7.3).

## 7.1 ‘Belief-about’ and epistemic-doxastic situations

In the language of ‘knowledge that’ and ‘belief that’ we can define several modal operators of the kind ‘having a belief about a proposition’ and ‘having knowledge about a proposition’. We show that the modalities ‘true belief about’ and ‘mere belief about’ play a particular role: they allow us to define in a combinatorial way all possible epistemic-doxastic situations about a contingent proposition  $\varphi$ . Moreover, we show how ‘knowledge that’ and ‘belief that’ can be defined from the modalities ‘true belief about’ and ‘mere belief about’.

We recall in Table 7.1 the three epistemic-doxastic logics EDL, EDL+ $BiBK$  and EDL+ $5(K_i)$  described in Chapter 2. In this section we only consider the logic EDL; we explore properties of our new operators in the two other logics in the next section.

### 7.1.1 From ‘belief-that’ to ‘belief-about’

Let us define the following modalities as abbreviations in the language  $Fml_{EDL}$ :

$$\begin{aligned}
 BA_i\varphi &= B_i\varphi \vee B_i\neg\varphi && \text{“}i\text{ has a belief about } \varphi\text{”} \\
 Kif_i\varphi &= K_i\varphi \vee K_i\neg\varphi && \text{“}i\text{ has knowledge about } \varphi\text{”} \\
 TBA_i\varphi &= (\varphi \wedge B_i\varphi) \vee (\neg\varphi \wedge B_i\neg\varphi) && \text{“}i\text{ has a true belief about } \varphi\text{”} \\
 MBA_i\varphi &= (B_i\varphi \wedge \neg K_i\varphi) \vee (B_i\neg\varphi \wedge \neg K_i\neg\varphi) && \text{“}i\text{ has a mere belief about } \varphi\text{”}
 \end{aligned}$$

Observe that by principles of propositional logic,  $TBA_i\varphi$  is equivalent to  $(\varphi \rightarrow B_i\varphi) \wedge (\neg\varphi \rightarrow B_i\neg\varphi)$ . Moreover,  $MBA_i\varphi$  is equivalent to  $(B_i\varphi \vee B_i\neg\varphi) \wedge \neg K_i\varphi \wedge \neg K_i\neg\varphi$  and hence to  $BA_i\varphi \wedge \neg Kif_i\varphi$ . (The first equivalence holds thanks to axioms  $KiB$  and  $D(B_i)$ .)

**Proposition 7.1.** *The following equivalences hold:*

$$\begin{aligned} BA_i\neg\varphi &\leftrightarrow BA_i\varphi, & TBA_i\neg\varphi &\leftrightarrow TBA_i\varphi, \\ Kif_i\neg\varphi &\leftrightarrow Kif_i\varphi, & MBA_i\neg\varphi &\leftrightarrow MBA_i\varphi. \end{aligned}$$

### 7.1.2 A ‘belief-about’ fragment of the language $Fml_{EDL}$

Let  $\mathcal{L}_{TBA,MBA}$  be the fragment of the language  $Fml_{EDL}$  where the only modal operators are  $TBA_i$  and  $MBA_i$ . Hence the grammar of  $\mathcal{L}_{TBA,MBA}$  is:

$$\mathcal{L}_{TBA,MBA} : \varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid TBA_i\varphi \mid MBA_i\varphi.$$

In the rest of the chapter we are going to investigate the properties of  $\mathcal{L}_{TBA,MBA}$ . In the rest of the present section we show that it has enough expressivity to account for all possible epistemic situations about a contingent proposition and to capture the ‘knowledge that’ and ‘belief that’ modalities. In the next section we investigate whether and how consecutive modalities can be reduced.

### 7.1.3 First-order epistemic-doxastic situations

Let  $\varphi$  be a contingent formula, i.e., a formula such that both  $\varphi$  and  $\neg\varphi$  are consistent. There are eight possible epistemic-doxastic situations that can be expressed in the traditional language of epistemic logic  $Fml_{EDL}$ , namely:

$\varphi \wedge K_i\varphi$	$\neg\varphi \wedge K_i\neg\varphi$
$\varphi \wedge B_i\varphi \wedge \neg K_i\varphi$	$\neg\varphi \wedge B_i\neg\varphi \wedge \neg K_i\neg\varphi$
$\varphi \wedge \neg B_i\varphi \wedge \neg B_i\neg\varphi$	$\neg\varphi \wedge \neg B_i\varphi \wedge \neg B_i\neg\varphi$
$\varphi \wedge B_i\neg\varphi$	$\neg\varphi \wedge B_i\varphi$

These distinctions can not only be expressed in  $Fml_{EDL}$ , but also in the fragment  $\mathcal{L}_{TBA,MBA}$ .

**Proposition 7.2.** *The following equivalences are theorems of EDL:*

$$\begin{aligned} \varphi \wedge K_i\varphi &\leftrightarrow \varphi \wedge TBA_i\varphi \wedge \neg MBA_i\varphi, \\ \neg\varphi \wedge K_i\neg\varphi &\leftrightarrow \neg\varphi \wedge TBA_i\neg\varphi \wedge \neg MBA_i\neg\varphi, \\ \varphi \wedge B_i\varphi \wedge \neg K_i\varphi &\leftrightarrow \varphi \wedge TBA_i\varphi \wedge MBA_i\varphi, \\ \neg\varphi \wedge B_i\neg\varphi \wedge \neg K_i\neg\varphi &\leftrightarrow \neg\varphi \wedge TBA_i\neg\varphi \wedge MBA_i\neg\varphi, \\ \varphi \wedge \neg B_i\varphi \wedge \neg B_i\neg\varphi &\leftrightarrow \varphi \wedge \neg TBA_i\varphi \wedge \neg MBA_i\varphi, \\ \neg\varphi \wedge \neg B_i\neg\varphi \wedge \neg B_i\varphi &\leftrightarrow \neg\varphi \wedge \neg TBA_i\neg\varphi \wedge \neg MBA_i\neg\varphi, \\ \varphi \wedge B_i\neg\varphi &\leftrightarrow \varphi \wedge \neg TBA_i\varphi \wedge MBA_i\varphi, \\ \neg\varphi \wedge B_i\varphi &\leftrightarrow \neg\varphi \wedge \neg TBA_i\neg\varphi \wedge MBA_i\neg\varphi. \end{aligned}$$

*Proof.* These equivalences can be proved with the  $KiB$  axiom together with principles of normal modal logics. The proof amounts to spelling out the definitions of  $TBA_i$  and  $MBA_i$  and applying the axioms  $KiB$ ,  $K(B_i)$ ,  $K(K_i)$  and the inference rules  $RN(B_i)$  and  $RN(K_i)$ . For example, the first equivalence can be

proved as follows:

$$\begin{aligned}
 \varphi \wedge TBA_i\varphi \wedge \neg MBA_i\varphi &\leftrightarrow \varphi \wedge B_i\varphi \wedge (B_i\varphi \rightarrow K_i\varphi) \\
 &\leftrightarrow \varphi \wedge B_i\varphi \wedge K_i\varphi \\
 &\leftrightarrow \varphi \wedge K_i\varphi \qquad \text{by } KiB
 \end{aligned}$$

□

Thanks to Proposition 7.2, the eight possible epistemic-doxastic situations can also be characterised in the fragment  $\mathcal{L}_{TBA,MBA}$ :

$\varphi \wedge TBA_i\varphi \wedge \neg MBA_i\varphi$	$\neg\varphi \wedge TBA_i\varphi \wedge \neg MBA_i\varphi$
$\varphi \wedge TBA_i\varphi \wedge MBA_i\varphi$	$\neg\varphi \wedge TBA_i\varphi \wedge MBA_i\varphi$
$\varphi \wedge \neg TBA_i\varphi \wedge \neg MBA_i\varphi$	$\neg\varphi \wedge \neg TBA_i\varphi \wedge \neg MBA_i\varphi$
$\varphi \wedge \neg TBA_i\varphi \wedge MBA_i\varphi$	$\neg\varphi \wedge \neg TBA_i\varphi \wedge MBA_i\varphi$

Hence we have characterised all possible epistemic-doxastic situations in terms of three independent components. This can be compared to the study of different normative positions in deontic logic as initiated by Kanger and Lindahl and studied more recently by Sergot [115].

#### 7.1.4 Higher-order epistemic-doxastic situations

We can generalise first-order epistemic-doxastic situations to higher orders. Let us demonstrate this by going through Sally's second-order epistemic-doxastic situations, i.e., her beliefs about Ann's beliefs.

1. Ann puts the marble in the basket. Sally then knows that Ann knows that the marble is in the basket, i.e., she has a belief about Ann's beliefs that is both true and not a mere belief:

$$\begin{aligned}
 &b \wedge TBA_Ab \wedge \neg MBA_Ab \wedge \\
 &TBA_Sb \wedge TBA_S TBA_Ab \wedge \neg MBA_S TBA_Ab \wedge \\
 &\neg MBA_Sb \wedge TBA_S MBA_Ab \wedge \neg MBA_S MBA_Ab.
 \end{aligned}$$

2. Sally leaves the room. Sally continues to believe that Ann sees the marble, but she no longer knows that:

$$\begin{aligned}
 &b \wedge TBA_Ab \wedge \neg MBA_Ab \wedge \\
 &TBA_Sb \wedge TBA_S TBA_Ab \wedge MBA_S TBA_Ab \wedge \\
 &MBA_Sb \wedge TBA_S MBA_Ab \wedge MBA_S MBA_Ab.
 \end{aligned}$$

3. Ann transfers the marble to the box. Sally's belief that Ann knows where the marble is remains true:

$$\begin{aligned}
 &\neg b \wedge TBA_Ab \wedge \neg MBA_Ab \wedge \\
 &\neg TBA_Sb \wedge TBA_S TBA_Ab \wedge MBA_S TBA_Ab \wedge \\
 &MBA_Sb \wedge TBA_S MBA_Ab \wedge MBA_S MBA_Ab.
 \end{aligned}$$

4. Sally re-enters and looks inside the basket. We consider several possibilities as to the evolution of her beliefs about Ann's beliefs:

- (a) Sally and Ann look into the basket together. Now Sally knows that Ann also knows that the marble isn't there:

$$\begin{aligned} & \neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge \\ & TBA_S b \wedge TBA_S TBA_A b \wedge \neg MBA_S TBA_A b \wedge \\ & \neg MBA_S b \wedge TBA_S MBA_A b \wedge \neg MBA_S MBA_A b. \end{aligned}$$

- (b) Sally looks into the basket without Ann. She may then believe that Ann still thinks that the marble is in the basket when it is not. Therefore Sally's beliefs about Ann's beliefs become untrue:

$$\begin{aligned} & \neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge \\ & TBA_S b \wedge \neg TBA_S TBA_A b \wedge MBA_S TBA_A b \wedge \\ & \neg MBA_S b \wedge \neg TBA_S MBA_A b \wedge MBA_S MBA_A b. \end{aligned}$$

- (c) Sally looks into the basket without Ann. She sees that the marble is not there, but believes that Ann is the one who took it out, and therefore that Ann still knows the location of the marble. As long as she has no confirmation, this remains a mere belief:

$$\begin{aligned} & \neg b \wedge TBA_A b \wedge \neg MBA_A b \wedge \\ & \neg TBA_S b \wedge TBA_S TBA_A b \wedge MBA_S TBA_A b \wedge \\ & MBA_S b \wedge TBA_S MBA_A b \wedge MBA_S MBA_A b. \end{aligned}$$

### 7.1.5 From 'belief-about' to 'belief-that'

The definition of 'about' modalities from 'that' modalities of Section 7.1 is straightforward. We now consider the other way around: expressing 'that' modalities using 'about' modalities. We already know that the formula  $K_i\varphi$  is equivalent to  $\varphi \wedge Kif_i\varphi$ . It is also known that the 'belief about' modality alone cannot express the belief-that modality [54]. We show now that the fragment  $\mathcal{L}_{TBA, MBA}$  is fully expressive: together, the two modalities of true belief  $TBA_i$  and of mere belief  $MBA_i$  are enough to express 'belief that' and 'knowledge that'.

**Proposition 7.3.** *The following equivalences are theorems of EDL:*

$$\begin{aligned} Kif_i\varphi & \leftrightarrow TBA_i\varphi \wedge \neg MBA_i\varphi, \\ K_i\varphi & \leftrightarrow \varphi \wedge TBA_i\varphi \wedge \neg MBA_i\varphi, \\ BA_i\varphi & \leftrightarrow TBA_i\varphi \vee MBA_i\varphi, \\ B_i\varphi & \leftrightarrow (\varphi \wedge TBA_i\varphi) \vee (\neg\varphi \wedge \neg TBA_i\varphi \wedge MBA_i\varphi). \end{aligned}$$

*Proof.* This follows from Proposition 7.2. □

## 7.2 Reduction of 'about' modalities

In this section we explore the interplay between the different modalities, as governed in particular by principles of introspection. We begin by listing some equivalences of the base logic EDL, then we investigate some more properties of its two extensions  $EDL+BiBK$  and  $EDL+5(K_i)$ .

### 7.2.1 Properties of EDL

Our first group of equivalences is about traditional operators followed by  $TBA_i$  and  $MBA_i$  and their negations.

**Proposition 7.4.** *The following equivalences are theorems of EDL:*

$$\begin{aligned} B_i TBA_i \varphi &\leftrightarrow BA_i \varphi, & K_i TBA_i \varphi &\leftrightarrow K_i f_i \varphi, \\ B_i \neg TBA_i \varphi &\leftrightarrow \neg BA_i \varphi, & K_i \neg TBA_i \varphi &\leftrightarrow \neg BA_i \varphi, \\ BA_i TBA_i \varphi &\leftrightarrow \top. \end{aligned}$$

*Proof.* We first prove the equivalences of the first column.

$$\begin{aligned} B_i TBA_i \varphi &\leftrightarrow B_i(\neg\varphi \vee B_i\varphi) \wedge B_i(\varphi \vee B_i\neg\varphi) \\ &\leftrightarrow (B_i\neg\varphi \vee B_i\varphi) \wedge (B_i\varphi \vee B_i\neg\varphi) && \text{by KD45}(B_i) \\ &\leftrightarrow BA_i \varphi \\ B_i \neg TBA_i \varphi &\leftrightarrow B_i\neg((\varphi \wedge B_i\varphi) \vee (\neg\varphi \wedge B_i\neg\varphi)) \\ &\leftrightarrow B_i(\neg\varphi \vee \neg B_i\varphi) \wedge B_i(\varphi \vee \neg B_i\neg\varphi) \\ &\leftrightarrow (B_i\neg\varphi \vee \neg B_i\varphi) \wedge (B_i\varphi \vee \neg B_i\neg\varphi) && \text{by KD45}(B_i) \\ &\leftrightarrow \neg B_i\varphi \wedge \neg B_i\neg\varphi && \text{by KD}(B_i) \\ &\leftrightarrow \neg BA_i \varphi \end{aligned}$$

The last equivalence of the first column follows immediately from the two above results.

We now move on to the second column.

$$\begin{aligned} K_i TBA_i \varphi &\leftrightarrow K_i((\varphi \rightarrow B_i\varphi) \wedge (\neg\varphi \rightarrow B_i\neg\varphi)) \\ &\leftrightarrow K_i(\varphi \rightarrow B_i\varphi) \wedge K_i(\neg\varphi \rightarrow B_i\neg\varphi) \\ &\leftrightarrow (K_i\neg\varphi \vee B_i\varphi) \wedge (K_i\varphi \vee B_i\neg\varphi) && \text{by Remark 2.1} \\ &\leftrightarrow K_i\neg\varphi \vee K_i\varphi \\ &\leftrightarrow K_i f_i \varphi \\ K_i \neg TBA_i \varphi &\leftrightarrow K_i\neg((\varphi \wedge B_i\varphi) \vee (\neg\varphi \wedge B_i\neg\varphi)) \\ &\leftrightarrow K_i(\neg\varphi \vee \neg B_i\varphi) \wedge K_i(\varphi \vee \neg B_i\neg\varphi) \\ &\leftrightarrow (K_i\neg\varphi \vee \neg B_i\varphi) \wedge (K_i\varphi \vee \neg B_i\neg\varphi) && \text{by Remark 2.1} \\ &\leftrightarrow \neg B_i\varphi \wedge \neg B_i\neg\varphi \\ &\leftrightarrow \neg BA_i \varphi \end{aligned}$$

This ends the proof.  $\square$

The above equivalences allow us to reduce consecutive modalities  $TBA_i TBA_i$  and  $MBA_i TBA_i$ .

