Epistemic Effectivity: Neighbourhood Semantics for Coalitional Ability and Group Knowledge

by

Pål Grønås Drange

MSc Thesis (Masteroppgave)

for the degree

Master of Science in Informatics

(Master i informatikk)

Faculty of Mathematics and Natural Sciences
University of Bergen

November, 2011

Det matematisk-naturvitenskapelige fakultet
Universitetet i Bergen

Abstract

In this thesis, we investigate coalitional ability under imperfect knowledge. Coalitional ability is a measure of how powerful a group of agents are, in terms of which outcomes they can achieve in some game. This has been studied using effectivity functions, which has a very natural interpretation in neighbourhood semantics in modal logic. On the other hand, coalitional ability has been studied in great details recently in a variety of papers, but not in the context of effectivity functions and neighbourhood semantics.

A very natural question to ask is "what is a coalition effective for under imperfect knowledge?". In this thesis we study that question and analyse the problem using effectivity functions. The main result is a representation theorem, completely characterising the effectivity functions corresponding to epistemic game structures under a variant of α-effectivity that uses distributed knowledge. Along the way, we also give a new neighbourhood semantics for standard epistemic logic with different variants of group knowledge.

Preface

This is a master thesis in modal (epistemic) logic using game theoretic concepts. It is a monograph, not a text book. This means that I will not try to teach the reader what logic is all about, not even what modal logic is. This means that to really enjoy this thesis, the reader should have a background in logic, with a good understanding of modal logic. Furthermore, I also assume the reader is familiar with the basic concepts, notions, semantics, language and nomenclature of game theory, ranging from strategic games, through solution concepts to cooperative games, meaning I will not spend too much time trying to *teach the reader* what the fuzz of game theory is all about. However, I *do attempt to define* everything I use and refer to, e.g. games (any type of such) and their contents (actions, players, outcome), different types of modal structures; frames and models, how formulae are evaluated, etc. despite the fact that this will probably be known to the reader already.

It is my hope that the reader will enjoy reading this thesis as much as I enjoyed writing it.

Acknowledgement

My greatest thanks go to my family $\{\text{Helge}, \text{Kari}, \text{Lars}, \text{Jonas}\}$[1] $\cup \{\text{Pia}\}$ who has always been there for me, always encouraging me to do what I wanted most. I will always be indebted to you. Thank you.

Academically, there are two persons who I owe everything; My academic thanks go to my supervisor, Professor Thomas Ågotnes, who showed me the beauty of modal logic and sent me to Amsterdam. To say that Erik Parmann [51] has always been there for me, would be a grave understatement. I can with great confidence say that without him, I would never have taken a master's degree in logic (sort of), something that would be a great loss for me. He has for the past five years, always been there, helping, entertaining and challenging me. Thank you.

I cannot express my gratitude to the institute of logic, language and computation at the University of Amsterdam, and the students there for teaching me absolutely everything I know about logic. My absolute deepest thanks go to Professor Yde Venema, Johannes Marti and Matthew Wampler-Doty for teaching me so much. My thanks go to all the students I spent my time with there, but especially I thank Paula, Gabriela and Charlotte for making my entire stay in Amsterdam completely perfect. Thank you.

Finally, thanks to all my closest friends, especially Kenneth, Mats, Espen, Lasse & Elisabeth (you are all very dear to me), the entire algorithms department and last but not least, my very own logic group, Truls, Piotr, Sjur, Yì et Paul Simon. Thank you.

Remark 0.1. To everyone not on this list; I either forgot you, or deliberately left you out, possibly due to space requirements. Please contact me if you are uncertain which one it is. $\dashv$

[1] A natural question to ask is why my first name does not start with the letter i, as that would make our family names iterate over the letters h to l. It seems my parents missed this golden opportunity.

Contents

Part I

Preliminaries

Chapter 1

Introduction

1.1 Reasoning about coalitional ability and imperfect knowledge

Two major fields within artificial intelligence, computer science and economy are the fields that deal with knowledge, and with strategic ability [56]. Analysing knowledge in computers are used to reason about distributed computing and all kinds of cooperation between computers and programs, but reasoning about knowledge also helps us to obtain a deeper understanding of what agents, e.g. robots or computer systems, can achieve under certain conditions. In the other field, known as game theory, we reason about strategic ability by introducing the concept of actions, and the outcome of simultaneously performed actions. Some work has been done fusing these two fields, however the logical analysis of this field is still young. Knowledge in game theory has been studied to a great extent and dates back to at least [43] and [60]. For a more general introduction to this field, see e.g. [49], specifically Extensive games with imperfect information[1].

We will use *epistemic modal logic*, from now on simply epistemic logic, to express knowledge, or more precisely, lack of knowledge for agents, and we will use different logics to express *strategic ability*, also called *effectivity*. There are a few logics for dealing with this fusion, the most known of them being ATEL [65, 30], but this logic has some problems that we will come back to.

Distributed knowledge is one of the most known notions of group knowledge, and in many cases it is by far the most natural one. Distributed knowledge models what a group of agents can come to know if they are allowed to exchange information[2].

[1]We will refer to information as *knowledge*, as it fits better in the framework of epistemic logic. However, the reader can think of (imperfect) knowledge as having (imperfect) information.

[2]There are problems with distributed knowledge that makes that sentence somehow borderline false, if not across the borderline. There are models that actually makes distributed knowledge *stronger* than what agents can come to learn if they are allowed to share information or knowledge. These models are the models that lie outside the class of full communication

Example 1.1. Anna and Bill meet in the elevator, they want to discuss the weather, but none of them knows what the weather is. However, Bill just saw Carlotta, a close friend of Anna, wearing her red coat. Anna knows that *if* Carlotta wears her red coat *then* it is raining. One might argue that Anna and Bill together know that it is raining. $\dashv$

In the example above, it is clear that they might find out that it is raining, given the knowledge they already have. This is the crux of distributed knowledge.

But despite the fact that it is so natural, for different reasons, the study of distributed knowledge has been somewhat neglected, [67] does not even mention it, nor does [65]. It is clear that distributed knowledge is not suited for every task we want group knowledge to model, in the case of strategic abilities, or abilities of what a group of agents can achieve if they are allowed to cooperate, it is certainly the most natural.

For instance, consider the game bughouse chess (a variant of chess where two teams of two plays against each other on two different chess boards). In this game, the teammates are allowed to communicate, and by cooperating and sharing knowledge, they are indeed using distributed knowledge to make decisions. For another example, consider a web server wishing to serve a user a web page. Perhaps this web server knows the wish of the user; she wants to visit the main page, but the web server does not know what the main page should contain, so the web server must contact the database server for this information. Their knowledge combined, the distributed way, is what the user wants to see.

In fact, one can argue that most strategic situations where two parts *cooperate*, if there is knowledge involved, and communication allowed, the *only* natural form of group knowledge is distributed knowledge.

1.2 Motivation

Several logics have been proposed to reason about coalitional ability, amongst them are ATL [10], STIT [15] and CL [53]. The latter is essentially the nexttime fragment of ATL using neighbourhood semantics. There are also several logics that deals with coalitional ability under imperfect knowledge, but no logic deals with the effectivity functions of coalitional abilities under imperfect knowledge. Furthermore, the most known logic for reasoning about coalitional abilities under imperfect knowledge, ATEL, has many drawbacks, which we discuss in Section 2.5.4.

The main contribution of this thesis is to introduce a neighbourhood semantic for distributed knowledge and coalitional ability under imperfect knowledge. The logic itself, called *DCL* (Distributed knowledge Coalition Logic) is a fragment of ATOL[3] and more generally of the constructive strategic logic. Creating

models, and are studied in great details in [55]. One might argue that if one are to model "real world situations", however, we only meet full communication models.

[3]DCL is the nexttime fragment of bounded memory strategies with distributed observational power of ATOL. DCL is contained by a simple translation, which is given in Section 5.2.

a neighbourhood semantic for this logic has been an open task. Marc Pauly gave in [53] a representation theorem, from now on Pauly's representation theorem, for effectivity functions for strategic games, but it has been open to give such a representation theorem for games with imperfect knowledge. This thesis gives a result for distributed knowledge, describing properties an effectivity function must have when distributed knowledge is assumed.

1.3 Overview of the thesis

In Chapter 2 we introduce modal logic with both its language and several different semantic systems for modal logic, amongst them are the *relational systems* and the *set systems* or *neighbourhood systems*. We take a look at some applications of modal logic, especially epistemic logic. We proceed to introduce game theory, a rather young mathematical field widely used to analyse multi-agent situations before we introduce coalition logic, a logic designed to talk about power and ability in multi-agent systems. We will in the latter section discuss a result linking relational systems and set systems.

In Chapter 3, we take a closer look on one specific form of *group knowledge* in epistemology, namely *distributed knowledge*, and we investigate how distributed knowledge behaves in set systems, or neighbourhood systems, a line of research that has been widely neglected. We give a representation theorem linking relational systems and neighbourhood systems by giving some simple restrictions on the set system. We will also discuss the completeness of the language $\mathcal{L}_{\mathcal{D}}$ with respect to the just mentioned restricted class of neighbourhood systems.

To generalize the results in the previous mentioned chapter, we introduce in Chapter 4 a logic with all three most used forms of group knowledge in a neighbourhood semantic setting, and we give a representation theorem, and thereby getting soundness and completeness for a specific class of neighbourhood models, with respect to the language S5 with mutual knowledge, common knowledge and distributed knowledge.

After having spent some time with epistemic logic, we will in Chapter 5 discuss several logics that relate strategic ability and knowledge, e.g. ATEL, ATOL, STIT and their relationship to coalition logic.

We when combine knowledge with effectivity functions, something that seems to have been overlooked in this field, and we give a representation theorem linking a specific relational logic with knowledge, and coalition logic with knowledge. We do a thorough analysis of distributed knowledge in coalition logic and give our representation theorem and see how Coalition Logic behaves in contrast to how Distributed knowledge Coalition logic behaves. All this in Chapter 6.

Finally we will in Chapter 7, review our results and discuss possible extensions and natural ways to proceed next.

Chapter 2

Background

In this section, we go into details about the aforementioned tools and fields that are relevant for this thesis. We will see what modal logic is, its advantages and disadvantages, different semantic systems and several “epistemic modal logics”. Furthermore, we will take a small detour into the field of game theory to get an introduction to the chapters that will cover modal logics for strategic abilities. Finally, we will recall Pauly’s representation theorem [53] for strategic games and discuss the theorem of truly playable games [31].

2.1 Notation and concepts

When discussing multi-agent systems, we need to be able to refer to the set of agents we are discussing. This set of agents will be denoted by $\mathcal{A}$, and is always assumed to be finite and non-empty. We will also sometimes assume that $\mathcal{A}$ is on the form $\{1, 2, \ldots, n\}$ for some $n \in \omega$. This is not required, but it makes some proofs simpler.

When we are using a logic, we also need to be able to refer to propositional letters. The set of propositional letters takes the names $\textsc{Prop}$, P, Π and Φ in the literature. In this thesis we will stick to Φ. Furthermore, it is assumed that Φ is always at most countable. This, combined with the fact that the agent set is finite, makes all our languages countable. We will always denote elements of Φ by p, q, r, p', p_1, etc.

In neighbourhood semantics (hypergraphs, set systems or coalitional games), we will refer to *neighbourhood functions.* These functions will always be denoted by the letter E, for *effectivity*, or the Greek letter ν, for *neighbourhood.*

We will use some main structures, and the most used are the relational modal models, referred to as $\mathcal{M}$, the set system version, being the neighbourhood modal models, referred to as $\mathcal{N}$ and strategic games, referred to as $\mathcal{G}$. We briefly mention algebras, and they are denoted by $\mathfrak{A}$.