**Proposition 7.5.** *The following equivalences hold in EDL:*

$$\begin{aligned} TBA_i TBA_i \varphi &\leftrightarrow TBA_i \varphi \vee \neg MBA_i \varphi, \\ MBA_i TBA_i \varphi &\leftrightarrow MBA_i \varphi. \end{aligned}$$

*Proof.* For the first equivalence:

$$\begin{aligned}
 TBA_i TBA_i \varphi &\leftrightarrow (TBA_i \varphi \wedge B_i TBA_i \varphi) \vee (\neg TBA_i \varphi \wedge B_i \neg TBA_i \varphi) \\
 &\leftrightarrow (TBA_i \varphi \wedge BA_i \varphi) \vee (\neg TBA_i \varphi \wedge \neg BA_i \varphi) && \text{by Proposition 7.4} \\
 &\leftrightarrow TBA_i \varphi \vee (\neg TBA_i \varphi \wedge \neg MBA_i \varphi) && \text{by Proposition 7.3} \\
 &\leftrightarrow TBA_i \varphi \vee \neg MBA_i \varphi
 \end{aligned}$$

For the second equivalence:

$$\begin{aligned}
 MBA_i TBA_i \varphi &\leftrightarrow BA_i TBA_i \varphi \wedge \neg K_i TBA_i \varphi \wedge \neg K_i \neg TBA_i \varphi \\
 &\leftrightarrow \top \wedge \neg K_i \varphi \wedge BA_i \varphi && \text{by Proposition 7.4} \\
 &\leftrightarrow MBA_i \varphi
 \end{aligned}$$

This ends the proof.  $\square$

We conjecture that the logic EDL is not strong enough to allow us to reduce consecutive modalities  $TBA_i MBA_i$  and  $MBA_i MBA_i$ . We do not establish this formally and show instead that such reductions exist for the two extensions of EDL,  $EDL+BiBK$  and  $EDL+5(K_i)$ .

## 7.2.2 Properties of $EDL+BiBK$

We once again begin by investigating the interactions between traditional operators and  $TBA_i$  and  $MBA_i$ .

**Proposition 7.6.** *The following equivalences hold in  $EDL+BiBK$ :*

$$\begin{aligned}
 B_i MBA_i \varphi &\leftrightarrow \perp, & K_i MBA_i \varphi &\leftrightarrow \perp, \\
 B_i \neg MBA_i \varphi &\leftrightarrow \top, & K_i \neg MBA_i \varphi &\leftrightarrow \neg MBA_i \varphi.
 \end{aligned}$$

*Proof.* We start with the left column.

$$\begin{aligned}
 B_i MBA_i \varphi &\leftrightarrow B_i((B_i \varphi \vee B_i \neg \varphi) \wedge \neg K_i \varphi \wedge \neg K_i \neg \varphi) \\
 &\leftrightarrow B_i(B_i \varphi \vee B_i \neg \varphi) \wedge B_i \neg K_i \varphi \wedge B_i \neg K_i \neg \varphi \\
 &\leftrightarrow (B_i \varphi \vee B_i \neg \varphi) \wedge B_i \neg K_i \varphi \wedge B_i \neg K_i \neg \varphi \\
 &\leftrightarrow (B_i K_i \varphi \vee B_i K_i \neg \varphi) \wedge B_i \neg K_i \varphi \wedge B_i \neg K_i \neg \varphi && \text{by } BiBK \\
 &\leftrightarrow \perp \\
 B_i \neg MBA_i \varphi &\leftrightarrow B_i((B_i \varphi \rightarrow K_i \varphi) \wedge (B_i \neg \varphi \rightarrow K_i \neg \varphi)) \\
 &\leftrightarrow B_i(B_i \varphi \rightarrow K_i \varphi) \wedge B_i(B_i \neg \varphi \rightarrow K_i \neg \varphi) \\
 &\leftrightarrow (B_i \varphi \rightarrow B_i K_i \varphi) \wedge (B_i \neg \varphi \rightarrow B_i K_i \neg \varphi) && \text{by KD45}(B_i) \\
 &\leftrightarrow \top && \text{by } BiBK
 \end{aligned}$$

It remains to prove the right column.

$$\begin{aligned}
 K_i MBA_i \varphi &\rightarrow B_i MBA_i \varphi \\
 &\rightarrow \perp \\
 K_i \neg MBA_i \varphi &\leftrightarrow K_i((B_i \varphi \rightarrow K_i \varphi) \wedge (B_i \neg \varphi \rightarrow K_i \neg \varphi)) \\
 &\leftrightarrow (B_i \varphi \rightarrow K_i \varphi) \wedge (B_i \neg \varphi \rightarrow K_i \neg \varphi) && \text{by Remark 2.1} \\
 &\leftrightarrow \neg MBA_i \varphi
 \end{aligned}$$

$\square$

Together with Proposition 7.5, the next result establishes that in  $\text{EDL}+BiBK$  all combinations of  $TBA_i$  and  $MBA_i$  can be reduced:

**Proposition 7.7.** *The following equivalences hold in  $\text{EDL}+BiBK$ :*

$$\begin{aligned} TBA_iMBA_i\varphi &\leftrightarrow \neg MBA_i\varphi, \\ MBA_iMBA_i\varphi &\leftrightarrow MBA_i\varphi. \end{aligned}$$

*Proof.* The proof makes use of Proposition 7.6:

$$\begin{aligned} TBA_iMBA_i\varphi &\leftrightarrow (MBA_i\varphi \wedge B_iMBA_i\varphi) \vee (\neg MBA_i\varphi \wedge B_i\neg MBA_i\varphi) \\ &\leftrightarrow (MBA_i\varphi \wedge \perp) \vee (\neg MBA_i\varphi \wedge \top) \\ &\leftrightarrow \neg MBA_i\varphi \\ MBA_iMBA_i\varphi &\leftrightarrow (B_iMBA_i\varphi \wedge \neg K_iMBA_i\varphi) \vee (B_i\neg MBA_i\varphi \wedge \neg K_i\neg MBA_i\varphi) \\ &\leftrightarrow (\perp \wedge \neg \perp) \vee (\top \wedge \neg \neg MBA_i\varphi) \\ &\leftrightarrow MBA_i\varphi \end{aligned}$$

□

### 7.2.3 Properties of $\text{EDL}+5(K_i)$

We show once again that all combinations of  $TBA_i$  and  $MBA_i$  can be reduced, now considering  $\text{EDL}+5(K_i)$ . In this logic the reductions are quite straightforward.

**Proposition 7.8.** *The following equivalences hold in  $\text{EDL}+5(K_i)$ :*

$$\begin{aligned} Kif_iMBA_i\varphi &\leftrightarrow \top, \\ MBA_iMBA_i\varphi &\leftrightarrow \perp, \\ TBA_iMBA_i\varphi &\leftrightarrow \top. \end{aligned}$$

*Proof.* The introspective principles tell us that  $MBA_i\varphi \rightarrow K_iMBA_i\varphi$  and that  $\neg MBA_i\varphi \rightarrow K_i\neg MBA_i\varphi$ , hence the first equivalence. From there, we can show:

$$\begin{aligned} MBA_iMBA_i\varphi &\leftrightarrow BA_iMBA_i\varphi \wedge \neg Kif_iMBA_i\varphi \\ &\leftrightarrow BA_iMBA_i\varphi \wedge \perp \\ &\leftrightarrow \perp \end{aligned}$$

Following the same reasoning as for the first equivalence, we also have that  $MBA_i\varphi \rightarrow B_iMBA_i\varphi$  and  $\neg MBA_i\varphi \rightarrow B_i\neg MBA_i\varphi$ . Therefore:

$$\begin{aligned} TBA_iMBA_i\varphi &\leftrightarrow (MBA_i\varphi \wedge B_iMBA_i\varphi) \vee (\neg MBA_i\varphi \wedge B_i\neg MBA_i\varphi) \\ &\leftrightarrow MBA_i\varphi \vee \neg MBA_i\varphi \\ &\leftrightarrow \top \end{aligned}$$

□

These last two reductions follow the intuition that in  $\text{EDL}+5(K_i)$  the agents can always tell whether or not their beliefs are based on knowledge, whereas in  $\text{EDL}+BiBK$  agents do not consider the possibility that their beliefs are mere beliefs.

### 7.3 Conclusion

We have studied two modalities of ‘belief about’ in the framework of three epistemic-doxastic logics whose base logic combines KD45 for the ‘belief that’ modality and S4 for the ‘knowledge that’ modality. Our ‘true belief about’ and ‘mere belief about’ modalities can express in a combinatorial way all eight possible epistemic situations. They are also expressive enough to capture the ‘belief that’ and ‘knowledge that’ modalities.

For all three logics, an axiomatization of the theorems in the language  $\mathcal{L}_{TBA, MBA}$  can be obtained in a very simple manner, namely by taking the traditional axiomatization of Table 7.1 and substituting  $K_i$  and  $B_i$  by their definitions in terms of  $TBA_i$  and  $MBA_i$  of Section 7.1.5. This is straightforward, but we do not find this very informative because the resulting axioms are complicated, particularly as the modal operators  $TBA_i$  and  $MBA_i$  neither satisfy the monotony axioms  $TBA_i(\varphi \wedge \psi) \rightarrow (TBA_i\varphi \wedge TBA_i\psi)$  and  $MBA_i(\varphi \wedge \psi) \rightarrow (MBA_i\varphi \wedge MBA_i\psi)$ , nor the conjunction axioms  $(TBA_i\varphi \wedge TBA_i\psi) \rightarrow TBA_i(\varphi \wedge \psi)$  and  $(MBA_i\varphi \wedge MBA_i\psi) \rightarrow MBA_i(\varphi \wedge \psi)$ .

One of the perspectives for future work is the definition of lightweight fragments of our language, as done for  $B_i$  by Muise et al. [107] and as we have done for  $K_i$  with EL-O. In the same spirit we can define *epistemic-doxastic atoms* as sequences of modalities  $TBA_i$  and  $MBA_i$  that are followed by a propositional variable: they are of the form

$$M_1 \cdots M_dp$$

where  $p \in Prop$  and where  $M_k$  is either  $TBA_i$  or  $MBA_i$ , for some  $i$ . We therefore expect lightweight epistemic-doxastic logic to be able to account for the evolution of belief and knowledge in applications where autonomous agents have to be equipped with a theory of mind in order to reason about other agents.

Just as the switch from ‘knowing that’ to ‘knowing whether’ makes an extension to ‘knowing what’ quite natural, we can also imagine the operators of true and mere belief being applied to constants: our beliefs are not limited to propositions and one might have a true belief or a mere belief about what the letter says or who the murderer is. Considering a set of constants  $Cst$ , the language of true and mere beliefs then becomes:

$$\mathcal{L}_{TBA, MBA}^C : \varphi ::= p \mid TBA_ic \mid MBA_ic \mid \neg\varphi \mid \varphi \wedge \varphi \mid TBA_i\varphi \mid MBA_i\varphi.$$

where  $c \in Cst$ .

Here is an example: When reading a murder mystery, a reader  $r$  starts with no belief or knowledge who the culprit might be. This can be expressed as  $\neg MBA_{r,c} \wedge \neg TBA_{r,c}$ , where  $c$  represents the identity of the culprit. After a few clues are given, the author’s goal is usually to mislead the reader: that is, the author wishes for  $MBA_{r,c} \wedge \neg TBA_{r,c}$  to be the case. Hence the reader should have a theory ( $MBA_{r,c}$ ), but that theory should be false ( $\neg TBA_{r,c}$ ). The reader, however, wishes to outwit the author and figure out who the real culprit is before their theory is confirmed by the final reveal: they wish for  $MBA_{r,c} \wedge TBA_{r,c}$  to be true. By the end of the book, all is explained, and the reader knows who the culprit really was:  $\neg MBA_{r,c} \wedge TBA_{r,c}$  is true.



## Chapter 8

# On knowledge and belief bases

### Sur les bases de connaissances et de croyances

Les représentations de la connaissance et de la croyance que nous avons étudiées dans cette thèse sont *symboliques* : les états sont des ensembles de formules à partir desquelles la valeur de vérité de n'importe quelle formule peut être directement calculée. Les états EL-O peuvent être vus comme des bases de connaissances contenant les connaissances de tous les agents à cet état ainsi que les vérités ontiques. Cette manière de représenter les connaissances des agents par des ensembles de formules n'est pas nouvelle, et présente des avantages en termes de vérification de modèles [120, 21, 22]. Ce type d'approche est encore plus prévalente lorsqu'on considère les croyances plutôt que les connaissances [82, 68, 113, 20, 3, 114], quoique principalement étudiée dans des contextes avec un seul agent, où le problème est de maintenir la cohérence d'une base de croyances malgré l'apparition de nouvelles informations.

Dans la lignée de ces approches avec bases de croyances, la Logique des Attitudes Doxastiques (LDA) de Lorini, introduite dans [92] et développée en tant que "famille de logiques LDA" dans [94], présente un système relativement général pour raisonner sur les croyances de plusieurs agents, en se fondant sur un mélange de bases de croyances représentant les *croyances explicites* des agents et une sémantique de mondes possibles (appelés *alternatives doxastiques*) permettant le calcul de *croyances implicites*. Contrairement aux modèles de Kripke standard, les relations d'accessibilité dans les modèles LDA sont calculées à partir des croyances explicites des agents, ce qui rend les modèles plus compacts : si les modèles de Kripke sont d'autant plus grands que l'incertitude est élevée, l'inverse est vrai pour les bases de croyances. De plus, les propriétés des croyances telles que l'introspection sont garanties par des propriétés des relations d'accessibilité. Il est donc possible de modéliser différents types de croyance ainsi que des notions de croyance de groupe, telles que la croyance distribuée [72], simplement en modifiant la définition de ces relations d'accessibilité.

Dans ce chapitre nous tirons profit de la modularité de LDA et montrons que les logiques présentées dans cette thèse ainsi que plusieurs approches similaires de la littérature peuvent être traduites dans LDA en ajustant les définitions

des alternatives doxastiques ainsi que du socle commun (l'ensemble de tous les mondes possibles dans un modèle donné). Nous donnons d'abord une vue d'ensemble des logiques LDA, puis montrons comment y traduire EL-O dans la Section 8.2 en passant de bases de connaissances à des 'bases de visions' que nous définissons. Les modèles résultants sont similaires aux structures de connaissances de [21], que nous traduisons aussi dans LDA dans la Section 8.3 afin de clarifier la relation entre les deux. Dans la Section 8.4 nous couvrons les 'bases de connaissances épistémiques propres' (PEKBs) de [84, 107], avant de revenir aux croyances et d'étudier un fragment de la logique des croyances correctes et simples croyances du Chapitre 7 dans la Section 8.5.

We finish this thesis by providing a basis for comparison between our work and a number of related approaches through embeddings into the framework of Lorini's Logic of Doxastic Attitudes.

The contents of this chapter are as of yet unpublished. References are made to the following publication:

- Herzig, A., Lorini, E., Perrotin, E., Romero, F. & Schwarzentruher, F. (2020). *A logic of explicit and implicit distributed belief*. 24th European Conference on Artificial Intelligence (ECAI 2020) (pp.753-760).

## Contents

<b>8.1</b>	<b>An overview of LDA</b>	<b>116</b>
8.1.1	Basic language and semantics	116
8.1.2	Other kinds of belief and knowledge	117
8.1.3	Group belief and knowledge	117
<b>8.2</b>	<b>EL-O: Vision bases</b>	<b>118</b>
8.2.1	'Vision' bases: language and semantics	118
8.2.2	EL-O to vision bases	121
8.2.3	Vision bases to EL-O	122
<b>8.3</b>	<b>Relation to knowledge structures</b>	<b>123</b>
<b>8.4</b>	<b>Relation to PEKBs</b>	<b>124</b>
<b>8.5</b>	<b>True and mere beliefs</b>	<b>124</b>
8.5.1	TM bases	124
8.5.2	S4.2 contexts and semantics	126
8.5.3	Equivalence of the semantics	127
<b>8.6</b>	<b>Discussion and conclusion</b>	<b>129</b>

The representations of knowledge and belief that we have studied in this thesis are *symbolic*: states are sets of formulas from which truths are directly computed. We can see EL-O states as knowledge bases containing the knowledge of all agents at that state as well as ontic truths. This approach of representing agents' minds through sets of formulas is not new, as it presents advantages when it comes to model checking [120, 21, 22]. It is even more prevalent in doxastic logics [82, 68, 113, 20, 3, 114], though mostly studied in single-agent settings where the focus is maintaining consistency of a belief base in the face of new information.

Following this ‘belief base’ approach to belief representation, Lorini’s Logic of Doxastic Attitudes (LDA), introduced in [92] and further developed as a “family of LDA logics” in [94], provides quite a general framework for reasoning about multiple agents’ beliefs, through a mixture of belief bases representing the agents’ *explicit beliefs* and possible worlds (so-called *doxastic alternatives*) semantics allowing for computation of *implicit beliefs*. Contrary to standard Kripke models, accessibility relations in LDA models are derived from agents’ explicit beliefs, which makes the framework fairly compact: while standard Kripke models are bigger the less information agents have, the opposite is the case for belief bases. Moreover, properties of belief such as introspection are ensured through properties on accessibility relations. Hence changing the computation of these relations allows us to model different kinds of belief as well as group belief notions such as distributed belief [72].

In this chapter we take advantage of the modularity of LDA and show that the frameworks presented in this thesis as well as several related approaches can all be embedded into LDA through slight adjustments in the definitions of doxastic alternatives and common ground (the set of possible worlds in a given model). We start by giving an overview of the LDA framework. We then show in Section 8.2 how to embed EL-O into this framework by going from belief bases to what we call ‘vision bases’. This leads to a framework very similar to the knowledge structures of [21], which we also embed into the LDA framework in section 8.3 in order to get a better idea of the relation between the two. In Section 8.4 we go over the proper epistemic knowledge bases of [84, 107], before returning to beliefs and considering the logic of true and mere beliefs of Chapter 7 in Section 8.5. We conclude in Section 8.6.

## 8.1 An overview of LDA

In this approach, each agent has a non-deductively closed belief base consisting of their explicit beliefs. An implicit belief is a statement that can be inferred from these explicit beliefs. In [92, 94], the idea of using belief bases as a semantics for multi-agent epistemic logic which was put forth. It was further developed in a series of papers with the aim of capturing multi-agent belief dynamics [94, 96], higher-order epistemic reasoning [93] and graded belief [97], of elucidating the connection between distributed belief and belief merging [72] and between belief bases and propositional awareness [98].

### 8.1.1 Basic language and semantics

We start by recalling the basic language and semantics of LDA. We assume as before that  $Prop$  is a countably infinite set of propositions and  $Agt = \{1, \dots, n\}$  is a finite set of agents. The languages  $\mathcal{L}_0(Prop, Agt)$  and  $\mathcal{L}_1(Prop, Agt)$  are defined as follows:

$$\begin{aligned}\mathcal{L}_0(Prop, Agt) \ni \nu &::= p \mid \neg\nu \mid \nu \wedge \nu \mid \Delta_i\nu \\ \mathcal{L}_1(Prop, Agt) \ni \varphi &::= \nu \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_i\varphi\end{aligned}$$

where  $p \in Prop$  and  $i \in Agt$ .  $\Delta_i\nu$  reads: “agent  $i$  *explicitly* believes that  $\nu$ ”, while  $\Box_i\varphi$  reads: “agent  $i$  *implicitly* believes that  $\varphi$ ”.

A *multi-agent belief base* is a tuple  $B = (B_1, \dots, B_n, V)$  where for every  $i \in \text{Agt}$ ,  $B_i \subseteq \mathcal{L}_0$  is agent  $i$ 's belief base and  $V \subseteq \text{Prop}$  is the actual state of the world. We call  $\mathbf{B}$  the set of all multi-agent belief bases. Formulas of  $\mathcal{L}_0(\text{Prop}, \text{Agt})$  are interpreted in multi-agent belief bases as follows:

$$\begin{aligned} B \models p & & \text{iff } p \in V; \\ B \models \Delta_i \nu & & \text{iff } \nu \in B_i; \end{aligned}$$

and as usual for boolean operators.

The interpretation of  $\mathcal{L}_1(\text{Prop}, \text{Agt})$  formulas involves quantification over a given set of belief bases, through the notion of *doxastic alternatives*: if  $B, B' \in \mathbf{B}$  are two multi-agent belief bases,  $B\mathcal{R}_i B'$  if for all  $\nu \in B_i$ ,  $B' \models \nu$ . A *context* is any set of multi-agent belief bases.

Formulas of the full language  $\mathcal{L}_1(\text{Prop}, \text{Agt})$  are then interpreted as follows: if  $B$  is a belief base and  $Cxt$  is a context,

$$\begin{aligned} (B, Cxt) \models \nu & & \text{iff } B \models \nu; \\ (B, Cxt) \models \Box_i \varphi & & \text{iff } (B', Cxt) \models \varphi \text{ for all } B' \in Cxt \text{ s.t. } B\mathcal{R}_i B'; \end{aligned}$$

the interpretation of boolean operators being standard.

### 8.1.2 Other kinds of belief and knowledge

Different properties of implicit belief can be represented by adjusting the definition of the relations  $R_i$ . For example, to add positive introspection of implicit belief we can add  $B_i \subseteq B'_i$  to the condition for  $B\mathcal{R}_i B'$ . For negative introspection we add the condition  $B'_i \subseteq B_i$ . The D axiom corresponds to individual belief bases being consistent.

If we wish to consider standard knowledge ('knowing that'), every  $B_i$  must contain only information that is true in  $B$ : that is, any belief base  $B$  must be such that if  $\nu \in B_i$  then  $B \models \nu$ . Then  $\Delta_i \nu$  can be read as " $i$  explicitly knows that  $\nu$ " and  $\Box_i \varphi$  as " $i$  implicitly knows that  $\varphi$ ".

It is actually more convenient to consider knowing *whether*: it suffices then to change " $B' \models \nu$ " to " $B$  and  $B'$  agree on  $\nu$ " in the condition for  $B\mathcal{R}_i B'$ , and truth of agents' knowledge need not be verified as it is now built in.  $\Delta_i \nu$  then reads " $i$  explicitly knows whether  $\nu$ " and  $\Box_i \varphi$  reads " $i$  implicitly knows whether  $\varphi$ ". The interpretation of the latter operator also changes to:

$$\begin{aligned} (B, Cxt) \models \Box_i \varphi & & \text{iff } (B, Cxt) \text{ and } (B', Cxt) \text{ agree on } \alpha \\ & & \text{for all } B' \in Cxt \text{ s.t. } B\mathcal{R}_i B'. \end{aligned}$$

### 8.1.3 Group belief and knowledge

The operator of implicit belief  $\Box_i$  can be generalized to a 'group belief' operator  $\Box_G$ , where  $G \subseteq \text{Agt}$  is a group of agents and  $\Box_i$  becomes identified with  $\Box_{\{i\}}$ . Three kinds of group belief and knowledge are generally studied: shared, distributed, and common belief and knowledge; see [72] for a study of  $\Box_G$  as an operator of distributed belief. When considering knowledge, standard common

knowledge (which follows the induction principle) can quite naturally be defined as follows:

$$(B, Cxt) \models \Box_G \varphi \quad \text{iff } (B', Cxt) \models \varphi \text{ for all } B' \in Cxt \text{ s.t. } BR_G B'$$

for ‘knowing that’ and

$$(B, Cxt) \models \Box_G \varphi \quad \text{iff } (B, Cxt) \text{ and } (B', Cxt) \text{ agree on } \alpha \\ \text{for all } B' \in Cxt \text{ s.t. } BR_G B'$$

for ‘knowing whether’, where  $\mathcal{R}_G$  is in both cases the transitive closure of  $\bigcup_{i \in G} \mathcal{R}_i$ .

## 8.2 EL-O: Vision bases

In this section we give an embedding of EL-O into the framework of LDA.

### 8.2.1 ‘Vision’ bases: language and semantics

As EL-O is about ‘knowing whether’, when embedding it into the LDA framework it is convenient to take the ‘knowing whether’ version of the language of LDA as a target language. We therefore restrict the language as follows: the language of explicit knowledge becomes

$$\mathcal{L}_0^{\text{Vis}} \ni \nu ::= p \mid \Delta_i \nu$$

where  $p \in Prop$ , while the language of implicit knowledge is restricted to  $\mathcal{L}_{\text{bool}}(ATM)$ . Here  $\Delta_i \nu$  reads “agent  $i$  explicitly observes whether or not  $\nu$  is true”, while  $S_i \alpha$  expresses implicit observation and  $JS \alpha$  expresses implicit joint observation (that is,  $S_i$  is the equivalent of  $\Box_i$  and  $JS$  that of  $\Box_{Agt}$  in the language of the previous section). One can view explicit observation as representing what the agent ‘physically’ sees, while implicit observation is what the agent deduces from what they see.

We now give the semantics of what we call vision bases. These semantics follow those given in the previous section for logics of knowing whether and commonly knowing whether, but are adapted to the restricted language. A *multi-agent vision base* is a tuple  $B = (B_1, \dots, B_n, V)$  where for every  $i \in Agt$ ,  $B_i \subseteq \mathcal{L}_0^{\text{Vis}}$  is agent  $i$ ’s vision base and  $V \subseteq Prop$  is the actual state of the world. The set of all multi-agent vision bases is noted  $\mathbf{B}^{\text{Vis}}$ . The semantics for  $\mathcal{L}_0^{\text{Vis}}$  is as follows:

$$\begin{array}{ll} B \models p & \text{iff } p \in V; \\ B \models \Delta_i \nu & \text{iff } \nu \in B_i. \end{array}$$

The definition of epistemic alternatives for vision bases is that given in Section 8.1.2 for ‘knowing whether’: if  $B, B' \in \mathbf{B}^{\text{Vis}}$ , then  $BR_i^{\text{Vis}} B'$  if  $B_i = B'_i$  and  $B$  and  $B'$  agree on all  $\nu \in B_i$ . The relation  $\mathcal{R}_i^{\text{Vis}}$  is clearly an equivalence relation for all  $i$ .

Once again, a context is any set of multi-agent vision bases. We call  $\mathbf{Cxt}^{\text{Vis}}$  the set of all contexts. If  $B$  is a vision base and  $Cxt$  is a context, truth of a

$Vis_1$	$S_i S_i \alpha$
$Vis_2$	$JS JS \alpha$
$Vis_3$	$JS S_i S_i \alpha$
$Vis_4$	$JS \alpha \rightarrow S_i \alpha$
$Vis_5$	$JS \alpha \rightarrow JS S_i \alpha$

Table 8.1: The EL-O axioms

formula of  $\mathcal{L}_{\text{bool}}(ATM)$  in  $(B, Cxt)$  is defined as follows:

$$\begin{aligned}
(B, Cxt) \models p & \quad \text{iff } B \models p; \\
(B, Cxt) \models S_i \alpha & \quad \text{iff } (B, Cxt) \text{ and } (B', Cxt) \text{ agree on } \alpha \\
& \quad \text{for all } (B', Cxt) \text{ s.t. } BR_i^{\text{vis}} B'; \\
(B, Cxt) \models JS \alpha & \quad \text{iff } (B, Cxt) \text{ and } (B', Cxt) \text{ agree on } \alpha \\
& \quad \text{for all } B' \in Cxt;
\end{aligned}$$

and as usual for boolean operators. Notice that the truth condition for  $(B, Cxt) \models JS \alpha$  differs from the one we have given in Section 8.1.3 for the ‘common knowledge whether’ operator. This is because, as discussed in previous chapters, joint vision in EL-O is a nonstandard form of common knowledge which does not obey the induction principle.

We now consider in what way the logic of vision bases relates to EL-O. A first interesting result pertains to the axiomatization of EL-O.

**Proposition 8.1.** *The EL-O axioms, recalled in Table 8.1, are valid in the semantics of vision bases.*

*Proof.* Let  $(B, Cxt)$  be a model,  $i$  be an agent and  $\alpha$  an atom.

$Vis_1$ : Let  $B' \in Cxt$  be such that  $BR_i^{\text{vis}} B'$ . We want to show that  $B$  and  $B'$  agree on  $S_i \alpha$ . Suppose there exists  $B''$  such that  $B' \mathcal{R}_i^{\text{vis}} B''$  and  $(B', Cxt)$  and  $(B'', Cxt)$  disagree on  $\alpha$ . Then  $BR_i^{\text{vis}} B''$  and  $(B, Cxt)$  and  $(B'', Cxt)$  disagree on  $\alpha$ . Likewise, if  $(B, Cxt)$  and  $(B'', Cxt)$  disagree on  $\alpha$  for some  $B''$  such that  $BR_i^{\text{vis}} B''$ , then  $B' \mathcal{R}_i^{\text{vis}} B''$  and  $(B', Cxt)$  and  $(B'', Cxt)$  disagree on  $\alpha$ . Hence  $B$  and  $B'$  agree on  $S_i \alpha$ .

$Vis_2$ : By the semantics of  $JS$ , all  $B'$  such that  $B' \in Cxt$  must agree on  $JS \alpha$ .

$Vis_3$ : By  $Vis_1$ , all  $B'$  such that  $B' \in Cxt$  verify  $(B', Cxt) \models S_i S_i \alpha$ .

$Vis_4, Vis_5$ : suppose that all  $B'$  such that  $B' \in Cxt$  agree on  $\alpha$ . Then in particular for any  $B', B'' \in Cxt$  such that  $B' \mathcal{R}_i^{\text{vis}} B''$ ,  $(B', Cxt)$  and  $(B'', Cxt)$  agree on  $\alpha$ . Therefore all  $B'$  such that  $B' \in Cxt$  verify  $(B', Cxt) \models S_i \alpha$ .  $\square$

The restricted nature of the EL-O language has two consequences when considering a translation to vision bases: the first, which we have already addressed, is to restrict what agents may explicitly see, that is, what kinds of formulas may be found in the agents’ vision bases. The second is to restrict what agents may commonly know, and how much reasoning power is available to them. This is reflected through the choice of possible contexts for vision bases.

We now consider the class  $\mathbf{Cxt}^{\text{EL-O}} \subseteq \mathbf{Cxt}^{\text{vis}}$  of contexts of the shape  $Cxt(P^+, P^-, N^+, N^-)$  where  $P^+, P^- \subseteq Prop$ ,  $N^+, N^- \subseteq \mathcal{L}_0 \setminus Prop$ ,  $P^+ \cap P^- =$

$\emptyset$ ,  $N^+ \cap N^- = \emptyset$ , and

$$\begin{aligned} \text{Cxt}(P^+, P^-, N^+, N^-) = \{ & (B_1, \dots, B_n, V) \in \mathbf{B}^{\text{Vis}} : \\ & P^+ \subseteq V, P^- \cap V = \emptyset, \\ & \forall \Delta_i \nu \in N^+, \nu \in B_i \\ & \text{and } \forall \Delta_i \nu \in N^-, \nu \notin B_i \}. \end{aligned}$$

In order to establish further properties of these vision bases w.r.t. to EL-O, we first define translations between elements of  $\mathcal{L}_0^{\text{Vis}}$  and EL-O atoms in which there are no occurrences of the operator  $JS$ : if  $\alpha$  is such an atom, we call  $\alpha^E$  the proposition  $\alpha$  in which all operators of the form  $S_i$  is replaced by the corresponding  $\Delta_i i$ , and if  $\nu \in \mathcal{L}_0^{\text{Vis}}$ , we call  $\nu^I$  the proposition  $\nu$  in which all operators of the form  $\Delta_i$  are replaced by the corresponding  $S_i$ . Formally:

$$\begin{aligned} p^E &= p^I = p; \\ (S_i \alpha)^E &= \Delta_i(\alpha^E); \\ (\Delta_i \nu)^I &= S_i(\nu^I). \end{aligned}$$

**Proposition 8.2.** *For all  $\alpha \in \text{ATM}$ , for all  $\text{cxt} \in \mathbf{Cxt}^{\text{EL-O}}$  and all  $B = (B_1, \dots, B_n, v) \in \text{Cxt}$ ,  $(B, \text{Cxt}) \models S_i \alpha$  iff one of the following holds:*

1.  $S_i \alpha \in I\text{-ATM}$
2.  $(B, \text{Cxt}) \models JS \alpha$
3. *there is no  $JS$  in  $\alpha$  and  $\alpha^E \in B_i$ .*

*Proof.* The right-to-left direction follows from the previous proposition and the definition of epistemic alternatives. We show the left-to-right direction by induction on  $\alpha$ . Suppose that neither of the three conditions holds. Construct  $B'$  in the following way:

- If  $\alpha = p$  for some  $p \in \text{Prop}$ ,  $B' = (B_1, \dots, B_n, V')$  where  $V' = V \cup \{p\}$  if  $p \notin V$ ,  $V' = V \setminus \{p\}$  if  $p \in V$ ;
- If  $\alpha = S_j \nu'$  for some  $j$  and  $\beta$ ,  $B' = (B_1, \dots, B'_j, \dots, B_n, V)$  where  $B'_j = B_j \cup \{\nu'\}$  if  $\nu' \notin B_j$ ,  $B'_j = B_j \setminus \{\nu'\}$  if  $\beta \in B_j$ .