If we discuss basic set theoretic concepts and functions that not directly touch on the structures we are discussing, we will use $\mathcal{U}$ to denote some arbitrary set, read as “universe”.

Whenever we have a structure $\mathbb{A}$ of any kind that has a state space (sometimes referred to as a domain, carrier, or a set of *possible worlds*, in the literature), we will denote by $S(\mathbb{A})$ the state space of the structure. In general we will mainly use this for relational and neighbourhood models or frames.

2.2 Tools for sets

Definition 2.1 (Powerset)**.** Given a set $\mathcal{U}$, we will denote by $2^{\mathcal{U}} = \{X \subseteq \mathcal{U}\}$ the *powerset* of $\mathcal{U}$, which is the set of all subsets of $\mathcal{U}$. $\dashv$

It might be of value to note that if $\mathcal{U}$ is finite, $2^{\mathcal{U}}$ is finite, and if $\mathcal{U}$ is infinite, $2^{\mathcal{U}}$ is strictly greater than $\mathcal{U}$, specifically if $\mathcal{U}$ is countable, $2^{\mathcal{U}}$ is uncountable. This is the main reason we want our set of agents, $\mathcal{A}$ to be finite; it keeps the number of coalitions finite, and all logic languages with $|2^{\mathcal{A}}|$ many modalities countable.

Theorem 2.2 (Cantor)**.** *If $\mathcal{U}$ is any infinite set, then $|2^{\mathcal{U}}| > |\mathcal{U}|$.*

Proof. First we prove that $|2^{\mathcal{U}}| \geq |\mathcal{U}|$. So to see that $2^{\mathcal{U}}$ is at least as big as $\mathcal{U}$, consider the injective function $\cdot : u \mapsto \{u\}$. Assume now, for the sake of contradiction, that the lemma is false. Then it must be the case that $|2^{\mathcal{U}}| = |\mathcal{U}|$. This means that there is a bijective function $\varphi : \mathcal{U} \to 2^{\mathcal{U}}$. Consider then the set $X = \{u \in \mathcal{U} \mid u \notin \varphi(u)\}$. Since $X \in 2^{\mathcal{U}}$ and φ is onto $2^{\mathcal{U}}$, there is an element $x \in \mathcal{U}$ such that $\varphi(x) = X$. To reach the contradiction we ask us is $x \in X$? Assume $x \in X$. Then it must be the case that $x \in \varphi(x)$, hence $x \notin X$. So assume $x \notin X$. Then $x \notin \varphi(x)$, but this means that $x \in X$. So it cannot be the case that $x \in X$, nor that $x \notin X$, which is a contradiction. We therefore conclude that there is no such bijective function, and hence that $|2^{\mathcal{U}}| > |\mathcal{U}|$. $\square$

When using epistemic logic and in general discuss knowledge (or lack thereof), and in modal logics in general, we will need to discuss *relations*. It is assumed that the reader is familiar with the basic concepts of relations. However, throughout this thesis, we will often need to discuss a family of relations, one relation for each agent $i \in \mathcal{A}$. We will follow the convention of [49] and write $(R_i)_{i\in\mathcal{A}}$ to denote a *profile* or *family* of relations. An *equivalence relation* is a relation which is reflexive, symmetric and transitive (it is assumed the reader is familiar with these concepts, if not, consult [16, 35]). When using equivalence relations, we will often denote the relation by the symbol $\sim$ instead of R, and when a relation denoted by $\sim$ is given, it can be assumed it is an equivalence relation. Sometimes, given a family of equivalence relations $(\sim_i)_{i\in\mathcal{A}}$, we will need to discuss a “lifted” version, where we get relations for coalitions, i.e., given a coalition $A \subseteq \mathcal{A}$, we will need to refer to $\sim_A$ as a relation that *lifts* the subprofile $(\sim_i)_{i\in A}$. We will assume we either know how to lift them, or when needed, define how to do so explicitly.

Definition 2.3 (Upset)**.** If $\mathcal{U}$ is any set and $S \subseteq \mathcal{U}$, we write $S\uparrow^{\mathcal{U}}$ to be the set $\{T \subseteq \mathcal{U} \mid S \subseteq T\}$. $\dashv$

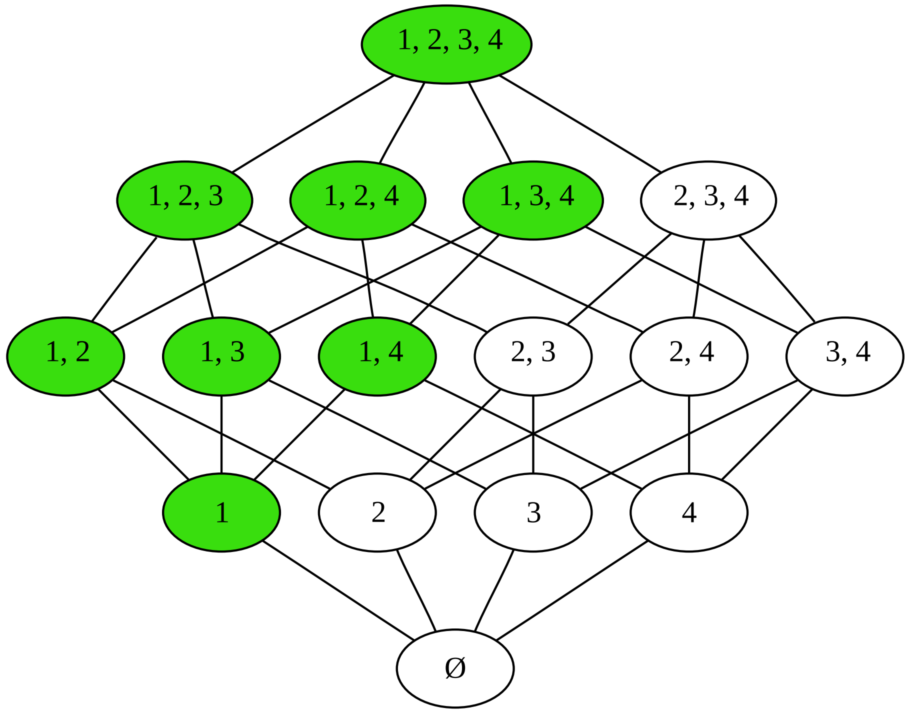


Figure 2.1: A principal filter on $2^{\{1,2,3,4\}}$ generated by $\{1\}$, in fact it is the upper set $\{1\}\uparrow$.

We call this the upper set, or simply the upset of S. The upwards arrow notation is parametrised on some universe, and is usually omitted when clear from context, e.g. when used on some set of states in a model, the uparrow is always assumed to be parametrised on the full set of states in the model. Furthermore, we consider the uparrow to have the least possible operator precedence, meaning that $X \cup Y \uparrow$ is to be read as $(X \cup Y)\uparrow$.

Definition 2.4 (Filter)**.** If $\mathcal{U}$ is a set, then $\mathcal{F} \subseteq 2^{\mathcal{U}}$ is a *filter* on $2^{\mathcal{U}}$ provided that

- $\mathcal{F} \neq \emptyset$,
- for every x and y in $\mathcal{F}$, $x \cap y \in \mathcal{F}$ ($\mathcal{F}$ is closed under binary intersection) and
- for every $x \in \mathcal{F}$ and $y \in 2^{\mathcal{U}}$, $x \subseteq y$ implies that $y \in \mathcal{F}$ ($\mathcal{F}$ is an upper set, or upwards closed).

We furthermore say that a filter is *proper* if it is not equal to the whole of $2^{\mathcal{U}}$. A filter is *principal* if it contains a least element, and is *free* or non-principal if it does not. Equivalently, a filter $\mathcal{F}$ is free if and only if $\bigcap \mathcal{F} = \emptyset$. We say that a (principal) filter $\mathcal{F}$ is *generated by a set* G if and only if $\mathcal{F} = G\uparrow$. $\dashv$

Definition 2.5 (Nonmonotonic core)**.** Let $\mathcal{F} \subseteq 2^{\mathcal{U}}$ be a family of sets. We define the *nonmonotonic core* of $\mathcal{F}$, written $\mathcal{F}^{nc}$ to be

$$\mathcal{F}^{nc} = \{X \in \mathcal{F} \mid \forall Y \in \mathcal{F}.Y \subseteq X \Rightarrow Y = X\} = \{X \in \mathcal{F} \mid \neg\exists Y \in \mathcal{F}.Y \subset X\},$$

that is to say that $\mathcal{F}^{nc}$ is the collection of the *inclusion minimal* elements of $\mathcal{F}$. We say that the *nonmonotonic core* of $\mathcal{F}$ is *complete* whenever for all $X \in \mathcal{F}$ there is a $Y \in \mathcal{F}^{nc}$ such that $Y \subseteq X$. $\dashv$

Note that it is not always the case that the nonmonotonic core is complete for a family of sets. If there are finitely many sets, or all sets in the family are finite, it is the case, but there are families with an empty nonmonotonic core, as the following example shows.

Example 2.6 (Fréchet filter)**.** Let $\mathcal{F} = \{X \subseteq \omega \mid X \text{ is cofinite}\}$. First, we show that $\mathcal{F}$ is a filter.

- Obviously, $\mathcal{F} \neq \emptyset$ since there indeed are cofinite sets.
- Second, if X and Y are cofinite, then $X \cap Y$ is cofinite. Why? Since $\overline{X}$ and $\overline{Y}$ are finite, we have that $X \cap Y = \overline{\overline{X} \cup \overline{Y}}$ is the complement of the union of two finite sets, which is the complement of a finite set.
- Third, $\mathcal{F}$ is closed under supersets. If $X \in \mathcal{F}$ and $Y \supseteq X$, this means that since there are only finitely many elements not in X, and every element in X is in Y, only finitely many elements are not in Y.

But we have that $\mathcal{F}^{nc} = \emptyset$, for assume otherwise, that $X \in \mathcal{F}^{nc}$. Since $\emptyset$ is not cofinite, let $x \in X$. Now, $X \setminus \{x\}$ is cofinite, so it is in $\mathcal{F}$, but it is also a strict subset of X, hence $X \notin \mathcal{F}^{nc}$ \`. We conclude that $\mathcal{F}^{nc}$ is not complete. $\dashv$

Lemma 2.7 (Folklore)**.** *Let $\mathcal{F} \subseteq 2^{\mathcal{U}}$ be a set of sets. If $\mathcal{F}$ is closed under binary intersection, then $\mathcal{F}$ is closed under finite intersection.*

Before we prove this, we state the lemma more precise: If for any $X = \{X_1, X_2\}$ with $X_1 \in \mathcal{F}$ and $X_2 \in \mathcal{F}$, $\bigcap_{i=1,2} X_i \in \mathcal{F}$, then it holds for any $Y = \{Y_1, \ldots, Y_k\}$ that if every $Y_i \in \mathcal{F}$ for $i \leq k$ that $\bigcap_{i \leq k} Y_i \in \mathcal{F}$.