We can then show in both cases that  $B' \in \text{Cxt}$ ,  $B\mathcal{R}_i^{\text{Vis}} B'$  and  $B$  and  $B'$  disagree on  $\alpha$  (using the induction hypothesis in the second case). Hence  $(B, \text{Cxt}) \not\models S_i \alpha$ .  $\square$

**Proposition 8.3.** *For any  $\alpha \in \text{ATM}$ ,  $\text{Cxt} = \text{Cxt}(P^+, P^-, N^+, N^-) \in \mathbf{Cxt}^{\text{Vis}}$  and  $B \in \text{Cxt}$ ,  $(B, \text{Cxt}) \models JS \alpha$  iff  $JS \alpha \in I\text{-ATM}$  or there exists  $\nu \in P^+ \cup P^- \cup N^+ \cup N^-$  such that  $JS \nu^I \Rightarrow JS \alpha$ .*

*Proof.* We show this by induction on the length of  $\alpha$ . It is clear that for any  $p \in \text{Prop}$ ,  $(B, \text{Cxt}) \models JS p$  iff  $p \in P^+ \cup P^-$ .

Suppose that  $(B, \text{Cxt}) \models JS S_i \alpha$  with  $JS S_i \alpha \notin I\text{-ATM}$ . If  $(B, \text{Cxt}) \models JS \alpha$  then by the induction hypothesis there exists  $\nu \in P^+ \cup P^- \cup N^+ \cup N^-$  such that  $JS \nu^I \Rightarrow JS \alpha$ , and then  $JS \nu^I \Rightarrow JS S_i \alpha$ . Otherwise, we show that  $\alpha^E \in P^+ \cup P^- \cup N^+ \cup N^-$ :

- If  $(B, Cxt) \models S_i \alpha$  then as  $(B, Cxt) \models JS S_i \alpha$  and  $(B, Cxt) \not\models JS \alpha$ , we have that  $(B', Cxt) \models S_i \alpha$  and  $(B', Cxt) \not\models JS \alpha$  for all  $B' \in Cxt$ . Hence for all  $B' \in Cxt$  we are in the third case of Proposition 8.2, that is,  $\alpha^E \in B'_i$ . Then necessarily  $\alpha \in N^+$ .
- If  $(B, Cxt) \not\models S_i \alpha$  then  $(B', Cxt) \not\models S_i \alpha$  for all  $B' \in Cxt$ , hence  $\alpha^E \notin B'_i$  for all  $B' \in Cxt$ . Then necessarily  $\alpha \in N^-$ .

Finally, suppose that there exists  $\nu \in P^+ \cup P^- \cup N^+ \cup N^-$  such that  $JS \nu^I \Rightarrow JS \alpha$ . We show that  $(B, Cxt) \models JS \nu^I$ , which will entail  $(B, Cxt) \models JS \alpha$  by Proposition 8.1. It is clear that  $(B, Cxt) \models JS \alpha$  if  $\nu$  is in  $P^+$ ,  $P^-$  or  $N^+$ . If  $\nu$  is in  $N^-$  then  $\nu$  is of the form  $\Delta_i \mu$  and with  $\mu \notin B'_i$  for all  $B' \in Cxt$ . There are then two cases:

- If  $(B', Cxt) \models S_i \mu^I$  for some  $B' \in Cxt$  then by Proposition 8.2, either  $S_i \mu^I \in I-ATM$  or  $(B', Cxt) \models JS \mu^I$ . In both cases we conclude that  $(B, Cxt) \models JS S_i \mu^I$ .
- Otherwise  $(B', Cxt) \not\models S_i \mu^I$  for all  $B' \in Cxt$ , hence  $(B, Cxt) \models JS S_i \mu^I$  by the semantics of the  $JS$  operator.

Hence in all cases  $(B, Cxt) \models JS \nu^I$ , which concludes the proof.  $\square$

### 8.2.2 EL-O to vision bases

For any EL-O state, we can construct a corresponding vision model which has exactly the same validities in the following manner. If  $s \subseteq ATM$  is a state, we define  $B^s = (B_1^s, \dots, B_n^s, V^s)$  such that for all  $\nu \in \mathcal{L}_0$ ,  $\nu \in B_i^s$  iff  $S_i \nu^I \in s$ , and for all  $p \in Prop$ ,  $p \in V^s$  iff  $p \in s$ . We also define  $Cxt^s = Cxt(P^+, P^-, N^+, N^-)$  where

- $P^+ = \{p : JS p \in s \text{ and } p \in s\}$ ;
- $P^- = \{p : JS p \in s \text{ and } p \notin s\}$ ;
- $N^+ = \{(S_i \alpha)^E : JS S_i \alpha \in s \text{ and } S_i \alpha \in s\}$ ;
- $N^- = \{(S_i \alpha)^E : JS S_i \alpha \in s \text{ and } S_i \alpha \notin s\}$ .

**Lemma 8.1.** *For all  $\alpha$  in  $\mathcal{L}_{\text{bool}}(ATM)$ , for all  $s \in ATM$ , if  $(B^s, Cxt^s) \models \alpha$  then  $s \models \alpha$ .*

*Proof.* We show this for every form of  $\alpha$ . First,  $s \models p$  iff  $p \in s$  iff  $p \in V^s$  iff  $(B^s, Cxt^s) \models p$ .

If  $(B^s, Cxt^s) \models JS \alpha$ , we know by Proposition 8.3 that either  $JS \alpha \in I-ATM$  or there exists  $\nu \in P^+ \cup P^- \cup N^+ \cup N^-$  such that  $JS \nu^I \Rightarrow JS \alpha$ . In the first case, clearly  $s \models JS \alpha$ ; in the second case,  $JS \nu^I \in s$  by definition of  $Cxt^s$ , hence  $s \models JS \alpha$ .

Finally, suppose that  $(B^s, Cxt^s) \models S_i \alpha$ . By Proposition 8.2, we know that either  $S_i \alpha \in I-ATM$ ,  $(B, Cxt) \models JS \alpha$ , or  $\alpha^E \in B_i^s$ . In the first case, clearly  $s \models S_i \alpha$ . In the second case, we have already shown that  $s \models JS \alpha$ , hence  $s \models S_i \alpha$ . In the third case,  $S_i \alpha \in s$  by definition of  $B_i^s$ , hence  $s \models S_i \alpha$ .  $\square$

**Lemma 8.2.** *For all  $\alpha$  in  $ATM$ , for all  $s \in ATM$ , if  $s \models \alpha$  then  $(B^s, Cxt^s) \models \alpha$ .*



*Proof.* We prove this for every form of  $\alpha$ . First,  $s \models p$  iff  $p \in s$  iff  $p \in V^s$  iff  $(B^s, Cxt^s) \models p$ .

Suppose now that  $s \models JS \alpha$ . Then  $JS \alpha \in s^{\Rightarrow} \cup I-ATM$ . If  $JS \alpha \in I-ATM$  then  $(B^s, Cxt^s) \models JS \alpha$  by Proposition 8.1. If  $JS \alpha \in s^{\Rightarrow}$ , then there exists  $\beta$  such that  $JS \beta \in s$  and  $JS \beta \Rightarrow JS \alpha$ . Then by definition of  $Cxt^s$  we have that  $\beta^E \in P^+ \cup P^- \cup N^+ \cup N^-$  and therefore  $(B^s, Cxt^s) \models JS \alpha$  by Proposition 8.3.

Finally, suppose that  $s \models S_i \alpha$ . Then  $S_i \alpha \in s^{\Rightarrow} \cup I-ATM$ . If  $S_i \alpha \in I-ATM$  then  $(B^s, Cxt^s) \models S_i \alpha$  by Proposition 8.1. If  $S_i \alpha \in s^{\Rightarrow}$ , then either  $S_i \alpha \in s$  or  $JS \beta \Rightarrow S_i \alpha$  for some  $JS \beta \in s$ . In the first case by definition of  $B^s$  we have  $\alpha^E \in B_i$ . In the second case we have already shown that  $(B^s, Cxt^s) \models JS \beta$ . By Proposition 8.1 this gives us that  $(B^s, Cxt^s) \models S_i \alpha$ .  $\square$

The following result can be shown by induction on  $\varphi$ :

**Proposition 8.4.** *For all  $\varphi$  in  $\mathcal{L}_{\text{bool}}(ATM)$ , for all  $s \in ATM$ ,  $s \models \varphi$  iff  $(B^s, Cxt^s) \models \varphi$ .*

*Proof.* By a straightforward induction on  $\varphi$ , using the two lemmas.  $\square$

### 8.2.3 Vision bases to EL-O

We now show the other direction: for any pointed vision model with a context in  $\mathbf{Cxt}^{\text{EL-O}}$ , we can construct a corresponding EL-O state with the same validities.

We first define for any  $Cxt = Cxt(P^+, P^-, N^+, N^-) \in \mathbf{Cxt}^{\text{EL-O}}$  and  $B = (B_1, \dots, B_n, V) \in Cxt$  the state  $s^{(B, Cxt)} \subseteq ATM$  as follows:

$$s^{(B, Cxt)} = V \cup \{S_i \nu^I \mid i \in \text{Agt}, \nu \in B_i\} \cup \{JS \nu^I \mid \nu \in P^+ \cup P^- \cup N^+ \cup N^-\}$$

**Lemma 8.3.** *For any  $\alpha \in ATM$ , for any  $Cxt \in \mathbf{Cxt}^{\text{EL-O}}$  and any  $B \in Cxt$ , if  $(B, Cxt) \models \alpha$  then  $s^{(B, Cxt)} \models \alpha$ .*

*Proof.* By induction on  $\alpha$ , where  $B = (B_1, \dots, B_n, V)$ :

If  $(B, Cxt) \models p$  then  $p \in V$  and therefore  $p \in s^{(B, Cxt)}$ .

If  $(B, Cxt) \models JS \beta$  then by Proposition 8.3, either  $JS \beta \in I-ATM$ , in which case we immediately have that  $s^{(B, Cxt)} \models JS \beta$ , or there exists  $\nu \in P^+ \cup P^- \cup N^+ \cup N^-$  such that  $JS \nu^I \Rightarrow JS \beta$ . Because  $\nu \in P^+ \cup P^- \cup N^+ \cup N^-$ , we have  $JS \nu^I \in s^{(B, Cxt)}$ . Therefore  $s^{(B, Cxt)} \models JS \nu^I$  and  $s^{(B, Cxt)} \models JS \beta$ .

If  $(B, Cxt) \models S_i \beta$  then either  $\beta \in B_i$  or  $(B, Cxt) \models JS \beta$  or  $S_i \beta \in I-ATM$ . In the first case  $S_i \beta \in s^{(B, Cxt)}$ . In the second case, by the same reasoning as above, we have that  $s^{(B, Cxt)} \models JS \beta$ , and therefore  $s^{(B, Cxt)} \models S_i \beta$ . In the third case we immediately have that  $s^{(B, Cxt)} \models S_i \beta$ .  $\square$

**Lemma 8.4.** *For any  $\alpha \in ATM$ , for any  $Cxt \in \mathbf{Cxt}^{\text{EL-O}}$  and any  $B \in Cxt$ , if  $s^{(B, Cxt)} \models \alpha$  then  $(B, Cxt) \models \alpha$ .*

*Proof.* By induction on  $\alpha$ , where  $B = (B_1, \dots, B_n, V)$ :

If  $s^{(B, Cxt)} \models p$  then  $p \in s$  and therefore  $p \in V$ .

If  $s^{(B, Cxt)} \models JS \beta$  then either  $JS \beta \in I-ATM$  or there exists  $\gamma$  such that  $JS \gamma \in s^{(B, Cxt)}$  and  $JS \gamma \Rightarrow JS \beta$ . In the first case  $(B, Cxt) \models JS \beta$ . In the second case by definition of  $s^{(B, Cxt)}$  we have  $(B, Cxt) \models JS \gamma$  and therefore  $(B, Cxt) \models JS \beta$ .

If  $s^{(B,Cxt)} \models S_i \beta$  then either  $S_i \beta \in I\text{-ATM}$  or  $S_i \beta \in s$  or there exists  $\gamma$  such that  $JS \gamma \in s^{(B,Cxt)}$  and  $JS \gamma \Rightarrow S_i \beta$ . In the first case  $(B, Cxt) \models S_i \beta$ . In the second case  $\beta^E \in B_i$  and therefore  $(B, Cxt) \models S_i \beta$ . In the third case we have by the same reasoning as above that  $(B, Cxt) \models JS \gamma$ , and therefore  $(B, Cxt) \models S_i \beta$ .  $\square$

**Proposition 8.5.** *For any  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$ , for any  $Cxt \in \mathbf{Cxt}^{\text{EL-O}}$  and any  $B = (B_1, \dots, B_n, V) \in Cxt$ ,  $s^{(B,Cxt)} \models \varphi$  iff  $(B, Cxt) \models \varphi$ .*

*Proof.* By induction on  $\varphi$  using the two previous lemmas.  $\square$

We now conclude for the relation of EL-O to vision base models:

**Proposition 8.6.** *For any  $\varphi \in \mathcal{L}_{\text{bool}}(\text{ATM})$ ,  $\vdash_{\text{EL-O}} \varphi$  iff  $\models_{\mathbf{Cxt}^{\text{EL-O}}} \varphi$  (where  $\models_{\mathbf{Cxt}^{\text{EL-O}}} \varphi$  is defined as: for all  $Cxt \in \mathbf{Cxt}^{\text{EL-O}}$ , for all  $B \in Cxt$ ,  $(B, Cxt) \models \varphi$ ).*

### 8.3 Relation to knowledge structures

Knowledge structures are defined in [21] as follows:

$$\mathcal{F} = (\text{Voc}, \theta, O_1, \dots, O_n)$$

where  $\text{Voc}$  is a finite subset of  $\text{Prop}$ ,  $\theta$  is a boolean formula over  $\text{Voc}$ , and for each agent  $i$ ,  $O_i$  is a subset of  $\text{Voc}$  representing that agent's *observable variables*, that is, the variables that  $i$  observes. We omit the dynamic components of the original presentation, though it would be interesting to examine parallels between dynamic versions of vision base models and knowledge structures as well. In order to either avoid ambiguity with vision base notations, or to show parallels between the two we adapt some of the notations of the original presentation.

States are assignments over (*i.e.* subsets of)  $\text{Voc}$  satisfying  $\theta$ . While a more general relation between states is defined in order to cover common knowledge for groups of agents, the indistinguishability relation for any agent  $i$  can be redefined as follows:  $s \mathcal{R}_i^{\text{KS}} s'$  iff  $s \cap O_i = s' \cap O_i$ . The semantics over these worlds are standard, with no distinction between explicit and implicit observation.

The purpose of the rest of this section is to show that knowledge structures can also be embedded into vision base models. This turns out to be quite straightforward, as many of the elements of both representations are very similar. As before, we consider a restriction on contexts of  $\mathbf{Cxt}^{\text{Vis}}$ , which we call  $\mathbf{Cxt}^{\text{KS}}$ , and show that given this restriction, there is an exact correspondence between vision models and knowledge structures.

For any knowledge structure  $\mathcal{F} = (\text{Voc}, \theta, O_1, \dots, O_n)$ , we define a corresponding context

$$\text{Cxt}(\mathcal{F}) = \{(O_1, \dots, O_n, V) : V \subseteq \text{Voc}, V \models \theta\}.$$

We call  $\mathbf{Cxt}^{\text{KS}}$  the set of contexts of the form  $\text{Cxt}(\mathcal{F})$ .

**Lemma 8.5.** *Let  $\mathcal{F} = (\text{Voc}, \theta, O_1, \dots, O_n)$  be a knowledge structure, and consider the context  $\text{Cxt}(\mathcal{F})$ . Then for any states  $s$  and  $s'$  of the knowledge structure,  $s \mathcal{R}_i^{\text{KS}} s'$  iff  $(O_1, \dots, O_n, s) \mathcal{R}_i^{\text{Vis}}(O_1, \dots, O_n, s')$ .*

*Proof.* The proof is straightforward, following the definitions.  $\square$

**Proposition 8.7.** *Let  $\mathcal{F} = (\text{Voc}, \theta, O_1, \dots, O_n)$  be a knowledge structure, and let  $s$  be a state of  $\mathcal{F}$ . Then for any formula  $\varphi$ ,  $((O_1, \dots, O_n, s), \text{Cxt}(\mathcal{F})) \models \varphi$  iff  $(\mathcal{F}, s) \models \varphi$ .*

*Proof.* The proof is straightforward, by induction on  $\varphi$ .  $\square$

## 8.4 Relation to PEKBs

We come back to the proper epistemic knowledge bases (PEKBs) mentioned in Section 6.4.1. Just like EL-O, PEKBs are fairly straightforward to translate into the framework of belief bases because they mainly rely on a restriction of the language. Moreover, common knowledge is not considered.

Recall that PEKBs are sets, or (equivalently) conjunctions, of restricted modal literals, defined as follows:

$$\lambda ::= p \mid \neg\lambda \mid K_i\lambda$$

This translates into the language of belief bases as:

$$\begin{aligned} \mathcal{L}_0^{\text{PEKB}} \ni \nu &::= p \mid \neg\nu \mid \Delta_i\nu \\ \mathcal{L}_1^{\text{PEKB}} \ni \varphi &::= r \mid \varphi \wedge \varphi \end{aligned}$$

Semantics are then standard, and the restricted reasoning capabilities of agents are preserved by considering the set of all multi-agent bases as a context. To reflect the logic of [84], we take the most general semantics for belief bases, corresponding to the modal logic K; if we wish to follow [107], we take the semantics corresponding to KD45 models, as described in Section 8.1.2.

## 8.5 True and mere beliefs

We now adapt belief base models to the S4.2 version of the logic of true and mere beliefs of Chapter 7 with the following restricted language:

$$\begin{aligned} \text{ATM}^{\text{TM}} \ni \alpha &::= p \mid TBA_i\alpha \mid MBA_i\alpha \\ \mathcal{L}_1^{\text{TM}} \ni \varphi &::= \alpha \mid \neg\varphi \mid \varphi \wedge \varphi \end{aligned}$$

We further restrict the language so that for any  $\alpha$ , there are no two consecutive operators relating to the same agent, that is, no  $O_iO'_i$  where  $i$  is an agent and  $O, O' \in \{TBA, MBA\}$ .

### 8.5.1 TM bases

Each agent now has two bases: one for true beliefs, and one for mere beliefs. The language of explicit beliefs is as follows:

$$\mathcal{L}_0^{\text{TM}} \ni \nu ::= p \mid \theta_i\nu \mid \mu_i\nu$$

where  $\theta_i\nu$  reads “ $i$  has a true explicit belief about  $\nu$ ” and  $\mu_i\nu$  reads “ $i$  has a mere explicit belief about  $\nu$ ”.

A *multi-agent true and mere belief base*, or TM base for short, is a tuple  $B = (T_1, M_1, \dots, T_n, M_n, V)$  where for every  $i \in \text{Agt}$  ( $\text{Agt} = \{1, \dots, n\}$ ),  $T_i, M_i \subseteq \mathcal{L}_0^{\text{TM}}$  are agent  $i$ 's true and mere belief bases respectively and  $V \subseteq \text{Prop}$  is the actual (ontic) state of the world. We furthermore require that for any agent  $i$ ,  $T_i$  and  $M_i$  contain no formula starting with  $\theta_i$  or  $\mu_i$ , that is, that agents have no explicit beliefs about their own explicit beliefs. We call  $\mathbf{B}^{\text{TM}}$  the set of all TM bases.

The semantics for formulas of  $\mathcal{L}_0^{\text{TM}}$  is as follows:

$$\begin{array}{ll} B \models p & \text{iff } p \in V; \\ B \models \theta_i\nu & \text{iff } \nu \in T_i; \\ B \models \mu_i\nu & \text{iff } \nu \in M_i. \end{array}$$

We now need to define epistemic alternatives for TM bases. Consider  $B, B' \in \mathbf{B}^{\text{TM}}$ . Then  $B\mathcal{R}_i^{\text{TM}}B'$  iff the following hold:

1.  $T_i \cup M_i = T'_i \cup M'_i$ ;
2.  $T_i \setminus M_i \subseteq T'_i \setminus M'_i$ ;
3.  $B$  and  $B'$  agree on  $T_i \cap T'_i$  and on  $(M_i \setminus T_i) \cap (M'_i \setminus T'_i)$ ;
4.  $B$  and  $B'$  disagree on  $T_i\Delta T'_i$ ;

where  $T_i\Delta T'_i = (T_i \setminus T'_i) \cup (T'_i \setminus T_i)$ . Intuitively, an agent in state  $B$  believes that another state  $B'$  is possible if:

- The agent holds beliefs about the exact same objects (this corresponds to the interaction axioms  $B_i\varphi \rightarrow K_iB_i\varphi$  and  $\neg B_i\varphi \rightarrow K_i\neg B_i\varphi$ );
- The agent still knows in  $B'$  what they know in  $B$  (this corresponds to positive introspection for knowledge);
- The agent's explicit beliefs are the same in  $B$  and  $B'$ : if the agent's belief about a formula  $\nu$  is either true in both  $B$  and  $B'$  or false in both  $B$  and  $B'$  then  $B$  and  $B'$  should agree on  $\varphi$ , and if the agent's belief about  $\varphi$  is true in  $B$  and false in  $B'$  or vice versa then  $B$  and  $B'$  should disagree on  $\nu$ .

**Proposition 8.8.** *For all  $i$ ,  $\mathcal{R}_i^{\text{TM}}$  is a reflexive and transitive relation.*

*Proof.* Reflexivity is immediate. For transitivity, consider  $B, B'$  and  $B''$  such that  $B\mathcal{R}_i^{\text{TM}}B'$  and  $B'\mathcal{R}_i^{\text{TM}}B''$ . Then:

1.  $T_i \cup M_i = T'_i \cup M'_i = T''_i \cup M''_i$ .
2.  $T_i \setminus M_i \subseteq T'_i \setminus M'_i \subseteq T''_i \setminus M''_i$ .
3. Suppose that  $\nu \in T_i \cap T''_i$ . If  $\nu \in T'_i$ , then  $\nu \in T_i \cap T'_i$ , hence  $B$  and  $B'$  agree on  $\nu$ , and  $\nu \in T'_i \cap T''_i$ , hence  $B'$  and  $B''$  agree on  $\nu$ . Therefore  $B$  and  $B''$  agree on  $\nu$ . If  $\nu \notin T'_i$ , then  $\nu \in T_i\Delta T'_i$ , hence  $B$  and  $B'$  disagree on  $\nu$ , and  $\nu \in T'_i\Delta T''_i$ , hence  $B'$  and  $B''$  disagree on  $\nu$ . Therefore  $B$  and  $B''$  agree on  $\nu$ .

Suppose now that  $\nu \in (M_i \setminus T_i) \cap (M_i'' \setminus T_i'')$ . If  $\nu \in T_i'$ , then  $\nu \in T_i \Delta T_i'$ , hence  $B$  and  $B'$  disagree on  $\nu$ , and  $\nu \in T_i' \Delta T_i''$ , hence  $B'$  and  $B''$  disagree on  $\nu$ . Therefore  $B$  and  $B''$  agree on  $\nu$ . If  $\nu \notin T_i'$ , then  $\nu \in (M_i \setminus T_i) \cap (M_i' \setminus T_i')$ , hence  $B$  and  $B'$  agree on  $\nu$ , and  $\nu \in (M_i' \setminus T_i') \cap (M_i'' \setminus T_i'')$ , hence  $B'$  and  $B''$  agree on  $\nu$ . Therefore  $B$  and  $B''$  agree on  $\nu$ .

4. Suppose  $\nu \in T_i \Delta T_i''$ . Consider for example  $\nu \in T_i \setminus T_i''$ ; the other case is symmetrical. If  $\nu \in T_i'$ , then  $\nu \in T_i \cap T_i'$ , so  $B$  and  $B'$  agree on  $\nu$ , and  $\nu \in T_i' \Delta T_i''$ , so  $B'$  and  $B''$  disagree on  $\nu$ . Therefore  $B$  and  $B''$  disagree on  $\nu$ . If  $\nu \notin T_i'$ , then  $\nu \in T_i \Delta T_i'$ , so  $B$  and  $B'$  disagree on  $\nu$ , and  $\nu \in (M_i' \setminus T_i') \cap (M_i'' \setminus T_i'')$ , so  $B'$  and  $B''$  agree on  $\nu$ . Therefore  $B$  and  $B''$  disagree on  $\nu$ .

Hence  $B \mathcal{R}_i^{\text{TM}} B''$ . □

### 8.5.2 S4.2 contexts and semantics

In order to reflect properties of S4.2 knowledge, we consider the set of contexts  $\mathbf{Cxt}^{\text{TM}}$  which has the following property: for any  $Cxt \in \mathbf{Cxt}^{\text{TM}}$  and any  $i \in \text{Agt}$ , if  $B = (T_1, M_1, \dots, T_n, M_n, V) \in Cxt$  then there exists a  $B^i = (T_1^i, M_1^i, \dots, T_n^i, M_n^i, V^i) \in Cxt$  such that  $T_i^i = T_i \cup M_i$ ,  $M_i^i = \emptyset$ , and  $B$  and  $B^i$  agree on  $T_i$  and disagree on  $M_i \setminus T_i$ .

Though we only require any such state to be in the considered contexts, we show how to construct one given a TM base  $(T_1, M_1, \dots, T_n, M_n, V)$ . For this we take:

- $T_j^i = \{\nu : \theta_j \nu \in T_i \text{ and } \nu \in T_j\} \cup \{\nu : \theta_j \nu \in M_i \setminus T_i \text{ and } \nu \notin T_j\}$  for  $j \neq i$ ;
- $M_j^i = \{\nu : \mu_j \nu \in T_i \text{ and } \nu \in T_j\} \cup \{\nu : \mu_j \nu \in M_i \setminus T_i \text{ and } \nu \notin T_j\}$  for  $j \neq i$ ;
- $V^i = (T_i \cap V) \cup ((M_i \setminus T_i) \setminus V)$ .

As  $T_i$  and  $M_i$  contain no atoms starting with  $\theta_i$  or  $\mu_i$ , it is easily shown that this TM base agrees with  $B$  on  $T_i$  and disagrees with  $B$  on  $M_i \setminus T_i$ .

*Proof.* This is shown by induction on the structure of  $\nu \in T_i \cup M_i$ . If  $p \in T_i$  for some  $p \in \text{Prop}$ , then  $p \in V^i$  iff  $p \in V$ , and therefore  $B$  and  $B^i$  agree on  $p$ . If  $p \in M_i \setminus T_i$ , then  $p \in V^i$  iff  $p \notin V$ , so  $B$  and  $B^i$  disagree on  $p$ . The argument is similar for  $\nu = \theta_i \nu'$  and  $\nu = \mu_i \nu'$ . Note that this only works for the restricted language in which there are no boolean connectors in formulas of  $\mathcal{L}_0$ . □

Recall that, as described in Chapter 2 (Section 2.1.4), S4.2 models have a final cluster for every  $R_i$  and that an agent believes  $\varphi$  at a state iff  $\varphi$  is true in all states of the corresponding final cluster. We now show that any context satisfying this property yields an S4.2 model, and that in particular the  $B^i$  described above are the elements of the final cluster for the relation  $\mathcal{R}_i^{\text{TM}}$ .

**Proposition 8.9.** *If  $Cxt \in \mathbf{Cxt}^{\text{TM}}$  then for all  $i$ ,  $\mathcal{R}_i^{\text{TM}}$  is a reflexive, transitive and confluent relation.*

*Proof.* Reflexivity and transitivity has been shown in the general case. Confluence is shown using the property required for contexts in  $\mathbf{Cxt}^{\mathbf{TM}}$ . Consider a TM base  $B = (T_1, M_1 \dots, T_n, M_n, V) \in \mathit{Cxt}$  and a corresponding  $B^i = (T_1^i, M_1^i, \dots, T_n^i, M_n^i, V^i) \in \mathit{Cxt}$ . For any  $B' = (T'_1, M'_1 \dots, T'_n, M'_n, V') \in \mathit{Cxt}$ , we show that if  $B\mathcal{R}_i^{\mathbf{TM}}B'$  then  $B'\mathcal{R}_i^{\mathbf{TM}}B^i$ .

1.  $T_i \cup M_i = T'_i \cup M'_i$ .
2.  $T'_i \setminus M'_i \subseteq T'_i \subseteq T_i \cup M_i$ .
3. If  $\nu \in T'_i \cap (T_i \cup M_i)$  then  $\nu \in T_i \cap T'_i$  hence  $B$  and  $B'$  agree on  $\nu$ , and  $\nu \in T_i$ , hence  $B$  and  $B^i$  agree on  $\nu$ . Therefore  $B'$  and  $B^i$  agree on  $\nu$ .
4. If  $\nu \in T'_i \Delta (T_i \cup M_i)$  then as  $T'_i \subseteq (T_i \cup M_i)$  it is the case that  $\nu \notin T'_i$ . If  $\nu \in T_i$  then  $\nu \in T_i \Delta T'_i$  therefore  $B$  and  $B'$  disagree on  $\nu$ , and  $B$  and  $B^i$  agree on  $\nu$ . Hence  $B'$  and  $B^i$  disagree on  $\nu$ . If  $\nu \notin T_i$  then  $\nu \in (M_i \setminus T_i) \cap (M'_i \setminus T'_i)$  and therefore  $B$  and  $B'$  agree on  $\nu$  and  $B$  and  $B^i$  disagree on  $\nu$ . Hence  $B'$  and  $B^i$  disagree on  $\nu$ .

□

Let us examine the shape of such a model for a single agent. It is a lattice with a final cluster. Each ‘level’ corresponds to a given number of propositions known, and knowledge increases when going towards the final cluster in the lattice. Clusters correspond to identical knowledge. The bottom of the lattice represents no knowledge, while the top of the lattice represents the world in which everything is known, which is the final cluster of believed states.

The language of implicit beliefs that we consider is  $\mathcal{L}_1^{\mathbf{TM}}$ , described at the beginning of this section. In order to define the semantics, the simplest way is to start from the standard epistemic language  $\mathit{Fml}_{\mathbf{EL}}$ . The semantics for this language are standard:

$$\begin{aligned} (B, \mathit{Cxt}) \models p & \quad \text{iff } B \models p; \\ (B, \mathit{Cxt}) \models K_i \varphi & \quad \text{iff } (B', \mathit{Cxt}) \models \varphi \text{ for all } B' \in \mathit{Cxt} \text{ s.t. } B\mathcal{R}_i^{\mathbf{TM}}B'; \end{aligned}$$

and as usual for boolean operators.

We can then define the operators  $B_i$ ,  $BA_i$ ,  $TBA_i$ ,  $MBA_i$  and their semantics in the standard manner:

$$\begin{aligned} B_i \varphi & \leftrightarrow \neg K_i \neg K_i \varphi; \\ BA_i \varphi & \leftrightarrow B_i \varphi \vee B_i \neg \varphi \\ TBA_i \varphi & \leftrightarrow (B_i \varphi \wedge \varphi) \vee (B_i \neg \varphi \wedge \neg \varphi); \\ MBA_i \varphi & \leftrightarrow (B_i \varphi \vee B_i \neg \varphi) \wedge \neg K_i \varphi \wedge \neg K_i \neg \varphi. \end{aligned}$$

### 8.5.3 Equivalence of the semantics

We now consider the canonical context  $\mathbf{B}^{\mathbf{TM}}$ , that is, the set of all possible TM bases. We show that given a set of atoms  $s \subseteq \mathit{ATM}^{\mathbf{TM}}$  and its direct translation into a TM base  $B^s$ , for any formula  $\varphi \in \mathcal{L}_1^{\mathbf{TM}}$ ,  $s \models^{\text{CPC}} \varphi$  iff  $(B^s, \mathbf{B}^{\mathbf{TM}}) \models \varphi$ . This relies on an induction on  $\varphi$ ; the rest of the section is dedicated to proving the cases of the true and mere belief operators, the other cases being straightforward.