Proof. Let $\mathcal{F}$ be a binary intersection closed family of sets. We prove this by induction on the size of the family we intersect. For the base case, $n = 2$, this holds by the fact that $\mathcal{F}$ is closed under binary intersection. The induction hypothesis is that if $\mathcal{F}$ is closed under intersection of families of size up to n, then it is closed under intersection of families of size up to $n + 1$. So let $X = \{X_1, \ldots, X_n\}$ be a family of sets of size n, and X_{n+1} be any set, such that $X_i \in \mathcal{F}$ for $i \leq n + 1$. By induction hypothesis, $\bigcap_{i \leq n} X_i \in \mathcal{F}$, and by binary intersection, $(\bigcap_{i \leq n} X_i \cap X_{n+1}) \in \mathcal{F}$. Hence for any family of sets $Y = \{Y_1, \ldots, Y_{n+1}\}$, if each $Y_i \in \mathcal{F}$ for $i \leq n + 1$, then $\bigcap_{i \leq n+1} Y_i \in \mathcal{F}$. $\square$

2.2.1 Effectivity functions

One of the main concepts in this thesis is that of an effectivity function. An effectivity function takes many forms, the simplest form being, given a universe $\mathcal{U}$, $E : \mathcal{U} \to 2^{2^{\mathcal{U}}}$, that is, E maps elements of $\mathcal{U}$ to sets of sets of $\mathcal{U}$. However, we often need to have one effectivity function per agent, or even per coalition, in which the effectivity function takes the form $E : \mathcal{U} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{U}}}$. When the latter is the case, we sometimes refer to E as a family of effectivity functions, $(E_u)_{u \in \mathcal{U}}$, where each of them is on the form $E_u : 2^{\mathcal{A}} \to 2^{2^{\mathcal{U}}}$.

Remark 2.8 (Effectivity functions and neighbourhood functions)*.* In Section 2.3.3, we will introduce *neighbourhood functions*, and the reader should then observe that a neighbourhood function and an effectivity function is two names for the same thing. The neighbourhood functions will be denoted ν (for neighbourhood) and will only be used when discussing pure neighbourhood frames and models. When we are talking about effectivity, in the sense of what agents and coalitions can achieve, we will use effectivity functions, which are denoted by E. ⊣

Definition 2.9 (Effectivity function)**.** Given a universe $\mathcal{U}$, an effectivity function is a function $E : \mathcal{U} \to 2^{2^{\mathcal{U}}}$. Furthermore, given a set of agents $\mathcal{A}$, we call $E : \mathcal{U} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{U}}}$ a coalitional effectivity function. If a function is on the form $E : 2^{\mathcal{A}} \to 2^{2^{\mathcal{U}}}$, we will call E a *simple* coalitional effectivity function. Finally, if $E : \mathcal{U} \times \mathcal{A} \to 2^{2^{\mathcal{U}}}$, we call E a single agent effectivity function. ⊣

Remark 2.10*.* We will in most of our analysis use coalitional effectivity functions, but for simplicity, we will refer to them simply as effectivity functions, when it is clear from the context that we are discussing coalitional effectivity functions. Observe that a coalitional effectivity function is a mapping from a universe $\mathcal{U}$ to simple coalitional effectivity functions. ⊣

Furthermore, when analysing different structures containing effectivity functions, we will also need to talk about how these effectivity functions look, or how a family of effectivity functions look and (co-)behave. We say e.g. that $E : \mathcal{U} \to 2^{2^{\mathcal{U}}}$ is a filter if for all $u \in \mathcal{U}$, $E(u)$ is a filter. Below follows a short list of different properties such effectivity functions can have. It is assumed that the universe now is a non-empty set $\mathcal{S}$, so we discuss families of effectivity functions of the type $E_s : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ for $s \in \mathcal{S}$.

Liveness We say that E is *alive*, or has the liveness property provided for every $A \subseteq \mathcal{A}$ and every $s \in \mathcal{S}$, $\emptyset \notin E_s(A)$.

Safety E is *safe* if for every $A \subseteq \mathcal{A}$ and every $s \in \mathcal{S}$ we have $\mathcal{S} \in E_s(A)$.

Superadditivity Let A and B be disjoint coalitions, i.e. $A \cap B = \emptyset$ and let $X \in E_s(A)$ and $Y \in E_s(B)$. E_s is *superadditive* provided $X \cap Y \in E_s(A \cup B)$. E is superadditive when for all $t \in \mathcal{S}$, E_t is superadditive.

Outcome monotonicity E is *outcome monotonic* provided that for all A and all s, if $X \in E_s(A)$ for some $X \subseteq \mathcal{S}$, then for every $Y \subseteq \mathcal{S}$, such that $X \subseteq Y$ we have $Y \in E_s(A)$.

Coalition monotonicity E is *coalition monotonic* provided that for all A and all s, if $X \in E_s(A)$ for some $X \subseteq \mathcal{S}$, then for every $B \subseteq \mathcal{A}$, such that $A \subseteq B$, we have $X \in E_s(B)$.

Regularity For all X, all A, if $X \in E_s(G)$, then $\overline{X} \notin E_s(\overline{A})$. We say that E is A-regular if it holds for a specific coalition, especially, we say that E is $\mathcal{A}$-regular if it holds for the grand coalition.

Maximality $\overline{X} \notin E_s(\overline{A})$ implies $X \in E_s(A)$. We say that E is A-maximal if it holds for a specific coalition, especially, we say that E is $\mathcal{A}$-maximal if it holds for the grand coalition.

Completeness of nonmonotonic core for $\emptyset$ E has a *complete nonmonotonic core* for $\emptyset$ if for all $s \in \mathcal{S}$ $E_s^{nc}(\emptyset)$ is complete. Furthermore, if $E_s(\emptyset)$ is outcome monotonic, $E_s(\emptyset)$ will be a principal filter over $\mathcal{S}$.[1]

Respecting knowledge E respects $(\sim_A)_{A \subseteq \mathcal{A}}$ if for all $A \subseteq \mathcal{A}$ and $s \sim_A t$, $E_s(A) = E_t(A)$.

A simple corollary follows from Lemma 2.7.

Corollary 2.11 (Finite superadditivity)**.** *Let E be an effectivity function, $A \subseteq \mathcal{A}$ and a partitioning of A into disjoint sets $A_1, \ldots, A_k$. Let $(X_i)_{i \leq k}$ be a family of outcome sets, such that $X_i \in E_s(A_i)$ for $i \leq k$. Then it holds that $\bigcap_{i \leq k} X_i \in E_s(A)$.*

Lemma 2.12. *Superadditivity, safety and liveness is sufficient for (i) coalition-monotonicity and (ii) regularity.*

Proof. (i) By safety, $E(A)$ is non-empty, so assume $X \in E(A)$ and that $A \subseteq B$. Let $A' = B \setminus A$, then by superadditivity we obtain that $X \in E(A \cup A') = E(B)$. (ii) Assume again $X \in E(A)$ and, for the sake of a contradiction, that $\overline{X} \in E(\overline{A})$. By superadditivity this gives that $\emptyset \in E(\mathcal{A})$ contradicting the liveness property. □

Before ending this section, we define what a *playable* and a *truly playable* effectivity function is. It should be noted that there is no good reason to distinguish between the two kinds. It should always be assumed we need truly playable effectivity functions. The reason we have two, is that in the papers they were introduced [52, 53], there was an error in a proof which was corrected in [31]; The error was a missing property that is needed when considering infinite games.

We will recall Pauly's representation theorem of Pauly and Goranko, Jamroga and Turrini below, and to do this, we need the following two concepts.

[1] Proof: $E_s(\emptyset)$ is closed under finite intersection by superadditivity and the fact that $\emptyset \cap \emptyset = \emptyset$, or the empty coalition is disjoint with itself.

Definition 2.13 (Playable effectivity function)**.** A simple coalitional effectivity function $E : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ for $\mathcal{A}$ over $\mathcal{S}$ is called a *playable effectivity function* if E is live, safe, $\mathcal{A}$-maximal, outcome monotonic and superadditive. $\dashv$

Corollary 2.14 (Pauly)**.** *Every playable effectivity function is regular and coalition monotonic.*

This is immediate from Lemma 2.12.

Definition 2.15 (Truly playable effectivity function)**.** A playable effectivity function E is called *truly playable* if the nonmonotonic core, $E^{nc}(\emptyset)$, of $E(\emptyset)$ is complete. $\dashv$

2.3 Modal logic

It is assumed that the reader is somewhat familiar with what a logic is and the distinction between a syntactic proof system and a semantic system, the two most known are the proposition logic and the first order logic.

2.3.1 Language

The language of modal logics is very simple. We will start by explaining the basic modal logic before introducing the general *similarity type*. The standard basic modal logic is the language $\mathcal{L}$ based on an (at most countable) set of propositions, Φ, and is generated by the grammar

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi,$$

where $p \in \Phi$, the box $\Box$ is a *modality*, and $\Box\varphi$ is read as "it's necessary that φ", "always φ", "it's known that φ", "it's provable that φ", and many more. Its *dual* is the *diamond* modality, $\Diamond\varphi$, and it is dual in the sense that $\Box\varphi \equiv \neg\Diamond\neg\varphi$. The diamond version is read as "it's possible that φ", "sometimes φ", "it's considered[2] that φ". Compare the duality between $\Box\varphi$ and $\Diamond\varphi$ to the standard duality between $\forall x\varphi$ and $\exists x\varphi$, i.e. $\forall x\varphi \equiv \neg\exists x\neg\varphi$.

However, we often want to to extend the language into a more complex modal logic, where we can have several modalities, and indeed, each modality can bind more than one formula. An example can be $\nabla(\varphi, \psi)$ where ∇ is a binary modality.

Definition 2.16 (Modal similarity type [16])**.** A modal similarity type is a pair $\tau = (O, \rho)$, where O is a nonempty set, and ρ is a function $\rho : O \to \omega$. The elements of O are called *modalities* or *modal operators*, and ρ is the *arity function*, i.e. if $\nabla \in O$ with $\rho(\nabla) = n$, then $\nabla(\varphi_1, \ldots, \varphi_n)$ is a well-formed formula when φ_i is for all $1 \leq i \leq n$. $\dashv$

[2]The best reading here is probably to read the dual as the definition, namely that "it is not known that φ is not the case"

Remark 2.17. In this thesis we will mainly work on modalities involving agents, and coalitions of agents. If $\mathcal{A}$ is a finite set of agents, our similarity types will be of the form $O_a = \bigcup_{i \in \mathcal{A}} \nabla_i$ and $\rho_a(\nabla) = 1$ for all $\nabla \in O_a$ or they will be of the type $O_A = \bigcup_{A \subseteq \mathcal{A}} \nabla_A$ with $\rho_A(\nabla) = 1$ for all $\nabla \in O_A$. $\dashv$

We will therefore in most cases omit references to similarity types and simply state that we have one modality for each coalition or agent.

2.3.2 Relational semantic – Kripke

The, without any competition, most used semantics for modal logics, is the relational semantics. This is also known as Kripke semantics and possible world semantics, named in honour of Saul Kripke after his work [42]. In this paper, we will refer to these semantics and models as relational semantics and relational models.

Definition 2.18 (Relational model)**.** A relational model for a modal similarity type $\tau = (O, \rho)$ is a structure $\mathcal{M} = (\mathcal{S}, (R_i)_{i \in O}, V)$ where $\mathcal{S}$ is called the *state space*, each R_i is an $1 + \rho(i)$-ary relation on $\mathcal{S}^{1+\rho(i)}$ and $V : \Phi \to 2^{\mathcal{S}}$ is the assignment function or *valuation function*. $\dashv$

The R_is make it apparent why this is called a relational model. When each R_i has arity 2 (each modality has arity 1), $\mathcal{M}$ is simply a vertex-labelled (V labels each vertex in $\mathcal{S}$) graph with several edge sets (each R_i is simply an edge set).