**Lemma 8.6.** *For any  $B = (T_1, M_1, \dots, T_n, M_n, V) \in \mathbf{B}^{\text{TM}}$ , any  $i$  and any  $\nu \in T_i \cup M_i$ ,  $(B, \mathbf{B}^{\text{TM}}) \models BA_i\nu$ , and if  $B^i$  is as defined above, then  $(B, \mathbf{B}^{\text{TM}}) \models B_i\nu$  iff  $B^i \models \nu$ , and  $(B, \mathbf{B}^{\text{TM}}) \models B_i\neg\nu$  iff  $B^i \not\models \nu$ .*

*Proof.* As  $B\mathcal{R}_i^{\text{TM}}B^i$ , to show that  $(B, \mathbf{B}^{\text{TM}}) \models B_i\nu$ , it suffices to show that  $(B^i, \mathbf{B}^{\text{TM}}) \models K_i\nu$ , and similarly for  $\neg\nu$ . Consider  $B' = (T'_1, M'_1, \dots, T'_n, M'_n, V)$  such that  $B^i\mathcal{R}_i^{\text{TM}}B'$ . We wish to show that  $B^i$  and  $B'$  agree on  $\nu$  for any  $\nu \in T_i \cup M_i$ . By definition of  $\mathcal{R}_i^{\text{TM}}$ , we know that  $T_i \cup M_i \subseteq T'_i$ , and that  $B^i$  and  $B'$  agree on  $(T_i \cup M_i) \cap T'_i$ , that is, on  $T_i \cup M_i$ , giving us the desired result.  $\square$

**Lemma 8.7.** *For any  $\nu \in \mathcal{L}_0$  and any  $B \in \mathbf{B}^{\text{TM}}$ , if  $(B, \mathbf{B}^{\text{TM}}) \models BA_i\nu$  then  $\nu \in T_i \cup M_i$ .*

*Proof.* If  $(B, \mathbf{B}^{\text{TM}}) \models BA_i\nu$  then for any  $B'$  such that  $B^i\mathcal{R}_i^{\text{TM}}B'$ ,  $B^i$  and  $B'$  must agree on  $\nu$ . Suppose that  $\nu \notin T_i \cup M_i$ . Then we can construct a state  $B' = (T'_1, M'_1, \dots, T'_n, M'_n, V)$  which differs from  $B^i$  only on the value of  $\nu$ . That is,  $B'$  is constructed in the following way:

- if  $\nu = p$  for some  $p \in \text{Prop}$ , then  $T'_k = T_k^i$  and  $M'_k = M_k^i$  for all  $k$ , and  $V'$  and  $V^i$  differ only on  $p$  ( $V' = V^i \cup \{p\}$  if  $p \notin V^i$  and  $V' = V^i \setminus \{p\}$  if  $p \in V^i$ );
- if  $\nu = \theta_j\nu'$  for some  $j$  and  $\nu'$ , then  $T'_k = T_k^i$  for all  $k \neq j$ ,  $T'_j$  and  $T_j^i$  differ only on  $\nu'$ ,  $M'_k = M_k^i$  for all  $k$ , and  $V' = V^i$ ;
- if  $\nu = \mu_j\nu'$  for some  $j$  and  $\nu'$ , then  $T'_k = T_k^i$  for all  $k$ ,  $M'_k = M_k^i$  for all  $k \neq j$ ,  $M'_j$  and  $M_j^i$  differ only on  $\nu'$ , and  $V' = V^i$ .

It is easily shown that  $B^i\mathcal{R}_i^{\text{TM}}B'$ , but  $B^i$  and  $B'$  disagree on  $\nu$ . This contradicts our first hypothesis. Therefore  $\nu \in T_i \cup M_i$ .  $\square$

**Proposition 8.10.** *For any  $\nu \in \mathcal{L}_{\text{TBA, MBA}}$ , for any  $B \in \mathbf{B}^{\text{TM}}$ ,  $(B, \mathbf{B}^{\text{TM}}) \models \theta_i\nu \leftrightarrow TBA_i\nu$  and  $(B, \mathbf{B}^{\text{TM}}) \models \mu_i\nu \leftrightarrow MBA_i\nu$ .*

*Proof.* Consider  $B = (T_1, M_1, \dots, T_n, M_n, V) \in \mathbf{B}^{\text{TM}}$  and  $p \in \text{Prop}$  and suppose that  $(B, \mathbf{B}^{\text{TM}}) \models \theta_i\nu$ . Then  $\nu \in T_i$  and, according to Lemma 8.6,  $(B, \mathbf{B}^{\text{TM}}) \models B_i\nu$  iff  $B^i \models \nu$ , and  $(B, \mathbf{B}^{\text{TM}}) \models B_i\neg\nu$  iff  $B^i \not\models \nu$ . But by definition of  $B^i$ ,  $B^i \models \nu$  iff  $B \models \nu$ . Therefore  $(B, \mathbf{B}^{\text{TM}}) \models T_i\nu$ .

Suppose now that  $(B, \mathbf{B}^{\text{TM}}) \models T_i\nu$ . Then by Lemma 8.7  $\nu \in T_i \cup M_i$ , and  $B$  and  $B^i$  agree on  $\nu$ . Therefore  $\nu \in T_i$ .

We move on to the case where  $(B, \mathbf{B}^{\text{TM}}) \models \mu_i\nu$ . Then  $\nu \in M_i$ , and Lemma 8.6 tells us that  $(B, \mathbf{B}^{\text{TM}}) \models BA_i\nu$ ; it remains to show that there exists  $B' \in \mathbf{B}^{\text{TM}}$  such that  $B\mathcal{R}_i^{\text{TM}}B'$  and  $B$  and  $B'$  disagree on  $\nu$ . We once again consider three cases depending on the form of  $\nu$ :

- if  $\nu = p$  for some  $p \in \text{Prop}$ , we consider  $B'$  such that  $T'_k = T_k$  for all  $k \neq i$ ,  $M'_k = M_k$  for all  $k$ ,  $T_i$  and  $T'_i$  differ only by  $p$ , and  $V$  and  $V'$  differ only by  $p$ ;
- if  $\nu = \theta_j\nu'$  for some  $j$  and  $\nu'$ , we consider  $B'$  such that  $T'_k = T_k$  for all  $k \neq i, j$ ,  $M'_k = M_k$  for all  $k$ ,  $V' = V$ ,  $T_i$  and  $T'_i$  differ only by  $\nu$ , and  $T_j$  and  $T'_j$  differ only by  $\nu'$ ;

- if  $\nu = \mu_j \nu'$  for some  $j$  and  $\nu'$ , we consider  $B'$  such that  $T'_k = T_k$  for all  $k \neq i$ ,  $M'_k = M_k$  for all  $k \neq j$ ,  $V' = V$ ,  $T_i$  and  $T'_i$  differ only by  $\nu$ , and  $M_j$  and  $M'_j$  differ only by  $\nu'$ .

Clearly  $B$  and  $B'$  disagree on  $\nu$  in all cases. More precisely,  $B$  and  $B'$  agree on all atoms except for  $\theta_i \nu$  and  $\nu$ . We now show that  $B\mathcal{R}_i^{\text{TM}}B'$ .

1.  $T_i \cup M_i = T'_i \cup M'_i$  as  $T_i$  and  $T'_i$  differ only by  $\nu$  and  $\nu \in M_i = M'_i$ ;
2.  $T_i \setminus M_i = T'_i \setminus M'_i$  for the same reason as above;
3. Consider  $\nu'' \in T_i \cap T'_i$ . Then  $\nu'' \neq \nu$ , and therefore  $B$  and  $B'$  agree on  $\nu$ . Consider now  $\nu'' \in (M_i \setminus T_i) \cap (M'_i \setminus T'_i)$ .  $\nu$  is in  $(M_i \setminus T_i)$  iff it is in  $T'_i$ , therefore  $\nu'' \neq \nu$  and  $B$  and  $B'$  agree on  $\nu$ .
4. Finally,  $\nu$  is the only member of  $T_i \Delta T'_i$ , and  $B$  and  $B'$  disagree on  $\nu$ .

Therefore  $B\mathcal{R}_i^{\text{TM}}B'$ , and we can conclude that  $(B, \mathbf{B}^{\text{TM}}) \models M_i \nu$ .

The last case is that in which  $(B, \mathbf{B}^{\text{TM}}) \models M_i \nu$ . Then by Lemma 8.7,  $\nu \in T_i \cup M_i$ ; moreover, there exists  $B''$  such that  $B\mathcal{R}_i^{\text{TM}}B''$  and  $B$  and  $B'$  disagree on  $\nu$ . Suppose that  $\nu \in T_i \setminus M_i$ . Then by definition of  $\mathcal{R}_i^{\text{TM}}$ ,  $\nu \in T'_i \setminus M'_i$  and  $B$  and  $B''$  agree on  $\nu$ , which contradicts our hypothesis. Therefore  $\nu \in M_i$ .  $\square$

## 8.6 Discussion and conclusion

We have shown how EL-O, knowledge structures, PEKBs, and the S4.2-based logic of true and mere belief can be embedded into Lorini's LDA framework. Our embeddings rely on small adjustments to the language and semantics of LDA.

We have shown in particular that both EL-O and knowledge structures can be embedded into a particular version of LDA models which we call vision models. As both the logic of vision models and that of knowledge structures have the same expressivity as standard epistemic logics, this result tells us that we could restrict our consideration of vision models to those where the context is in  $\mathbf{Cxt}^{\text{KS}}$ . However, allowing for full generality of contexts is not completely redundant, as it can allow for a more natural way of expressing simple situations with given higher-order knowledge, whereas this construction is not that clear for knowledge states.

Altogether, we have shown that LDA framework provides a general and flexible tool for reasoning about knowledge and beliefs by combining the compactness of symbolic approaches with the flexibility of possible worlds approaches. This makes it a very useful reference framework against which many other existing formalisms can be compared.



## Chapter 9

# Conclusion

Dans cette thèse nous avons défini et étudié quelques modélisations simples de raisonnement sur et planification avec de la connaissance et de la croyance dans des systèmes multiagents. En particulier, nous avons fait un premier pas vers une planification épistémique multi-agent réaliste avec la logique EL-O. Notre approche est symbolique et se base sur des atomes d’observabilité, autrement dit des agents avec raisonnement simplifié basé sur le concept de ‘savoir si’ plutôt que le traditionnel ‘savoir que’.

Du côté statique, nous avons présenté plusieurs versions de EL-O et montré qu’il s’agissait de fragments de différentes version de la logique standard S5 (avec et sans l’axiome d’induction pour la connaissance commune). Nous avons prouvé que le problème de satisfiabilité pour chacune de ces versions de EL-O est NP-complet, donc dans la même classe de complexité que le problème de satisfiabilité en logique classique propositionnelle. De plus, la nouvelle sémantique que nous avons défini pour EL-O dans le Chapitre 3 nous permet de travailler avec des modèles finis, ce qui s’avère être un élément crucial lorsqu’il s’agit de définir la planification dans EL-O et de montrer que le problème de décision de l’existence d’un plan étant donnée une tâche de planification est PSPACE-complet, encore une fois la même classe complexité que le problème propositionnel classique correspondant. Ce résultat tient à la fois pour la planification séquentielle et parallèle. Nous avons prouvé ceci en définissant une traduction polynomiale des tâches de planification EL-O vers des tâches de planification classiques, ce qui signifie que les planificateurs classiques peuvent être utilisés pour résoudre des problèmes de planification épistémique sans perte significative d’efficacité dès lors que ces problèmes peuvent être exprimés dans le formalisme de EL-O. Nous avons aussi défini une extension dynamique de EL-O qui nous permet de capturer la solvabilité des tâches de planifications comme un problème de vérification de modèle.

La complexité de la planification dans EL-O présente une amélioration significative par rapport à la planification DEL, qui est indécidable même pour des fragments relativement simples. Par exemple, si les actions modifient aussi le monde (et non seulement les connaissances des agents), alors le problème de l’existence d’un plan est indécidable dès que des opérateurs épistémiques sont autorisés à apparaître dans les préconditions ; si les actions sont purement épistémiques, alors le problème est indécidable dès que le nombre d’agents ou la profondeur épistémique dépasse 2. Nous avons aussi défendu l’idée que les

atomes d’observabilité fournissent une manière simple et finie de représenter des situations d’ignorance maximale, qui deviennent simplement des états contenant seulement des informations propositionnelles, tandis que dans la logique épistémique standard la représentation de telles situations nécessite des modèles de Kripke infinis. De manière générale, dans DEL, l’incertitude fait augmenter la taille des modèles, tandis que le contraire est vrai pour EL-O. De plus, nous avons montré que la correspondance entre les actions EL-O et les modèles à événements de DEL n’est pas si claire. Ceci est dû à une différence fondamentale des approches de modélisation entre EL-O et DEL : dans EL-O, on se préoccupe plutôt des effets directs d’une action (c.à.d. quels atomes deviennent vrais ou faux), alors que dans DEL, la modélisation s’articule autour de l’incertitude, à partir de laquelle la connaissance et son évolution sont déterminées. Bien sûr, puisque EL-O est une logique restreinte, toutes les actions ne peuvent pas y être modélisées, mais nous défendons l’idée que dans bien des cas, EL-O offre au modélisateur de tâches de planification une manière très naturelle de définir modèles et actions, loin de n’être qu’une version restreinte et “moindre” de DEL.

Comparé à d’autres logiques épistémiques simplifiées pour la planification épistémique, notre approche gagne en expressivité : EL-O permet de modéliser de la connaissance d’ordre supérieur, de la connaissance commune, des opérateurs moins restreints que les ‘littéraux modaux restreints’ de la littérature, et une plus grande liberté dans la définition des actions, qui ne sont pas limitées aux seules actions publiques ou semi-privées. Il est aussi possible d’incorporer dans EL-O une notion de ‘savoir ce qu’est’ (ou ‘connaître la valeur de’), dont l’importance dans les applications en intelligence artificielle est soulignée dans la littérature. Nous avons aussi comparé les logiques présentées dans cette thèse avec d’autres logiques épistémiques symboliques proches en encodant le tout dans le cadre très général des logiques LDA de Lorini. Il serait intéressant de se pencher également sur la question de l’applicabilité à la planification d’autres logiques épistémiques simplifiées développées pour résoudre d’autres problèmes tels que la vérification de modèles, en particulier celles de [56], et d’en comparer les résultats avec notre approche.

L’efficacité de EL-O repose sur la limitation des capacités de raisonnement des agents. En particulier, les agents ne peuvent connaître la valeur de vérité de disjonctions arbitraires. En conséquence, EL-O ne peut pas être utilisé pour modéliser des problèmes tels que celui des Muddy Children. De plus, comme développé dans la thèse, la généralisation de notre opérateur de vision jointe à des groupes quelconques d’agents n’est pas évidente et demeure un problème ouvert.

Les succès que nous avons rencontrés avec EL-O nous ont tout de même incités à étudier l’adaptabilité des éléments clés de cette approche à des logiques non seulement de connaissance mais aussi de croyance. À cet effet nous avons proposé deux nouveaux opérateurs, l’un de ‘croyance correcte’ et l’autre de ‘simple croyance’, à l’aide desquels toutes les situations épistémico-doxastiques peuvent être représentées de manière combinatoire. Cela permet de ne pas avoir à gérer les interactions entre les traditionnels opérateurs de connaissance et de croyance, par exemple le fait que la connaissance implique la croyance. Nous avons étudié le comportement de ces nouveaux opérateurs dans trois logiques épistémiques doxastiques standard. La prochaine étape pourrait être d’étudier des restrictions à des ‘atomes épistémico-doxastiques’ similaires aux atomes d’observation de EL-O, ou bien de se pencher sur le problème de la planifi-

cation avec des connaissances et des croyances. Il est intéressant de noter que lorsque l'on considère des opérateurs de 'croyance à propos de', les problèmes de cohérence que l'on retrouve au coeur de nombreuses approches de révision des croyances disparaissent : la question est plutôt de savoir si l'agent choisira ou non d'avoir une croyance à propos de la proposition en question, et de savoir si cette croyance sera vraie ou non. Autrement dit, la difficulté ne porte plus sur *comment* mettre à jour des bases de croyances, mais plutôt de décider *s'il faut ou non* les mettre à jour. Ceci mène à un changement de perspective que nous espérons étudier plus en profondeur dans des travaux futurs.