We refer to elements of $\mathcal{S}$ as "states" or "points". Other notions in the literature are "worlds", "vertices" or "nodes" and even "configurations". If $s \in \mathcal{S}$, we call the pair $(\mathcal{M}, s)$ a *pointed model*, and if $\mathcal{L}$ is a modal language with a similarity type matching $\mathcal{M}$'s signature, we define the truth of $\varphi \in \mathcal{L}$ in pointed models as follows:

$$
\begin{array}{ll}
(\mathcal{M}, s) \vDash_{\mathcal{RS}} p & \text{if and only if } s \in V(p) \\
(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\varphi & \text{if and only if not } (\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \\
(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \wedge \psi & \text{if and only if both } (\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \text{ and } (\mathcal{M}, s) \vDash_{\mathcal{RS}} \psi \\
(\mathcal{M}, s) \vDash_{\mathcal{RS}} \nabla(\varphi_1, \ldots, \varphi_n) & \text{if and only if for all } t_1, \ldots, t_n \in \mathcal{S} \\
 & \text{such that } R_i(s, t_1, \ldots, t_n),\ (\mathcal{M}, t_i) \vDash_{\mathcal{RS}} \varphi_i
\end{array}
$$

We will as remarked in this thesis only use an extended version of the basic similarity type, and then the last scheme above is exchanged for

$$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \nabla_A(\varphi) \text{ if and only if for all } t \in \mathcal{S} \text{ such that } R_A(s, t)\ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$$

Definition 2.19 (Truth set)**.** We define the *truth set* of a formula $\varphi \in \mathcal{L}$ in a model $\mathcal{M} = (\mathcal{S}, (R_i)_{i \in O}, V)$ as $[\![\varphi]\!]^{\mathcal{M}} = \{s \in \mathcal{S} \mid (\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi\}$, i.e. all states satisfying φ. $\dashv$

2.3.3 Neighbourhood semantic – Scott-Montague

In this section, we consider only the basic similarity type, i.e. the modal language $\mathcal{L}$ with only one single unary modality, $\Box$. Where the relational semantics was based on a relational model, i.e. a graph, the neighbourhood semantics is based on a generalised model of the relational, namely a hypergraph. Instead of having a vertex-labelled graph with several edgesets, we have a vertex-labelled hypergraph with several hyper-edgesets. The structure

$$\mathcal{N} = (\mathcal{S}, \nu, V)$$

is called a neighbourhood model with $\nu : \mathcal{S} \to 2^{2^{\mathcal{S}}}$, or sometimes in the literature a Scott-Montague model (after [47] and [58] who discussed them, seemingly independent [50]). Observe that ν is an effectivity function (recall Remark 2.8). A *pointed neighbourhood model* is a pair $(\mathcal{N}, s)$ where $s \in S(\mathcal{N})$. In the basic modal language $\mathcal{L}$, a formulae $\varphi \in \mathcal{L}$ is evaluated in pointed models as follows:

$$
\begin{array}{ll}
(\mathcal{N}, s) \vDash_{\mathcal{NS}} p & \text{if and only if } s \in V(p) \\
(\mathcal{N}, s) \vDash_{\mathcal{NS}} \neg\varphi & \text{if and only if not } (\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \\
(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \wedge \psi & \text{if and only if both } (\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \text{ and } (\mathcal{N}, s) \vDash_{\mathcal{NS}} \psi \\
(\mathcal{N}, s) \vDash_{\mathcal{NS}} \Box\varphi & \text{if and only if } [\![\varphi]\!]^{\mathcal{N}} \in \nu(s)
\end{array}
$$

where $[\![\varphi]\!]^{\mathcal{N}} = \{s \in S(\mathcal{N}) \mid (\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi\}$ is as in Definition 2.19, except for the $\vDash_{\mathcal{NS}}$-relation, meaning that we will use the same notation for both types of truth, and it should cause no confusion as the truth definition is implicit in the type of model we use.

Remark 2.20 (Strict versus weak neighbourhood semantics). The last point of the definition of truth makes this a *strict* semantical definition [11], in that the truth set of φ has to be a member of the effectivity function. The *weak* semantics is to demand the existence of a set X in the effectivity function such that $[\![\varphi]\!]^{\mathcal{N}} \subseteq X$. We will not go into details about this distinction in this thesis, but take the semantic definition to always be strict. For more on this distinction, see [50]. See also [62] for usages of the weak form. $\dashv$

It can be shown that neighbourhood semantics is a strict generalisation of the relational semantics. To see that neighbourhood semantics is at least as expressive as simple relational semantics, we observe the following trivial translation of frames:

Proposition 2.21. *If $\mathcal{M} = (\mathcal{S}, R, V)$ is a relational frame, then $\mathcal{N} = (\mathcal{S}, \nu_R, V)$ is a neighbourhood frame with the property that for all $\varphi \in \mathcal{L}$, and for all $s \in \mathcal{S}$, $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$ if and only if $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$, where*

$$\nu_R(s) = \{X \subseteq \mathcal{S} \mid R[s] \subseteq X\} = R[s]\uparrow .$$

However, the opposite direction does not hold, which we can simply see by observing that $\Box(p \to q) \to (\Box p \to \Box q)$ is *not* valid on neighbourhood frames[3]. But there is a restriction of the neighbourhood frames that corresponds to the relational frames:

Proposition 2.22. *If $\mathcal{N} = (\mathcal{S}, \nu, V)$ is a neighbourhood model, and for all $s \in \mathcal{S}$ we have that $\nu(s)$ is a generated filter by a single set, then there is a relational model $\mathcal{M} = (\mathcal{S}, R_\nu, V)$ such that for all $\varphi \in \mathcal{L}$, and for all $s \in \mathcal{S}$, $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$ if and only if $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$.*

Proof. This is proven in [50]. He shows that if ν is augmented, meaning a filter containing its core, then it has an equivalent relational model. $\Box$

2.3.4 Algebraic semantic

For completeness, we will very briefly mention the last of the three most used semantics for modal logic, namely *algebraic semantic*. The algebraic semantic has many nice features over the two above, one of them being its completeness for classes of logics where the relational semantics fails to be complete. For a more thorough introduction, see [16, 44, 45, 23].

Definition 2.23 (Modal algebra)**.** A structure $\mathfrak{A} = (A, \wedge, \vee, \top, \neg, \Box)$ is a *modal algebra* if A is closed under the unary operation $\Box$ and

- $(A, \wedge, \vee, \top, \neg)$ is a Boolean algebra
- for x and y in A, $\Box(x \vee y) = \Box x \vee \Box y$

$\dashv$

We also define $\Diamond x = \neg\Box\neg x$.

Definition 2.24 (Algebraic assignment)**.** An algebraic assignment for a modal algebra $\mathfrak{A} = (A, \wedge, \vee, \top, \neg, \Box)$ is a function $\theta : \Phi \to A$, and we define $\tilde{\theta} : \mathcal{L} \to A$ in the obvious way. $\dashv$

We take $\varphi \approx \psi$ if for all θ, $\tilde{\theta}(\varphi) = \tilde{\theta}(\psi)$ and we finally write $\vDash_{\mathcal{AS}} \varphi$ if $\varphi \approx \top$.

2.4 Epistemic logic

The by far most popular and widely accepted logic for epistemology is the logic S5. This logic will be assumed throughout the entire thesis without mentioning it, but we note that several of the results also work with other logic systems as background. A reader who is completely non-familiar with epistemic logic in a modal logical setting is encouraged to consult any one of [26, 37, 67].

[3] In neighbourhood models, the following rule is sound: If $\varphi \leftrightarrow \psi$ is valid, then $\Box\varphi \leftrightarrow \Box\psi$ is valid.

2.4.1 Motivation

Epistemic logic is a widely used tool to reason about knowledge, both in philosophy and in computer science. The earliest epistemic logics were single agent logics, and were basic modal logics with the only modality K and its dual $\widetilde{\mathrm{K}}$, read "it is known" and "it is considered", respectively. If p is a proposition, Kp is read "it is known that p". Many different axiomatizations have been suggested, and there are still several different axiomatizations in use to reason about knowledge, however, the most used is S5 (defined below). In this axiomatization, we can from Kp deduce that KKp, meaning "if it is known that p, then it is known that it is known that p". Another axiom to assume it the *truth axiom*, $\mathrm{K}p \to p$, meaning that if p is known, then p must indeed be the case, etc. In this thesis, however, we discuss multi-agent systems, and it is then natural to introduce one modality per agent; We write $\mathrm{K}_i p$ to mean "agent i knows p".

2.4.2 S5

In relational frames, we put no restriction on how the relations actually look. But by looking at only those frames that validates some axioms, we essentially do put restrictions on the relations. The single most used logic for knowledge is the system S5, or the logic where each relation is an equivalence relation. In case a relation is an equivalence relation, we follow the convention and use the symbol $\sim$ instead of R. The basic modal language $\mathcal{L}$ based on $\mathcal{A}$ and Φ will thus be the following

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i \varphi,$$

where $p \in \Phi$ and $i \in \mathcal{A}$. The following axiomatization is sound and complete with respect to relational frames with n equivalence relations:

All instantiations of propositional tautologies	
$\Box_i(p \to q) \to (\Box_i p \to \Box_i q)$	K
$\Box_i p \to p$	T
$\Box_i p \to \Box_i \Box_i p$	4
$p \to \Box_i \Diamond_i p$	5
From φ and $\varphi \to \psi$ infer ψ	Modus ponens
From φ infer $\Box\varphi$	Necessitation
From φ infer $\varphi[\psi/\chi]$	Uniform substitution

2.4.3 Knowledge: The individual case

We will in this section introduce language for individual knowledge, the semantics, and a sound and complete axiomatization.

Language

We continue our assumption of a fixed finite set of agents $\mathcal{A}$ and a countable set Φ of proposition letters.

Definition 2.25 (Language $\mathcal{L}_{\mathcal{K}}$)**.** The language $\mathcal{L}_{\mathcal{K}}$ for $\mathcal{A}$ and Φ is defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathrm{K}_i\varphi,$$

where $p \in \Phi$ and $i \in \mathcal{A}$. $\dashv$

This means that we have $|\mathcal{A}|$ many unary modalities. Examples of formulae from $\mathcal{L}_{\mathcal{K}}$ are $p \wedge \mathrm{K}_a p \wedge \mathrm{K}_a \neg \mathrm{K}_b p$ (when a and b are elements of $\mathcal{A}$ and $p \in \Phi$) and $\mathrm{K}_a \neg \mathrm{K}_a p$. Furthermore, we write $\widetilde{\mathrm{K}}_a$ for the dual, i.e. $\widetilde{\mathrm{K}}_a = \neg \mathrm{K}_a \neg$. $\widetilde{\mathrm{K}}_a p$ can be read as "a considers p possible". Other examples of formulae are $\mathrm{K}_a\mathrm{K}_a\mathrm{K}_a p \wedge \neg \mathrm{K}_a\mathrm{K}_a\mathrm{K}_a\mathrm{K}_a p$ and $p \wedge \mathrm{K}_a \neg p$. The latter sentence is in most epistemic logics[4] considered as "not true". We will see in the next section that it is indeed the case that we consider $p \wedge \mathrm{K}_a \neg p$ to be absurd.

Semantics

Usually, the logic $\mathcal{L}_{\mathcal{K}}$ is interpreted in relational models (see Section 2.3.2), but we will give brief semantic definitions for both relational models and neighbourhood models.

If $\mathcal{A}$ is our usual finite set of agents, then $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$ is a relational S5 model for $\mathcal{A}$ provided that for each $i \in \mathcal{A}$, $\sim_i$ is an equivalence relation over $\mathcal{S} \times \mathcal{S}$. If $s \in \mathcal{S}$, we call the pair $(\mathcal{M}, s)$ a pointed model, and if φ and ψ are formulae of $\mathcal{L}_{\mathcal{K}}$ and $i \in \mathcal{A}$, we evaluate the formulae in pointed models by the following rules:

$$\begin{array}{ll}
(\mathcal{M}, s) \vDash_{\mathcal{RS}} p & \text{if and only if } s \in V(p) \\
(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\varphi & \text{if and only if not } (\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \\
(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \wedge \psi & \text{if and only if } (\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \text{ and } (\mathcal{M}, s) \vDash_{\mathcal{RS}} \psi \\
(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{K}_i\varphi & \text{if and only if for all } t \sim_i s,\ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi
\end{array}$$

That last sentence can be read that $\mathrm{K}_i\varphi$ if φ is true in any *world agent i considers possible* and fits well with our intuition about knowledge and possible worlds.

We will now see how the neighbourhood models look. So suppose $\mathcal{A}$ is again our standard finite set of agents and Φ the set of proposition letters. If $\mathcal{S}$ is any set of states and $V : \Phi \to 2^{\mathcal{S}}$, we call $\mathcal{N} = (\mathcal{S}, (\nu_s)_{s\in\mathcal{S}}, V)$, an *individual epistemic neighbourhood model* if for all $s \in \mathcal{S}$, $\nu_s : \mathcal{A} \to 2^{2^{\mathcal{S}}}$ is a single agent simple effectivity function with the property that for all $i \in \mathcal{A}$, the following hold: Denote by $Z^i_s = \bigcap \nu_s(i)$,

- $s \in Z^i_s$,
- $t \in Z^i_s$ implies $s \in Z^i_t$ and

[4] Epistemic logics is about knowledge, however, in doxastic logics, the logics about beliefs (from ancient Greek for belief, $\delta o\xi\alpha$), it is considered plausible that an agent can believe false propositions.

- $\nu_s(i) = Z^i_s\uparrow$.

This restrictions on the effectivity functions make sure that $\nu_s(i)$ simply is the set of all upper sets of some equivalence class around s, or in other words, $\nu_s(i)$ is the filter generated by a set containing s.

Now, given such an individual epistemic neighbourhood model, with a point $s \in \mathcal{S}$, we call $(\mathcal{N}, s)$ a pointed individual epistemic neighbourhood model, and evaluate formulae as follows:

$$
\begin{array}{ll}
(\mathcal{N}, s) \vDash_{\mathcal{NS}} p & \text{if and only if } s \in V(p) \\
(\mathcal{N}, s) \vDash_{\mathcal{NS}} \neg\varphi & \text{if and only if not } (\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \\
(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \wedge \psi & \text{if and only if not } (\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \\
(\mathcal{N}, s) \vDash_{\mathcal{NS}} \mathrm{K}_i\varphi & \text{if and only if } [\![\varphi]\!]^{\mathcal{N}} \in \nu_s(i)
\end{array}
$$

Given these definitions, there is a straight forward translation from individual epistemic relational models to individual epistemic neighbourhood models (and back again). Given $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$, we let $Z^i_s = [s]_{\sim_i}$ and let $\nu_s(i) = Z^i_s\uparrow$ and given $\mathcal{N} = (\mathcal{S}, (\nu_s)_{s\in\mathcal{S}}, V)$, we let $[s]_i = Z^i_s$ and define $\sim_i$ from those equivalence classes. This is equivalent to saying that $s \sim_i t$ if and only if $t \in Z^i_s$.

Lemma 2.26. *Given individual epistemic relational model $\mathcal{M}$ and individual epistemic neighbourhood model $\mathcal{N}$, there are modally equivalent models $\mathcal{N}^{\mathcal{M}}$ and $\mathcal{M}^{\mathcal{N}}$, respectively.*

We will not prove this lemma as it is implied by Theorem 3.7, which deals with both the K and the D modality which is introduced below.

Axiomatization

Given the language $\mathcal{L}_{\mathcal{K}}$ over $\mathcal{A}$ and Φ (see Definition 2.25), the following axiomatization (taken from [67]) is sound and complete with respect to relational frames over $|\mathcal{A}|$ many equivalence relations. From Lemma 2.26 it follows that the same system is sound and complete with respect to singleton knowledge neighbourhood systems.

All instantiations of propositional tautologies	
$\mathrm{K}_i(\varphi \to \psi) \to (\mathrm{K}_i\varphi \to \mathrm{K}_i\psi)$	Distribution of K over $\to$
From φ and $\varphi \to \psi$ infer ψ	Modus ponens
$\mathrm{K}_i\varphi \to \varphi$	Truth
$\mathrm{K}_i\varphi \to \mathrm{K}_i\mathrm{K}_i\varphi$	Positive introspection
$\neg\mathrm{K}_i\varphi \to \mathrm{K}_i\neg\mathrm{K}_i\varphi$	Negative introspection

2.4.4 Knowledge: The group case

In this section we will go through the three most used forms of group knowledge, with a special focus on *distributed knowledge*, which is part of the main topic of this thesis. All three forms of group knowledge is based on the single agent case, and are all different ways of combining the knowledge of each agent. To make clear the problem:

Example 2.27. $\mathrm{K}_a p$ and $\neg\mathrm{K}_b p$. Does $\mathrm{K}_{\{a,b\}}p$? $\dashv$

In the example, a knows p, whereas b does not. Do they together know p? This is obviously only a matter of definition, and what we want to analyse and use group knowledge for. If a and b are freely allowed to communicate, it is obvious that a needs only whisper p to b in order to obtain $\mathrm{K}_b p$. If both a and b knows p, it sounds reasonable that $\mathrm{K}_{\{a,b\}}p$ (or does it?). However, if a and b are not allowed to talk (recall now the Prisoner's dilemma (Example 2.39), letting p mean "both are quiet"), then it is clear that things are different, seeing as how a simply is not allowed or able to whisper anything to b.

Example 2.28. $\mathrm{K}_c q$ and $\mathrm{K}_d(q \to r)$. Does $\mathrm{K}_{\{c,d\}}r$? $\dashv$

This example is elaborated in Example 1.1 and further analysed in Section 2.4.4. This example is even more interesting than the example above, as in this case, none of the agents individually know r, but they each know something that might lead to its discovery.

The concepts we will discuss in this section are *distributed knowledge*, *mutual knowledge* and *common knowledge*. Informally, distributed knowledge amongst a group is what they can figure out if they are allowed to communicate (not entirely true, see below), mutual knowledge is simply what everyone in the group knows, whereas common knowledge is what every agent in the group know that every agent in the group know that every agent in the group ad infinitum. If $A \subseteq \mathcal{A}$ is a group of agents, we write $\mathrm{D}_A\varphi$ to mean that φ is distributed knowledge amongst A, we write $\mathrm{E}_A\varphi$ to mean that φ is mutual knowledge amongst A (i.e. $\bigwedge_{i\in A}\mathrm{K}_i\varphi$) and we write $\mathrm{C}_A\varphi$ when it is common knowledge in A that φ, or:

$$\mathrm{C}_A\varphi \Leftrightarrow \bigwedge_{n\in\omega} \mathrm{E}^n_A\varphi,$$

where $\mathrm{E}^1_A\varphi = \mathrm{E}_A\varphi$ and $\mathrm{E}^n_A\varphi = \mathrm{E}_A\mathrm{E}^{n-1}_A\varphi$ for $n > 1$.

The following example is due to [67, 64], in which they explain the presence of mutual knowledge and lack of common knowledge.

Example 2.29 (Alco at the conference)**.** Suppose a group of people, $\mathcal{A}$, including Alco is at a conference in Barcelona. At some point, Alco gets bored of all the logical talk and decides to go visit the bar for a drink (as Alco usually does). Some time later there is an announcement in the conference room. Unknown to the rest of the participants, there is an intercom system, relaying all the messages to the bar. This means that Alco also gets to hear the announcement.

If we call the announcement φ, it is now the case that every participant knows φ; it is mutual knowledge amongst $\mathcal{A}$ that φ. After hearing φ, Alco takes off. However, Betty, afraid that Alco missed the announcement, goes to the bar to tell him. There, she learns that Alco indeed heard the message. So now she knows that everybody knows. However, it is not the case that Alco knows that she knows this. φ is not common knowledge, but indeed mutual knowledge. $\dashv$

In the following example, it becomes clear that common knowledge is a prerequisite for cooperation, but is impossible to obtain under certain (very natural) circumstances. The example is a folklore example about common knowledge, sometimes referred to as The two generals' problem, two armies problem and coordinated attack. This version of the example is taken from [67].

Example 2.30 (Byzantine generals)**.** Imagine two allied generals, a and b, standing on two mountain summits, with their enemy in the valley between them. It is generally known that a and b together can easily defeat the enemy, but if only one of them attacks, he will certainly lose the battle.

General a sends a messenger to b with the message $\mathtt{m}$ (= "I propose that we attack on the first day of the next month at 8 PM sharp"). It is not guaranteed, however, that the messenger will arrive. Suppose that the messenger does reach the other summit and delivers the message to b. Then $\mathrm{K}_b\mathtt{m}$ holds, and even $\mathrm{K}_b\mathrm{K}_a\mathtt{m}$. Will it be a good idea to attack? Certainly not, because a wants to know for certain that b will attack as well, and he does not know that yet. Thus, b sends the messenger back with an "okay" message. Suppose the messenger survives again. Then $\mathrm{K}_a\mathrm{K}_b\mathrm{K}_a\mathtt{m}$ holds. Will the generals attack now? Definitely not, because b does not know whether his "okay" has arrived, so $\mathrm{K}_b\mathrm{K}_a\mathrm{K}_b\mathtt{m}$ does not hold, and common knowledge of $\mathtt{m}$ has not yet been established. $\dashv$

We will in the following three sections introduce the different group knowledge concepts, but it might be helpful to know beforehand that the following is the case: $\mathrm{C}_A\varphi \Rightarrow \mathrm{E}_A\varphi \Rightarrow \mathrm{D}_A\varphi \Rightarrow \varphi$. None of the other directions hold in general; φ can be the case without φ being distributed knowledge in A, $\mathrm{D}_A\varphi$ can be the case without φ being mutual knowledge in A, $\mathrm{E}_A\varphi$ can be the case without even $\mathrm{E}_A\mathrm{E}_A\varphi$ being the case. Indeed, we can make the hierarchy slightly stronger:

$$\begin{aligned}
\mathrm{D}_A\varphi &\Rightarrow \varphi \\
\mathrm{E}_A\varphi &\Rightarrow \mathrm{D}_A\varphi \\
\mathrm{E}_A\mathrm{E}_A\varphi &\Rightarrow \mathrm{E}_A\varphi \\
\mathrm{E}_A\mathrm{E}_A\mathrm{E}_A\varphi &\Rightarrow \mathrm{E}_A\mathrm{E}_A\varphi \\
&\vdots \\
\mathrm{E}_A^{n+1}\varphi &\Rightarrow \mathrm{E}_A^{n}\varphi \\
&\vdots \\
\mathrm{C}_A\varphi &\Rightarrow \mathrm{E}_A^{n}\varphi \text{ for all } n \in \omega
\end{aligned}$$

Distributed knowledge

> "Intuitively, a formula φ is distributed knowledge among a group of agents A iff φ follows from the knowledge of all individual agents in A put together. Semantically, φ is distributed knowledge among A iff φ is true in all worlds that *every* agent in A considers possible."
> – Distributed knowledge, Floris Roelofsen [55]

Distributed knowledge[5] is a very natural property [28, 26, 55], however in a modal logic setting, distributed knowledge is in fact very non-natural and indeed slightly ill-behaved. We will see below that indeed, distributed knowledge is not even modally definable, meaning that modal logic might not be the right tool for reasoning about it. It should be mentioned, however, that if we consider the subset of models called *full communication models* [55], distributed knowledge behaves exactly as we want. We will not go into the topic of full communication models in this thesis, but mention that it is a very natural restriction on the class of models we would like to consider. The reader is encouraged to consult any one of [55, 29, 63].