Enfin, prenant un détour par les axiomatisations de logiques de connaissance commune, nous avons proposé une nouvelle axiomatisation pour la connaissance commune **S5** avec un axiome d'induction que nous trouvons plus intuitif que les solutions préexistantes. Notre axiomatisation est spécifique à la connaissance et fonctionne aussi bien pour les logiques avec l'opérateur 'savoir que' que pour celles avec 'savoir si', mais n'est pas adaptée aux croyances. Il serait intéressant d'étudier une généralisation de ce nouvel axiome aux logiques qui ne suivent pas l'axiome de vérité.

Nous travaillons dans cette thèse avec une approche centralisée dans laquelle un planificateur unique décide des actions de tous les agents. D'autres approches ont étudié le problème de bavardage classique d'un point de vue complètement différent, avec des agents tous autonomes. Plus généralement, prenons l'action de demander de l'information à un autre agent : cette situation a lieu lorsque le premier agent sait que le second agent connaît l'information en question. Cependant, si la planification est centralisée, seul le planificateur omniscient a théoriquement besoin de savoir que le second agent a l'information pour ordonner au premier agent de l'interroger, qu'importe ce que sait ce dernier. En ce sens, élargir notre approche à de la planification distribuée et autonome étendrait nettement son domaine d'application.

---

In this thesis we have defined and investigated a few simple frameworks for reasoning about and planning with knowledge and belief in multiagent systems. Most notably, we have made a first step towards realistic multi-agent epistemic planning through the EL-O framework. Our approach is a symbolic approach based on observability atoms, that is, simple reasoning revolving around the concept of 'knowing whether' rather than the traditional 'knowing that'.

On the static side, we have given a few versions of EL-O which we have shown to be fragments of different versions of **S5** (with and without the induction axiom for common knowledge) and for which we have shown the satisfiability problem to be NP-complete, just like the satisfiability problem in Classical Propositional Calculus. Moreover, the improved semantics that we have defined for EL-O in Chapter 3 allow us to work with finite models, which is instrumental in defining EL-O planning and showing that the problem of deciding the existence of a plan given a planning task is PSPACE-complete, once again the same complexity as its classical counterpart. This holds both for sequential and parallel planning. We have shown this by defining a polynomial translation of EL-O planning tasks into classical planning tasks, which means that classical planners may be used to solve epistemic planning problems with no significant loss of efficiency, so long as these problems can be expressed in the EL-O formalism. We have also defined a dynamic extension of EL-O which allows us to capture solvability of

planning tasks as a model-checking problem.

The complexity result for EL-O planning is a significant improvement over DEL planning, which is undecidable even for simple fragments. For example, if actions also change the world (and not only the agents' knowledge), then plan existence is undecidable whenever epistemic operators are allowed in pre-conditions; if actions are purely epistemic, then it is undecidable whenever two agents are involved or the epistemic depth exceeds 2 [9]. We have also argued that observability atoms provide a simple, finite way to represent situations of maximal ignorance, namely by states that only hold propositional information, while such situations require infinite Kripke models in standard epistemic logic. More generally, in DEL, models get bigger the more uncertainty there is, while the opposite is true in EL-O. Moreover, we have shown that the correspondence between EL-O actions and DEL event models is not that clear, due to a fundamental difference in approaches when defining such actions and models: in EL-O, the focus is on the effects of an action (i.e., which atoms become true or false), whereas in DEL, the focus is on the uncertainty, from which knowledge and the evolution thereof are computed. Of course, the restricted nature of EL-O means that every possible action may not be modeled in EL-O, but we have argued that in many cases, EL-O offers to the designer of a planning task a very natural way of defining models and actions, far from being but a "lesser", more restricted DEL.

In comparison to other lightweight epistemic logics and epistemic planning frameworks, our approach gains in expressivity: we have higher-order knowledge, a form of common knowledge, more reasoning power than the similar restricted modal literals of the literature, and great freedom in the definition of actions, which may be public, private, or anything in between. We also incorporate a notion of 'knowing what' (or 'knowing the value'), which has been argued to be of crucial importance in AI applications. We have further assessed the frameworks proposed in this thesis and other related symbolic approaches through encoding into the very general framework of Lorini's LDA logics. It would be interesting to further investigate the compared applicability to planning of other simplified epistemic logics designed for other problems such as model checking, in particular those of [56].

At the core of the efficiency of EL-O is the agents' limited reasoning power. In particular, agents cannot know arbitrary disjunctions. This means that EL-O cannot be used to model problems such as the muddy children problem. Moreover and as discussed in this thesis, it is not obvious how to generalize our joint vision operator to arbitrary groups of agents, and we leave this as an open problem for now.

Nevertheless, the success with EL-O has prompted us to investigate whether the key elements of the approach can be transferred over to a logic not only of knowledge but also of belief. To this end we have introduced two new operators, one of true belief and one of mere belief, which allow us to represent all epistemic-doxastic situations in a combinatorial manner. This allows us to avoid the interaction between the traditional knowledge and belief operators, such as knowledge implying belief. We have investigated the behavior of these new operators in three different standard epistemic-doxastic logics. The next step might be to investigate restrictions to 'epistemic-doxastic atoms' similar to the EL-O observability atoms, or look into planning with knowledge and belief. Interestingly when considering 'belief about' modalities, the consistency issue

at the core of many approaches to belief revision becomes void: the question is rather whether or not the agent will choose to have a belief about the proposition at hand, and whether or not that belief will be true. That is, the question is not of *how* to update belief bases but rather of *whether* to update them, leading to a change in perspective which we leave as future work to explore.

Finally, through a detour into the world of axiomatizations of common knowledge we have proposed a new alternative for **S5** common knowledge featuring an induction axiom for common knowledge which we find more intuitively appealing than existing solutions. Our axiomatization is tailored to knowledge and translates well to the ‘knowing whether’ setting, but we have shown that it is not suited for beliefs. An interesting avenue for future work would be a generalization of this new axiom to logics which do not obey the truth axiom.

In this thesis we have assumed a centralised approach in which a single planner decides the actions of all agents. Other approaches have studied the classical gossip problem from a completely different perspective, assuming that all agents are autonomous [5, 48, 123]. More generally, consider the action of asking another agent for information: we have argued that this happens when the first agent knows that the second agent has said information. However, when considering centralized planning only the omniscient planner would theoretically need to know that the second agent has the information in order to tell the first agent to ask the question, no matter what the first agent knows. In this regard considering distributed and autonomous planning would greatly widen the applicability of our framework.

# Bibliography

- [1] Eralp A. Akkoyunlu, Kattamuri Ekanadham, and Richard V. Hubert. Some constraints and tradeoffs in the design of network communications. In *Proceedings of the 5th ACM Symposium on Operating Systems Principles (SOSP 1975)*, pages 67–74. ACM Press, 1975.
- [2] Luca Alberucci and Gerhard Jäger. About cut elimination for logics of common knowledge. *Annals of Pure and Applied Logic*, 133(1-3):73–99, 2005.
- [3] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The journal of symbolic logic*, 50(2):510–530, 1985.
- [4] Sergei N. Artemov. Observable models. In Sergei N. Artemov and Anil Nerode, editors, *Logical Foundations of Computer Science - International Symposium, LFCS 2020, Deerfield Beach, FL, USA, January 4-7, 2020, Proceedings*, volume 11972 of *Lecture Notes in Computer Science*, pages 12–26. Springer, 2020.
- [5] Maduka Attamah, Hans van Ditmarsch, Davide Grossi, and Wiebe van der Hoek. Knowledge and gossip. In *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI 2014)*, pages 21–26, 2014.
- [6] Guillaume Aucher. Axioms .2 and .4 as interaction axioms. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.
- [7] Guillaume Aucher. Principles of Knowledge, Belief and Conditional Belief. In Manuel Rebuschi, Martine Batt, Gerhard Heinzmann, Franck Lihoreau, Michel Musiol, and Alain Trognon, editors, *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics*, pages 97–134. Springer, 2014.
- [8] Guillaume Aucher. Intricate axioms as interaction axioms. *Studia Logica*, 103(5):1035–1062, 2015.
- [9] Guillaume Aucher and Thomas Bolander. Undecidability in epistemic planning. In Francesca Rossi, editor, *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, pages 27–33. AAAI Press, 2013.

- [10] Robert Audi, editor. *The Cambridge Dictionary of Philosophy, Second Edition*. Cambridge University Press, 1999. entry ‘Action’.
- [11] Brenda Baker and Robert Shostak. Gossips and telephones. *Discrete Mathematics*, 2(3):191–193, 1972.
- [12] Matteo Baldoni. *Normal multimodal logics: Automatic deduction and logic programming extension*. PhD thesis, Università degli Studi di Torino, Dipartimento di Informatica, 1998.
- [13] Matteo Baldoni, Laura Giordano, and Alberto Martelli. A modal extension of logic programming: Modularity, beliefs and hypothetical reasoning. *J. Log. Comput.*, 8(5):597–635, 1998.
- [14] Alexandru Baltag. To know is to know the value of a variable. *Advances in modal logic*, 11:135–155, 2016.
- [15] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 1998)*, pages 43–56. Morgan Kaufmann Publishers, 1998.
- [16] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. *Logic and the foundations of game and decision theory (LOFT 7)*, 3:9–58, 2008.
- [17] Alexandru Baltag and Johan van Benthem. A simple logic of functional dependence. *Journal of Philosophical Logic*, pages 1–67, 2021.
- [18] Simon Baron-Cohen, Alan M. Leslie, and Uta Frith. Does the autistic child have a theory of mind? *Cognition*, 21(1):37–46, 1985.
- [19] Alex Bavelas. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.
- [20] Salem Benferhat, Didier Dubois, Henri Prade, and Mary-Anne Williams. A practical approach to revising prioritized knowledge bases. *Stud Logica*, 70(1):105–130, 2002.
- [21] Johan van Benthem, Jan van Eijck, Malvin Gattinger, and Kaile Su. Symbolic model checking for Dynamic Epistemic Logic. In Wiebe van der Hoek, Wesley H. Holliday, and Wen-fang Wang, editors, *Proceedings of the 5th International Conference on Logic, Rationality and Interaction (LORI 2015)*, pages 366–378. Springer-Verlag, 2015.
- [22] Johan van Benthem, Jan van Eijck, Malvin Gattinger, and Kaile Su. Symbolic model checking for Dynamic Epistemic Logic - S5 and beyond. *J. Log. Comput.*, 28(2):367–402, 2018.
- [23] Johan van Benthem, Jan van Eijck, and Barteld Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.

- [24] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.
- [25] Thomas Bolander. Seeing is believing: Formalising false-belief tasks in dynamic epistemic logic. In Hans van Ditmarsch and Gabriel Sandu, editors, *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pages 207–236. Springer International Publishing, Cham, 2018.
- [26] Thomas Bolander and Mikkel Birkegaard Andersen. Epistemic planning for single and multi-agent systems. *Journal of Applied Non-Classical Logics*, 21(1):9–34, 2011.
- [27] Thomas Bolander, Tristan Charrier, Sophie Pinchinat, and François Schwarzentruber. Del-based epistemic planning: Decidability and complexity. *Artif. Intell.*, 287:103304, 2020.
- [28] Thomas Bolander, Thorsten Engesser, Robert Mattmüller, and Bernhard Nebel. Better eager than lazy? How agent types impact the successfulness of implicit coordination. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018*, pages 445–453, 2018.
- [29] Thomas Bolander, Martin Holm Jensen, and François Schwarzentruber. Complexity results in epistemic planning. In Qiang Yang and Michael Wooldridge, editors, *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2791–2797. AAAI Press, 2015.
- [30] Thomas Bolander, Hans van Ditmarsch, Andreas Herzig, Emiliano Lorini, Pere Pardo, and François Schwarzentruber. Announcements to attentive agents. *Journal of Logic, Language and Information*, 25(1):1–35, 2016.
- [31] Giacomo Bonanno. On the logic of common belief. *Math. Log. Q.*, 42:305–311, 1996.
- [32] Samuel Bucheli, Roman Kuznets, and Thomas Studer. Two ways to common knowledge. *Electronic Notes in Theoretical Computer Science*, 262:83–98, 2010.
- [33] Tom Bylander. The computational complexity of propositional STRIPS planning. *Artificial Intelligence*, 69:165–204, 1994.
- [34] Aggeliki Chalki, Costas D Koutras, and Yorgos Zikos. A quick guided tour to the modal logic  $S4.2$ . *Logic Journal of the IGPL*, 26(4):429–451, 2018.
- [35] Aggeliki Chalki, Costas D Koutras, and Yorgos Zikos. A note on the complexity of  $S4.2$ . *Journal of Applied Non-Classical Logics*, pages 1–22, 2021.
- [36] Tristan Charrier, Andreas Herzig, Emiliano Lorini, Faustine Maffre, and François Schwarzentruber. Building epistemic logic from observations and public announcements. In Chitta Baral, James P. Delgrande, and Frank



- Wolter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016*, pages 268–277. AAAI Press, 2016.
- [37] Sébastien Lê Cong, Sophie Pinchinat, and François Schwarzentruber. Small undecidable problems in epistemic planning. In Jérôme Lang, editor, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 4780–4786. ijcai.org, 2018.
- [38] Martin Cooper, Andreas Herzig, Faustine Maffre, Frédéric Maris, Elise Perrotin, and Pierre Régnier. When ‘knowing whether’ is better than ‘knowing that’. In *13èmes Journées d’Intelligence Artificielle Fondamentale (JIAF 2019)*. hal-02302984, July 2019.
- [39] Martin Cooper, Andreas Herzig, Faustine Maffre, Frédéric Maris, Elise Perrotin, and Pierre Régnier. A lightweight epistemic logic and its application to planning. *Artificial Intelligence*, 298:103437, 2020.
- [40] Martin Cooper, Andreas Herzig, Faustine Maffre, Frédéric Maris, and Pierre Régnier. A simple account of multi-agent epistemic planning. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, pages 193–201, 2016.
- [41] Martin Cooper, Andreas Herzig, Faustine Maffre, Frédéric Maris, and Pierre Régnier. The epistemic gossip problem. *Discrete Mathematics*, 342(3):654–663, 2019.
- [42] Martin Cooper, Andreas Herzig, Frédéric Maris, Elise Perrotin, and Julien Vianey. Lightweight Parallel Multi-Agent Epistemic Planning. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning*, pages 274–283, 9 2020.
- [43] Martin Cooper, Andreas Herzig, Frédéric Maris, and Julien Vianey. Temporal epistemic gossip problems. In Marija Slavkovik, editor, *Multi-Agent Systems - 16th European Conference, EUMAS 2018, Bergen, Norway, December 6-7, 2018, Revised Selected Papers*, volume 11450 of *Lecture Notes in Computer Science*, pages 1–14. Springer, 2018.
- [44] Robin P Cubitt and Robert Sugden. Common knowledge, salience and convention: A reconstruction of david lewis’ game theory. *Economics & Philosophy*, 19(2):175–210, 2003.
- [45] Robert Demolombe and Maria del Pilar Pozos Parra. A simple and tractable extension of situation calculus to epistemic logic. In Zbigniew W. Ras and Setsuo Ohsuga, editors, *Foundations of Intelligent Systems, 12th International Symposium, ISMIS 2000, Charlotte, NC, USA, October 11-14, 2000, Proceedings*, volume 1932 of *Lecture Notes in Computer Science*, pages 515–524. Springer, 2000.
- [46] Lasse Dissing and Thomas Bolander. Implementing theory of mind on a robot using dynamic epistemic logic. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, (IJCAI 2020)*, pages 1615–1621, 2020.