The concept and the definition of distributed knowledge in modal models are based on what we want distributed to mean, namely "What agents can come to know if they can share information". Recall Example 1.1, in which we could argue that Anna and Bill together knew it was raining outside despite the fact that none of them individually knew this. Let the proposition p be "Carlotta is wearing her red rain coat", let r be the proposition "it is raining". Then in the example, Bill knew p ($\mathrm{K}_b p$) and Anne knew that if p then r ($\mathrm{K}_a(p \to r)$). We can model this with the model in Figure 2.2 (reflexive arrows are omitted). The rectangular state is the current.

In the figure we can also observe that if we consider only the lack of knowledge that both have, then the only two states they cannot distinguish is the states where Carlotta is not wearing her coat.

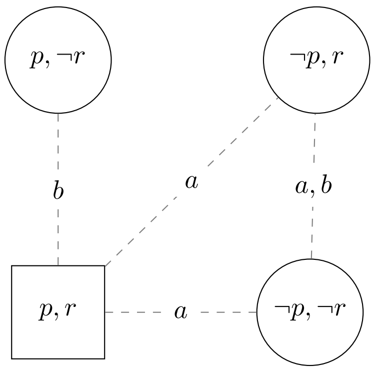
Formally, we define distributed knowledge for a coalition $A \subseteq \mathcal{A}$ in a point s in a model $\mathcal{M}$, as what is true in the equivalence class around s for the *intersection* of the equivalence classes of each of the agents in A, that is

$$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{D}_A\varphi \text{ if and only if for all } t \sim^{\mathrm{D}}_A s \ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi,$$

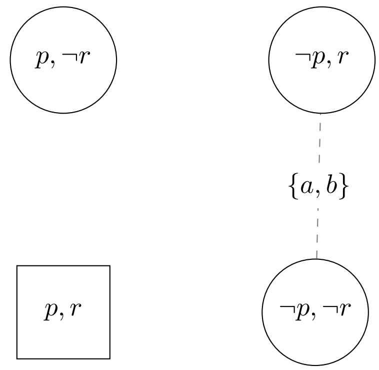
where $\sim^{\mathrm{D}}_A = \bigcap_{i \in A} \sim_i$ is the lifted relation for A.

Modal definability As we have mentioned several places, there is a problem with distributed knowledge when it comes to modal logic. The problem is that the D-operator is not modally definable in terms of the basic modal language, that is, given only relations per individual, we cannot define the D-operator in terms of the K-operators. This is in contrast to mutual knowledge, as we

[5] Sometimes, though rarely, referred to as *implicit knowledge*, in e.g. [33, 64]. That name refers to the fact that something is known *implicitly* among a group, without necessarily anyone knowing it explicitly. This is not to be confused with the way "implicit" is being used in the concept of *implicit belief* ([46]).



(a) Individual knowledge



(b) Distributed knowledge

Figure 2.2: Individual and distributed knowledge of the weather

will see can be expressed as a single conjunction over Ks. It can be shown (by induction on the length of φ) that two bisimilar models are modally equivalent. As we will see below, bisimilar models are not invariant under the D-operator.

Definition 2.31 (Bisimulation)**.** Let two relational models $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i \in \mathcal{A}}, V)$ and $\mathcal{M}' = (\mathcal{S}', (\sim'_i)_{i \in \mathcal{A}}, V')$ over $\mathcal{A}$ and Φ be given. A non-empty relation $Z \subseteq \mathcal{S} \times \mathcal{S}'$ is a *bisimulation* if and only if for all $s \in \mathcal{S}$ and $s' \in \mathcal{S}'$ such that $(s, s') \in Z$, the following properties hold:

Local harmony $s \in V(p)$ if and only if $s' \in V'(p)$ for all $p \in \Phi$,

Forth for all $i \in \mathcal{A}$, and all $t \in \mathcal{S}$ such that $s \sim_i t$, there is a $t' \in \mathcal{S}'$ such that $s' \sim'_i t'$ and $(t, t') \in Z$ and

Back for all $i \in \mathcal{A}$, and all $t' \in \mathcal{S}'$ such that $s' \sim_i t'$, there is a $t \in \mathcal{S}$ such that $s \sim_i t$ and $(t, t') \in Z$.

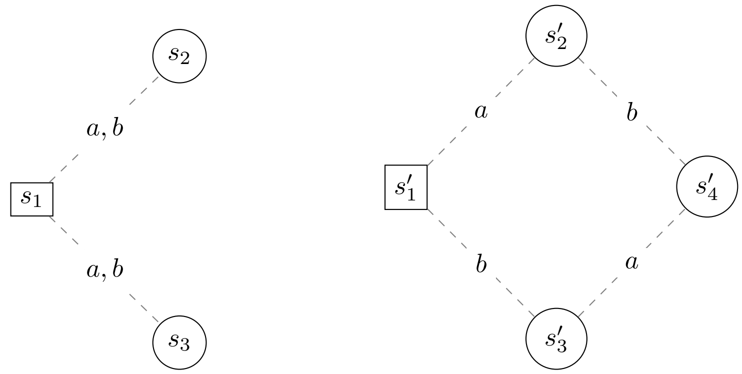
$\dashv$

The models in Figure 2.3 are bisimilar, but are not equivalent with respect to the D-modality. To see that they are bisimilar, we observe that the following is a bisimulation

$$Z = \{(s_1, s'_1), (s_2, s'_2), (s_3, s'_3), (s_1, s'_4)\}.$$

The reader can verify, that Z indeed is a bisimulation (e.g. s_1 forth: s_1 has an a-arrow to s_2 in $\mathcal{M}$, and s'_1 has a-arrow to s'_2 and s_2 and s'_2 are bisimilar). We can now observe that the following is the case:

- $(\mathcal{M}, s_1) \vDash_{\mathcal{RS}} \neg \mathrm{D}_{\{a,b\}} p$ since $s_1 \sim_{\{a,b\}} s_2$ and since $(\mathcal{M}, s_2) \vDash_{\mathcal{RS}} \neg p$ and that
- $(\mathcal{M}', s'_1) \vDash_{\mathcal{RS}} \mathrm{D}_{\{a,b\}} p$ since the only state in the $\sim_{\{a,b\}}$-relation to s'_1 is itself, and s'_1 indeed satisfies p.



(a) Model $\mathcal{M}$ with $V(p) = \{s_1\}$ and $V(q) = \{s_2, s_3\}$

(b) Model $\mathcal{M}'$ with $V'(p) = \{s'_1, s'_4\}$ and $V(q) = \{s'_2, s'_3\}$

Figure 2.3: Bisimilar models that are not equivalent with respect to the D operator. This shows that D is not invariant under bisimulation and that D is not modally definable in the basic modal language.

van Benthem asks in his 2006 paper, [61] if there is a natural extension of the notion of bisimulation which is characteristic for distributed knowledge:

> **Problem 1.** "Find a natural extended notion of bisimulation that is characteristic for epistemic logic enriched with $\mathrm{D}_G\varphi$."
> – van Benthem [61].

This question was solved in 2007, in a paper by Roelofsen [55], where he coined the term *collective bisimulation*. If two models are collective bisimilar, they are modally equivalent, also with respect to the D-modality, even when defined in terms of individual indistinguishability relations.

Definition 2.32 (Collective bisimulation)**.** Let $(\mathcal{M}, s)$ and $(\mathcal{M}', s')$ be two pointed relational models over the same agent set $\mathcal{A}$ and propositional letters Φ. A non-empty set $Z \subseteq \mathcal{S} \times \mathcal{S}'$ is a collective bisimulation if for all $(s, s') \in Z$

Local harmony $s \in V(p)$ if and only if $s' \in V'(p)$ for all $p \in \Phi$,

Collective forth for every coalition $A \subseteq \mathcal{A}$, for all $t \in \bigcap_{i \in A} R_i[s]$, there is a $t' \in \bigcap_{i \in A} R'_i[s']$ such that $(t, t') \in Z$ and reversely

Collective back for every coalition $A \subseteq \mathcal{A}$, for all $t' \in \bigcap_{i \in A} R'_i[s']$, there is a $t \in \bigcap_{i \in A} R_i[s]$ such that $(t, t') \in Z$,

where $R_i[s] = \{t \in \mathcal{S} \mid (s, t) \in R_i\}$. $\dashv$

As we can read directly off from this definition, this definition simply insists that it is not enough for the individual relations to satisfy the back and forth conditions; the intersection relations must satisfy the properties as well.

Example 2.33. It is clear that the two models in Figure 2.3 are not bisimilar. To see that, we immediately see that s_1 and s'_1 should be in Z, since they share the same propositional letters. But from s_1 there is an $\{a, b\}$-arrow to s_2. The only $\{a, b\}$-arrow going out from s'_1 is to s'_1 itself, which cannot be linked with s_2 by local harmony. $\dashv$

Roelofsen then proceeds to prove that this is indeed the notion that van Benthem was asking for. Consult [55] for details and full proof.

Mutual knowledge

Contrary to distributed and common knowledge, it seems the community has not agreed on a name for *mutual knowledge*. In the literature, we can therefore find the concept of mutual knowledge referred to as both *everybody knows*-knowledge, e.g. [16, 26, 30, 65, 67] and *general knowledge* [67, 64]. Since we need to choose one, we will in this thesis stick with the name used in [49, 40, 70], which is mutual knowledge. Firstly, it is less clumsy than "everybody knows"-knowledge, and it also seems like mutual knowledge is more often used in logic than general knowledge is.

Formally, we will here define mutual knowledge for a coalition $A \subseteq \mathcal{A}$ in a point s in a model $\mathcal{M}$ in two different ways, and then observe that they are equivalent. The syntactic way is to let $\mathrm{E}_A\varphi$ be defined as $\bigwedge_{i\in A} \mathrm{K}_i\varphi$, that is if $A = \{1, 2, \ldots, m\}$ we define $\mathrm{E}_A\varphi = \mathrm{K}_1\varphi \wedge \mathrm{K}_2\varphi \wedge \cdots \wedge K_m\varphi$. As long as our agent set is finite (which it is in our case), this formula will be finitary, and hence a well-formed formula. The semantic (or structural) way is to define it as what is true in the set of states around s for the *union* of the equivalence classes of each of the agents in A, that is

$$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{E}_A\varphi \text{ if and only if for all } t \sim^{\mathrm{E}}_A s \ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi,$$

where $\sim^{\mathrm{E}}_A = \bigcup_{i\in A} \sim_i$ is the lifted relation for A. It is very important to know that this relation is *not* an equivalence relation, we will however write the symbol $\sim$ instead of R since it fits better with the rest of the group relations, but we must treat it with care.

Now, let $(\mathcal{M}, s)$, φ and $A \subseteq \mathcal{A}$ be given. Assume that $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{E}_A\varphi$ be the case using the former definition. This holds if and only if $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \bigwedge_{i\in A} \mathrm{K}_i\varphi$. But this again holds if and only if, by definition of K, for all $i \in A$ and for all $t \in \mathcal{S}$, if $s \sim_i t$, then $(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$. But this is again the same as if for all $t \in \mathcal{S}$ if $s \sim^{\mathrm{E}}_i t$ then $(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$, which is the latter definition.

Common knowledge

> "Two people, 1 and 2, are said to have *common knowledge* of an event $\mathcal{E}$ if both know it, 1 knows that 2 knows it, 2 knows that 1 knows it, 1 knows that 2 knows that 1 knows it, and so on."

This is how Robert Aumann starts his paper Agreeing to disagree [12] from 1976. He was one of the first people studying common knowledge in a formal mathematical way, doing so with a set theoretic approach of partitioning probability spaces. In this section we will discuss common knowledge in a modal logical setting. The two approaches are equivalent.

As is the case with both mutual and distributed knowledge, for the singleton groups, the operator $\mathrm{C}_{\{i\}}$ coincides with K_i. Formally, we could define C to be the following

$$\mathrm{C}_A\varphi \Leftrightarrow \bigwedge_{n\in\omega} \mathrm{E}^n_A\varphi,$$

where $\mathrm{E}^1_A\varphi = \mathrm{E}_A\varphi$ and $\mathrm{E}^n_A\varphi = \mathrm{E}_A\mathrm{E}^{n-1}_A\varphi$ for $n > 1$. However, the right hand side of the formula is not a well-formed modal logic formula.[6] We therefore define common knowledge in a semantic manner. Let $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$ be a model, with $s \in \mathcal{S}$ and $A \subseteq \mathcal{A}$. We define truth of common knowledge of a formula φ for A in s as follows:

$$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{C}_A\varphi \text{ if and only if for all } t \sim^{\mathrm{C}}_A s\ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \mathrm{C}_A\varphi,$$

where $\sim^{\mathrm{C}}_A = (\bigcup_{i\in A} \sim_i)^*$ is the transitive closure. However, there is a problem with transitive closure in many finitary logics, as the transitive closure is essentially an infinitary operation. The punishment for this is that any logic with common knowledge defined like this is non-compact, which in the end means that we can forget about strong completeness.

Definition 2.34 (Compactness)**.** We say that a logic Λ over a language $\mathcal{L}$ is compact if it holds that for any nonempty set of formulae $\Gamma \subseteq \mathcal{L}$, if for all finite subsets $\Gamma' \subseteq_{<\omega} \Gamma$, we have that Γ' is satisfiable, then Γ is satisfiable. We can easily prove that common knowledge is not compact. $\dashv$

Lemma 2.35. *Common knowledge is not compact.*

Proof. Let $\Gamma = (\bigcup_{n\in\omega} \mathrm{E}^n_A\varphi) \cup \{\neg\mathrm{C}_A\varphi\}$. It is an easy exercise to verify that Γ is finitely satisfiable, but Γ itself is not. $\square$

2.5 Alternating-Time Temporal (Epistemic) Logic

The most used logic for reasoning about coalitional ability and power, is the logic Alternating-Time Temporal Logic (ATL), introduced in a series of papers.[7] Building their work on CTL and CTL*. We will quickly introduce the language and then see what the semantic systems look like. CTL contains *path modalities*, meaning that one can express statements as "there is a path in which φ is true somewhere" and "in all paths, φ is true somewhere". What ATL did, was to add coalition modalities making it possible to say "coalition A has the power to make sure that φ will be the case" and more. We skip the CTL language and semantics, and go directly to ATL. For an intro to CTL, consult [38, 54].

[6] It is not a well-formed formula in our language at least, as our language is finitary. In *infinitary* modal logic, the formula would be well-formed.

[7] It is worth mentioning that ATL and ATEL is far from the only logics reasoning about coalitional power or abilities. See e.g. [5, 4], using concepts from epistemic logic and game theory and STIT, [15, 36].

2.5.1 Alternating-time temporal logic – language

The language for ATL, $\mathcal{L}_{\mathcal{ATL}}$ is defined by the following syntax

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\!\langle A \rangle\!\rangle \mathrm{X}\varphi \mid \langle\!\langle A \rangle\!\rangle \mathrm{G}\varphi \mid \langle\!\langle A \rangle\!\rangle \varphi \mathrm{U} \varphi,$$

where $p \in \Phi$ is a proposition letter and $A \subseteq \mathcal{A}$. In the language $\langle\!\langle A \rangle\!\rangle$ is a coalition modality, and is always followed by a branching temporal operator, and each branching temporal operator is always following a coalition modality. The coalition modality $\langle\!\langle A \rangle\!\rangle$ is read as "A can force that", or "A has a strategy such that", and together with the branching temporal operators, $\langle\!\langle A \rangle\!\rangle \mathrm{X}\varphi$ is read as "A has a strategy to enforce that the next state will satisfy φ", $\langle\!\langle A \rangle\!\rangle \mathrm{G}\varphi$ is read as "A has a strategy to ensure φ being true always", $\langle\!\langle A \rangle\!\rangle \varphi \mathrm{U} \psi$ is read as "A has a strategy to ensure φ being the case until ψ is the case". We can also define the *future* operator as $\langle\!\langle A \rangle\!\rangle \mathrm{F}\varphi = \langle\!\langle A \rangle\!\rangle \top \mathrm{U} \varphi$. There is an extension of ATL called ATL* that allows branching temporal operators without being preceded by a coalition modality. We will not discuss this language here.

Examples of well-formed formulae of ATL are, if $\{a, b, c\} \subseteq \mathcal{A}$, and $\{p, q\} \subseteq \Phi$, $\neg\langle\!\langle \{a\} \rangle\!\rangle \mathrm{X}p$, $\neg\langle\!\langle \{a\} \rangle\!\rangle \mathrm{X}p \wedge \langle\!\langle \{a, b\} \rangle\!\rangle \mathrm{X}p$ and $(\neg\langle\!\langle \{a\} \rangle\!\rangle \mathrm{X}p) \wedge \langle\!\langle \{a, b\} \rangle\!\rangle p \mathrm{U}(\langle\!\langle \{a, b, c\} \rangle\!\rangle \mathrm{X}q)$.

2.5.2 Alternating-time temporal logic – semantics

There are at least two widely used semantics for ATL. The first that was mentioned is the ATS [9, 19]. This was the first semantics that Alur et al. introduced, and they used this in their first paper. In the followup papers, they introduced a new semantics, called CGS, that fixed some of the drawbacks of the ATSs. This was developed in [10], it also goes by the name MGM (Multi-player Game Model) in [30].

In this section we will introduce all notions we will need in order to define satisfiability in the semantics, in fact, we will define the semantic system we will use, namely that of a *Concurrent Game System* (CGS). A CGS over $\mathcal{A}$ and Φ is a structure containing a set of *states* $\mathcal{S}$, a set of *actions*, Σ, which is the collection of all possible actions. Some of the actions might not be applicable for every agent, or even in every state, and that is why we have a function δ; It assign to each agent a set of actions per state (modulo currying).

Given Σ, a set of actions, by σ_i we mean a function mapping i to an action (i.e. the simple function $\{(i, \alpha)\}$, where $\alpha \in \Sigma$). We call this simply an individual action. When $A \subseteq \mathcal{A}$, we denote by σ_A a function mapping each agent of A to individual actions, and we call σ_A a *collective action*. When $B \subseteq \mathcal{A}$ is disjoint from A, we denote by $\sigma_A \sigma_B$ the union of the two functions σ_A and σ_B. $\sigma_A \sigma_B$ will then be a collective action for $A \cup B$.

A function mapping states to individual actions is called a (memoryfree) *strategy*. We denote, for $i \in \mathcal{A}$, a strategy for i by f_i, and when $A \subseteq \mathcal{A}$ we call F_A a collective strategy, when F_A maps states to collective actions.

Remark 2.36 (Memoryless strategies)*.* In ATL and ATEL, and most certainly in the logic DCL (Chapter 6), memoryfree strategies and memorybased strategies coincide. We will therefore assume, unless otherwise is specified, that all strategies are memoryfree. On the other side, any proof in the literature (as far as the author is aware) of this correspondence is making the assumption that the state space is finite, an assumption we do not make in Chapter 6. This is not an issue, as memory in strategies only could make a difference when interpreting the path quantifiers of ATL; In DCL we only use the next time fragment, effectively making this a non-issue. $\dashv$

When we run a game in a system, we will necessarily obtain a sequence of states. We assume this sequence to be infinite[8], and we denote a sequence by λ. Essentially, a sequence λ is a function $\lambda : \omega \to \mathcal{S}$. For j a natural number, we write $\lambda(j)$ to mean the j'th element (state) in the sequence, with $\lambda(0)$ meaning our current state. A function mapping (finite) sequences of states, $\mathcal{S}^*$ to an action is called a *memory-based* strategy. Such a strategy can be either unbounded (it maps the entire history to a state), or bounded, either having a fixed number k such that it maps sequences of length at most k to a state, or other ways of incorporating memory.

Concurrent game systems

A concurrent game system (CGS) over $\mathcal{A}$ and Φ, is a tuple $\mathcal{M} = (\mathcal{S}, \Sigma, \delta, o, V)$ where $\mathcal{S}$ is a non-empty set of states and Σ a non-empty set of actions. $\delta : \mathcal{A} \times \mathcal{S} \to 2^{\Sigma}$ is a mapping that assigns to each agent and each state, a set of available actions. We will sometimes write $\delta_i(s)$ instead of $\delta(i, s)$. A collective action for $A \subseteq \mathcal{A}$ is a tuple σ_A where $\sigma_i \in \Sigma$ for each $i \in A$, and a collective action is a collective action for $\mathcal{A}$. A (memoryfree) strategy f_i for an agent i is a function mapping a state s to an individual action in $\delta_i(s)$. A collective strategy for a group A is a function F_A where $F_A(i)$ is a strategy for each agent $i \in A$. A collective strategy for $\mathcal{A}$ is simply called a strategy. Given a collective action for $\mathcal{A}$, $\sigma_{\mathcal{A}}$, $o(s, \sigma_{\mathcal{A}}) \mapsto t$, for $\sigma_i \in \delta_i(s)$ is a function mapping a state and a collective action to a state. We will often write $o_s(\sigma)$ for $o(s, \sigma)$. $V : \Phi \to 2^{\mathcal{S}}$ is the evaluation function. We define $o_s(\sigma_A)$, to be the set $\{t \in \mathcal{S} \mid \exists \sigma_{\overline{A}} \text{ and } t = o_s(\sigma_A \sigma_{\overline{A}})\}$, where $\overline{A} = \mathcal{A} \setminus A$. Thus, $o_s(\sigma_\emptyset) = o_s(\epsilon)$ is the set of all outcome states from s.

If F is any memoryfree collective strategy, we define the function $o^*(s, F)$ to be the path λ determined by repeated applications of o with the respective state, i.e. $\lambda(0) = s$, and for each $i \geq 0$, $\lambda(i+1) = o(\lambda(i), F(\lambda(i)))$. If F_A is a collective strategy for A, $o^*(s, F_A)$ gives the *set of all paths* subject to F_A.

Given a pointed CGS $(\mathcal{M}, s)$ and $\varphi \in \mathcal{L}_{\mathcal{ATL}}$, the evaluation of truth is as

[8] A sequence of states is always assumed to be infinite; Every collective action in every game leads to a new game: You can't stop playing

follows:

$(\mathcal{M}, s) \vDash_{\mathcal{ATL}} p$ iff $s \in V(p)$
$(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \neg\varphi$ iff not $(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \varphi \wedge \psi$ iff $(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \varphi$ and $(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \psi$
$(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \langle\langle A\rangle\rangle \mathrm{X}\varphi$ iff A has a collective action σ_A such that
for any $\sigma_{\overline{A}}$, we have that $(\mathcal{M}, o(\sigma_A\sigma_{\overline{A}})) \vDash_{\mathcal{ATL}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \langle\langle A\rangle\rangle \mathrm{G}\varphi$ iff A has a collective strategy F_A such that for any $F_{\overline{A}}$,
if $\lambda = o^*(s, F_A F_{\overline{A}})$, then for all $i \in \omega$ we have that $(\mathcal{M}, \lambda(i)) \vDash_{\mathcal{ATL}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{ATL}} \langle\langle A\rangle\rangle \varphi \mathrm{U}\psi$ iff A has a collective strategy F_A such that for any $F_{\overline{A}}$,
if $\lambda = o^*(s, F_A F_{\overline{A}})$, then there is an $i \in \omega$ such that
$(\mathcal{M}, \lambda(i)) \vDash_{\mathcal{ATL}} \psi$ and for all $j < i$, $(\mathcal{M}, \lambda(j)) \vDash_{\mathcal{ATL}} \varphi$

It can be mentioned that ATL has some peculiarities when it comes to some formulae being true in models (see e.g. Section 5.2.3), as we will see immediately below also is the case with ATEL. We will not go into details on this topic, but the interested reader can consult [3].