- [47] Hans P. van Ditmarsch. *Knowledge games*. PhD thesis, University of Groningen, 2000.
- [48] Hans van Ditmarsch, Davide Grossi, Andreas Herzig, Wiebe van der Hoek, and Louwe B. Kuijer. Parameters for epistemic gossip problems. In *Proc. LOFT 2016*, 2016.
- [49] Hans van Ditmarsch, Joseph Y Halpern, Wiebe van der Hoek, and Barteld Kooi. *Handbook of Epistemic Logic*. College Publications, 2015.
- [50] Hans van Ditmarsch, Andreas Herzig, Jérôme Lang, and Pierre Marquis. Introspective forgetting. *Synthese*, 169(2):405–423, 2009.
- [51] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer Publishing Company, Incorporated, 1st edition, 2007.
- [52] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
- [53] Jie Fan, Davide Grossi, Barteld Kooi, Xingchi Su, and Rineke Verbrugge. Commonly knowing whether. *arXiv preprint arXiv:2001.03945*, 2020.
- [54] Jie Fan, Yanjing Wang, and Hans van Ditmarsch. Contingency and knowing whether. *Rew. Symb. Logic*, 8(1):75–107, 2015.
- [55] Satoshi Fukuda. Formalizing common belief with no underlying assumption on individual beliefs. *Games and Economic Behavior*, 2020.
- [56] Malvin Gattinger. *New Directions in Model Checking Dynamic Epistemic Logic*. PhD thesis, University of Amsterdam, 2018.
- [57] Hector Geffner and Blai Bonet. *A Concise Introduction to Models and Methods for Automated Planning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2013.
- [58] Michael R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers, 1989.
- [59] Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of logic, language and information*, 6(2):147–169, 1997.
- [60] Jelle Douwe Gerbrandy et al. *Bisimulations on planet Kripke*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, 1999.
- [61] Edmund L. Gettier. Is justified true belief knowledge? *Analysis*, 23(6):121–123, 1963.
- [62] Paul Gochet and Pascal Gribomont. Epistemic logic. In D.M. Gabbay and J. Woods, editors, *Handbook of the History and Philosophy of Logic*, volume 4, pages 463–502. Elsevier Science, Amsterdam, 2003.
- [63] Jim Gray. Notes on data base operating systems. In *Operating Systems, An Advanced Course*, pages 393–481. Springer-Verlag, 1978.

- [64] Andras Hajnal, Eric C. B. Milner, and Endre Szemerédi. A cure for the telephone disease. *Canadian Mathematical Bulletin*, 15(3):447–450, 1972.
- [65] Joseph Y. Halpern and Yoram Moses. A guide to the modal logics of knowledge and belief: Preliminary draft. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI’85*, pages 480–490. Morgan Kaufmann Publishers Inc., 1985.
- [66] Joseph Y. Halpern and Yoram Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379, 1992.
- [67] Joseph Y Halpern and Rafael Pass. A knowledge-based analysis of the blockchain protocol. *arXiv preprint arXiv:1707.08751*, 2017.
- [68] Sven Ove Hansson. Theory contraction and base contraction unified. *Journal of Symbolic Logic*, 58(2):602–625, 1993.
- [69] Andreas Herzig. Logics of knowledge and action: critical analysis and challenges. *Auton. Agents Multi Agent Syst.*, 29(5):719–753, 2015.
- [70] Andreas Herzig, Emiliano Lorini, and Faustine Maffre. A poor man’s epistemic logic based on propositional assignment and higher-order observation. In Wiebe van der Hoek, Wesley H. Holliday, and Wen-fang Wang, editors, *Proceedings of the 5th International Conference on Logic, Rationality and Interaction (LORI 2015)*, pages 156–168. Springer Verlag, 2015.
- [71] Andreas Herzig, Emiliano Lorini, and Faustine Maffre. Possible worlds semantics based on observation and communication. In Hans van Ditmarsch and Gabriel Sandu, editors, *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pages 339–362. Springer International Publishing, Cham, 2018.
- [72] Andreas Herzig, Emiliano Lorini, Elise Perrotin, Fabián Romero, and François Schwarzentruber. A logic of explicit and implicit distributed belief. In *24th European Conference on Artificial Intelligence (ECAI 2020)*, pages 753–760. IOS Press, 2020.
- [73] Andreas Herzig and Faustine Maffre. How to share knowledge by gossiping. In *Proceedings of the 3rd International Conference on Agreement Technologies (AT 2015)*, pages 249–263. Springer-Verlag, 2016.
- [74] Andreas Herzig, Frédéric Maris, and Julien Vianey. Dynamic logic of parallel propositional assignments and its applications to planning. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5576–5582. ijcai.org, 2019.
- [75] Brian Hill and Francesca Poggiolesi. Common knowledge: finite calculus with a syntactic cut-elimination procedure. *Logique et Analyse*, 58(230):279–306, 2015.

- [76] Jaakko Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, 1962.
- [77] Wiebe van der Hoek and Michael Wooldridge. Tractable multiagent planning for epistemic goals. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3*, pages 1167–1174, 2002.
- [78] I. L. Humberstone. The logic of non-contingency. *Notre Dame Journal of Formal Logic*, 36(2):214–229, 1995.
- [79] Walter Knödel. New gossips and telephones. *Discrete Mathematics*, 13(1):95, 1975.
- [80] Filippos Kominis and Hector Geffner. Beliefs in multiagent planning: From one agent to many. In Ronen I. Brafman, Carmel Domshlak, Patrik Haslum, and Shlomo Zilberstein, editors, *Proceedings of the Twenty-Fifth International Conference on Automated Planning and Scheduling, ICAPS 2015, Jerusalem, Israel, June 7-11, 2015*, pages 147–155. AAAI Press, 2015.
- [81] Filippos Kominis and Hector Geffner. Multiagent online planning with nested beliefs and dialogue. In Laura Barbulescu, Jeremy Frank, Mausam, and Stephen F. Smith, editors, *Proceedings of the Twenty-Seventh International Conference on Automated Planning and Scheduling, ICAPS 2017, Pittsburgh, Pennsylvania, USA, June 18-23, 2017*, pages 186–194. AAAI Press, 2017.
- [82] K. Konolige. What awareness isn't: A sentential view of implicit and explicit belief. In *Proceedings of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge, Monterey, CA, USA, March 1986*, pages 241–250. Morgan Kaufmann, 1986.
- [83] Sarit Kraus and Daniel Lehmann. Knowledge, belief and time. *TCS*, 58:155–174, 1988.
- [84] Gerhard Lakemeyer and Yves Lespérance. Efficient reasoning in multiagent epistemic logics. In Luc De Raedt, Christian Bessière, Didier Dubois, Patrick Doherty, Paolo Frasconi, Fredrik Heintz, and Peter J. F. Lucas, editors, *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31, 2012*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, pages 498–503. IOS Press, 2012.
- [85] Leslie Lamport, Robert Shostak, and Marshall Pease. The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, 4(3):382–401, 1982.
- [86] Hyman G. Landau. The distribution of completion times for random communication in a task-oriented group. *The Bulletin of Mathematical Biophysics*, 16(3):187–201, 1954.

- [87] Daniel J. Lehmann. Knowledge, common Knowledge and related puzzles (extended summary). In *Proceedings of the 3rd Annual ACM Symposium on Principles of Distributed Computing (PODC 1984)*, pages 62–67, 1984.
- [88] Wolfgang Lenzen. Epistemologische Betrachtungen zu [S4;S5]. *Erkenntnis*, 14:33–56, 1979.
- [89] David Lewis, editor. *Convention: A Philosophical Study*. Harvard University Press, 1969.
- [90] Andrés Occhipinti Liberman, Andreas Achen, and Rasmus Kræmmer Rendsvig. Dynamic term-modal logics for first-order epistemic planning. *Artif. Intell.*, 286:103305, 2020.
- [91] Luc Lismont and Philippe Mongin. On the logic of common belief and common knowledge. *Theory and Decision*, 37(1):75–106, Jul 1994.
- [92] Emiliano Lorini. In praise of belief bases: Doing epistemic logic without possible worlds. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [93] Emiliano Lorini. Exploiting belief bases for building rich epistemic structures. In L. S. Moss, editor, *Proceedings of the Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2019)*, volume 297 of *EPTCS*, pages 332–353, 2019.
- [94] Emiliano Lorini. Rethinking epistemic logic with belief bases. *Artificial Intelligence*, 282:103233, 2020.
- [95] Emiliano Lorini and Andreas Herzig. Direct and indirect common belief. In Anita Konzelmann Ziv and Hans Bernhard Schmid, editors, *Institutions, Emotions, and Group Agents: Contributions to Social Ontology*, pages 355–372. Springer, 2014.
- [96] Emiliano Lorini and Fabián Romero. Decision procedures for epistemic logic exploiting belief bases. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2019)*, pages 944–952. IFAAMAS, 2019.
- [97] Emiliano Lorini and François Schwarzentruber. A computationally grounded logic of graded belief. In *Proceedings of the 17th European Conference on Logics in Artificial Intelligence*, LNCS. Springer, 2021.
- [98] Emiliano Lorini and Pengfei Song. Grounding awareness on belief bases. In M. A. Martins and I. Sedlár, editors, *Dynamic Logic. New Trends and Applications - Third International Workshop, DaLi 2020, Prague, Czech Republic, October 9-10, 2020, Revised Selected Papers*, volume 12569 of *Lecture Notes in Computer Science*, pages 170–186. Springer, 2020.
- [99] Faustine Maffre. *Ignorance is bliss: observability-based dynamic epistemic logics and their applications*. PhD thesis, Université Paul Sabatier-Toulouse III, 2016.

- [100] Bastien Maubert, Sophie Pinchinat, and François Schwarzentruber. Reachability games in dynamic epistemic logic. *arXiv preprint arXiv:1905.12422*, 2019.
- [101] John McCarthy. Epistemological problems of artificial intelligence. In Bonnie Lynn Webber and Nils J. Nilsson, editors, *Readings in Artificial Intelligence*, pages 459 – 465. Morgan Kaufmann Publishers, 1981.
- [102] John-Jules Ch. Meyer and Wiebe van der Hoek. *Epistemic logic for AI and computer science*. Number 41 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1995.
- [103] Tim Miller, Paolo Felli, Christian J. Muise, Adrian R. Pearce, and Liz Sonenberg. ‘Knowing whether’ in proper epistemic knowledge bases. In Dale Schuurmans and Michael P. Wellman, editors, *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 1044–1050. AAAI Press, 2016.
- [104] Robin Milner. An algebraic definition of simulation between programs. In D. C. Cooper, editor, *Proceedings of the 2nd International Joint Conference on Artificial Intelligence. London, UK, September 1-3, 1971*, pages 481–489. William Kaufmann, 1971.
- [105] Dov Monderer and Dov Samet. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1(2):170–190, 1989.
- [106] Hugh Montgomery and Richard Routley. Contingency and non-contingency bases for normal modal logics. *Logique et analyse*, 9(35/36):318–328, 1966.
- [107] Christian Muise, Vaishak Belle, Paolo Felli, Sheila A. McIlraith, Tim Miller, Adrian R. Pearce, and Liz Sonenberg. Planning over multi-agent epistemic states: A classical planning approach. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015)*, pages 3327–3334. AAAI Press, 2015.
- [108] Héctor Palacios and Hector Geffner. Compiling uncertainty away: Solving conformant planning problems using a classical planner (sometimes). In *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA*, pages 900–905. AAAI Press, 2006.
- [109] Ronald P. A. Petrick and Fahiem Bacchus. A knowledge-based approach to planning with incomplete information and sensing. In Malik Ghalab, Joachim Hertzberg, and Paolo Traverso, editors, *Proceedings of the Sixth International Conference on Artificial Intelligence Planning Systems, April 23-27, 2002, Toulouse, France*, pages 212–222. AAAI, 2002.
- [110] Ronald P. A. Petrick and Fahiem Bacchus. Extending the knowledge-based approach to planning with incomplete information and sensing. In Shlomo Zilberstein, Jana Koehler, and Sven Koenig, editors, *Proceedings of the Fourteenth International Conference on Automated Planning and*

- Scheduling (ICAPS 2004), June 3-7 2004, Whistler, British Columbia, Canada*, pages 2–11. AAAI, 2004.
- [111] Jan Plaza. Logics of public communications. *Synthese*, 158(2):165–179, 2007.
- [112] Jussi Rintanen, Keijo Heljanko, and Ilkka Niemelä. Planning as satisfiability: parallel plans and algorithms for plan search. *Artificial Intelligence*, 170(12-13):1031–1080, 2006.
- [113] H. Rott. Just because”: Taking belief bases seriously. In *Logic Colloquium*, volume 98, pages 387–408, 1998.
- [114] Krister Segerberg. Belief revision from the point of view of doxastic logic. *Logic Journal of the IGPL*, 3(4):535–553, 1995.
- [115] M. Sergot. A computational theory of normative positions. *ACM Trans. Comput. Log.*, 2:581–622, 2001.
- [116] Ilya Shapirovsky. On PSPACE-decidability in transitive modal logic. In Renate A. Schmidt, Ian Pratt-Hartmann, Mark Reynolds, and Heinrich Wansing, editors, *Advances in Modal Logic 5, papers from the fifth conference on "Advances in Modal logic," held in Manchester, UK, 9-11 September 2004*, pages 269–287. King’s College Publications, 2004.
- [117] Giacomo Sillari and Peter Vanderschraaf. Common knowledge. 2014.
- [118] Tran Cao Son and Chitta Baral. Formalizing sensing actions A transition function based approach. *Artif. Intell.*, 125(1-2):19–91, 2001.
- [119] Mark Steedman and Ronald PA Petrick. Planning dialog actions. In *Proceedings of the 8th SIGDIAL Workshop on Discourse and Dialogue (SIGdial 2007)*, pages 265–272, 2007.
- [120] Kaile Su, Abdul Sattar, and Xiangyu Luo. Model checking temporal logics of knowledge via OBDDs. *The Computer Journal*, 50(4):403–420, 2007.
- [121] Robert Tijdeman. On a telephone problem. *Nieuw Archief voor Wiskunde*, 19(3):188–192, 1971.
- [122] Johan Van Benthem. Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155, 2007.
- [123] Hans van Ditmarsch, Jan van Eijck, Pere Pardo, Rahim Ramezani, and François Schwarzentruber. Epistemic protocols for dynamic gossip. *J. Applied Logic*, 20:1–31, 2017.
- [124] Frans Voorbraak. *As Far as I Know. Epistemic Logic and Uncertainty*. PhD thesis, Universiteit Utrecht, 1993.
- [125] Yanjing Wang. Beyond knowing that: a new generation of epistemic logics. In *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pages 499–533. Springer, 2018.

- [126] Yanjing Wang and Jie Fan. Knowing that, knowing what, and public communication: Public announcement logic with kv operators. In *IJCAI*, volume 13, pages 1147–1154. Citeseer, 2013.
- [127] Heinz Wimmer and Josef Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition*, 13(1):103–128, 1983.



Dans cette thèse nous étudions un cadre simple dans lequel modéliser les croyances et les connaissances ainsi que leur évolution dans des systèmes multi-agents. La logique standard de représentation des connaissances est très expressive, mais au prix d'une haute complexité calculatoire. Nous proposons ici un cadre qui permet de capturer plus de situations que d'autres approches existantes tout en restant efficace. En particulier, nous considérons l'application de notre logique à la planification épistémique : étant données une situation initiale et des actions possibles, peut-on atteindre un but fixé ? Cela peut signifier savoir à qui poser des questions pour apprendre des informations, faire en sorte de ne pas être remarquée lorsque l'on lit le courrier de quelqu'un d'autre, ou empêcher quelqu'un d'entendre nos secrets. Nous considérons aussi de possibles extensions à des logiques de croyance, ainsi que les liens entre notre système et d'autres cadres proches.

---

In this thesis we study a lightweight framework in which to model knowledge and beliefs and the evolution thereof in multiagent systems. The standard logic used for this is very expressive, but this comes at a high cost in terms of computational efficiency. We here propose a framework which captures more than other existing approaches while remaining cost-effective. In particular, we show its applicability to epistemic planning: given an initial situation and some possible actions, can we find a way to reach our desired goal? This might mean knowing who to ask in order to learn something, making sure we aren't seen when reading someone else's mail, or preventing someone from overhearing our secrets. We also discuss possible extensions to logics of belief, and the relations between our framework and other related approaches.