2.5.3 Concurrent epistemic game systems

We will not use the same semantic structure as presented in [65] alternating epistemic transition system (AETS), we will use the CEGS version given in [40], that is, *concurrent epistemic game system* (CEGS)[9]. The CEGS is simply the extension of CGS, where we introduce relations $\sim_i$ for each $i \in \mathcal{A}$. Furthermore, we will assume the relations are equivalence relations.[10] This kind of a structure was in [57] called an *iCGS*, for *imperfect information concurrent game structure*, however we stick with CEGS as it fits better in with the conventions (viz. ATS/AETS [65], MGM/MEGM [30]). In addition to simply adding an equivalence relation $\sim_i \subseteq \mathcal{S} \times \mathcal{S}$ for each $i \in \mathcal{A}$, we put additional constraints on δ_i; We demand that the actions available for an agent i must be equivalent in indistinguishable states.

Definition 2.37 (Concurrent Epistemic Game System)**.** A concurrent epistemic game system, or CEGS for short, is a tuple $\mathcal{M} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}}, V)$, where

- $\mathcal{S}$ is a non-empty set of states,

[9] In [30] CEGS is referred to as MEGM (Multi-player epistemic game model), just as they call CGS by the name MGM.

[10] It might also be interesting to develop a logic over CEGSs where the indistinguishability relations are not necessarily equivalence relations, but maybe relations that are used in e.g. doxastic logic. This way we can model what agents and coalitions believe they can achieve, but not necessarily force. We must still make sure that actions are uniform, i.e. if $R_i(s,t)$ then if $\alpha \in \delta_s(i)$ is an action available in s, then $\alpha \in \delta_t(i)$, i.e. α must be available in t.

- Σ is a set of actions and
- $\delta : \mathcal{A} \times \mathcal{S} \to 2^{\Sigma}$ is a function assigning for each agent a set of available actions in a state. We pose the restriction that if $s \sim_i t$ then $\delta_i(s) = \delta_i(t)$.
- o is a function sending a collective action for $\mathcal{A}$, i.e. a tuple $\sigma \in \Sigma^{|\mathcal{A}|}$ and a state s to a new *outcome state*, i.e. $o : \mathcal{S} \times \Sigma^{|\mathcal{A}|} \to \mathcal{S}$. This function does not need to be total, $o_s(\sigma)$ needs only be defined when $\sigma \in \times_{i\in\mathcal{A}}\delta_i(s)$,
- $\sim_i \subseteq \mathcal{S} \times \mathcal{S}$ is an equivalence relation for each agent $i \in \mathcal{A}$ and
- $V : \Phi \to 2^{\mathcal{S}}$ is a valuation function.

We sometimes discuss such structures without the valuation function V; We then call the structure $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}})$ a concurrent epistemic game frame (CEGF), where all components are as in the definition above. $\dashv$

2.5.4 Alternating-time temporal epistemic logic

Alternating-time temporal epistemic logic was introduced in [65, 66] (see the survey in Section 5.2). In these papers they introduce ATEL and its semantics, however, there are several reasons we will deviate from their logic. Interestingly, they do not mention distributed knowledge at all, but do analyse mutual and common knowledge. It might therefore be interesting to analyse ATEL with the distributed knowledge operator. Before diving into the analysis of the problems with ATEL, we introduce the language and the semantics. We will deviate from the original papers, as they were based on a structure called AETS, which comes with its own drawbacks; We will use CEGS, the concurrent epistemic game systems, see Definition 2.37.

The ATEL language $\mathcal{L}_{\mathcal{ATEL}}$ over our usual agent set $\mathcal{A}$ and proposition set Φ, is defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \langle\!\langle A \rangle\!\rangle \mathrm{X}\varphi \mid \langle\!\langle A \rangle\!\rangle \mathrm{G}\varphi \mid \langle\!\langle A \rangle\!\rangle \varphi \mathrm{U}\psi \mid \mathrm{C}_A\varphi \mid \mathrm{E}_A\varphi,$$

where $A \subseteq \mathcal{A}$, $i \in \mathcal{A}$ and $p \in \Phi$. We use the shorthand $\mathrm{K}_i\varphi = \mathrm{C}_{\{i\}}\varphi$ for $i \in \mathcal{A}$ and by $\langle\!\langle A \rangle\!\rangle \mathrm{F}\varphi$ we mean $\langle\!\langle A \rangle\!\rangle \top \mathrm{U}\varphi$.

There are several known “problems” with ATEL, and they are all related to counter intuitive properties. The following example appeared as Example 1 in [36] (see also Example 5.2).

Example 2.38 (Blind man and light switch)**.** A blind man a is in a room with a light switch. The light switch controls a lamp in the room. a can choose to remain passive, or to press the light switch. Even though a does not know if the light is on or not, he has a strategy making sure the light is on, namely to either press the button (if the light is off) or remain passive (if the light is on). $\dashv$

When modelling this in ATEL, the system we make to represent this situation, will give a the power to ensure that the light is on, and that even a knows that he has an action to ensure that the light is on. This is highly counter intuitive, but is an immediate consequence of the fact that ATEL makes use of de dicto strategies, rather than de re strategies (Section 5.1.1). In the logic we will use in Section 6.5, we will see that in the same system, it would indeed *not* be the case that a can ensure the light being switched on. However, if a would not be blind, he would have a strategy that makes the light be on.

Also differing from [65], we will use only the language ATL (actually, we will use CL, which is the nexttime fragment of ATL), that is, we do not have any knowledge modalities. This means that we cannot express directly things about knowledge, e.g. we cannot say $\mathrm{K}_i\varphi$, but then again, this is not what we want to model, either.

2.6 Game theory

Game theory is, like modal logic, a relatively young mathematical research area, initiated by mainly von Neumann in an article dating back to 1928, and then later in the book with Morgenstern [71] and also extended in [48]. Its purpose is the analysis of strategies and power in multi-agent situations. The notion of "multi-agent situation" is supposed to be very broad, an agent can be anything that can *act* on some environment. Examples of agents range from the obvious human beings, animals and computers, to less obvious such as nature, randomness by e.g. dice etc. A strategy can be any "act" that might lead to a change in the environment.

2.6.1 Strategic games

Formally, we define a *strategic game* to be a tuple $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o)$, where Σ is the set of actions and for each agent i, $\delta_i \subseteq \Sigma$ is the set of available actions for i, $\mathcal{S}$ is a set of *possible outcomes* and o is a function sending *collective actions* to $\mathcal{S}$, i.e. $o : \times_{i\in\mathcal{A}}\delta_i \to \mathcal{S}$. By analysing o's behaviour, we can express power of agents, as what in $\mathcal{S}$ they can *force*, if they just choose the correct action. Furthermore, if we allow agents to cooperate, we can talk about coalitional power.

In game theory, we in addition to the above tuple, have a preference relation over $\mathcal{S}$, i.e. for each agent $\succsim_i \subseteq \mathcal{S} \times \mathcal{S}$ is a total reflexive and transitive ordering over the outcome. However, as we mention in Remark 2.44, we are not interested in the quantitative measure of the outcome, but a qualitative measure. For that reason, we will not discuss Nash equilibria here (see e.g. [48, 49] for a thorough analysis of Nash equilibria and other solution concepts of strategic games).

The following example is the de facto example in game theory, and this version is taken from [49].

Example 2.39 (The Prisoner's Dilemma)**.** Two suspects in a crime are put into separate cells. If they both confess, each will be sentenced to three years in prison. If only one of them confesses, he will be freed and used as a witness against the other, who will receive a sentence of four years. If neither confesses, they will both be convicted of a minor offence and spend one year in prison.

	Cooperate	Confess
Cooperate	$(3,3)$	$(0,4)$
Confess	$(4,0)$	$(1,1)$

⊣

This kind of table is called the *strategic form* or *normal form* of the game. It tells us that if both players chooses to cooperate (that is, both keep their mouth shut), they will get a utility of 3 each, whereas if they both choose to confess, they both get a utility of 1. It seems obvious that if both cooperates, they will be better off, however, the (only) Nash equilibrium is the situation where both confess. According to our modelling of a strategic game $\mathcal{S} = \{(3,3),(0,4),(4,0),(1,1)\}$, $\Sigma = \{\text{Cooperate}, \text{Confess}\}$ with $\delta_1 = \delta_2 = \Sigma$, and o works according to the table.

Example 2.40 (Mozart or Mahler)**.** Mozart or Mahler is a variant of the game Battle of the Sexes. In this game, both players get a utility of 2 if they both go to the Mozart concert, they get utility 1 if they go to the Mahler concert and they both get 0 if they choose to go to different concerts.

	Mozart	Mahler
Mozart	$(2,2)$	$(0,0)$
Mahler	$(0,0)$	$(1,1)$

This is a standard coordination text book example. It has two Nash equilibria, namely (Mozart, Mozart) and (Mahler, Mahler). ⊣

Definition 2.41 (α-effectivity)**.** Given a strategic game $\mathcal{G}$, we denote by $E^{\mathcal{G}} : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ the *α-effectivity function* of $\mathcal{G}$, defined by $X \in E^{\mathcal{G}}(A)$ if and only if there is a collective strategy σ_A such that for all strategies of the opposing coalition $\overline{A}$, $\sigma_{\overline{A}}$, $o(\sigma_A \sigma_{\overline{A}}) \in X$. ⊣

Proposition 2.42 (Goranko et al. [31])**.** *For every α-effectivity function $E^{\mathcal{G}} : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$, the following hold:*

1. *The nonmonotonic core of $E^{\mathcal{G}}(\emptyset)$ is the singleton set $\{Z\}$ where $Z = \{x \in \mathcal{S} \mid x = o(\sigma_{\mathcal{A}})$ for some $\sigma_{\mathcal{A}}\}$ and*
2. *$E^{\mathcal{G}}(\emptyset)$ is the principal filter generated by Z.*

2.6.2 Cooperative games

Contrary to the strategic games, where we formalise strategies and outcome functions on a single agent level, in cooperative games, we abstract away the strategies and outcome, and formalise the *power of coalitions*. It is a simple exercise to see that cooperative games subsumes strategic games, simply by looking only at the power of single agents, and then combining coalitions in a superadditive manner. A remark is in order here; We do not describe cooperative games here, but instead will focus on something called a coalition frame. The coalition frame has a different semantic than cooperative games. See Remark 2.43 for more on this distinction.

A coalition frame is a tuple $\mathcal{N} = (\mathcal{A}, E, \mathcal{S})$, where $E : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ is a coalitional effectivity function. Where the main ingredients of a strategic game was the strategies of each agent, and the outcome function, the main part of a coalition frame is the effectivity function. This function maps coalitions to sets of sets of elements from $\mathcal{S}$, with the meaning that if $X \in E(A)$, then A can force the outcome to be in X, but not necessarily which of the elements it actually will be.

Remark 2.43 (Coalition frame versus coalitional game). In [49], a coalitional game is defined as a tuple $(\mathcal{A}, X, V, (\succsim_i)_{i\in\mathcal{A}})$, where X is a nonempty set of *consequences*, $V : 2^{\mathcal{A}} \to 2^X$ and $\succsim_i \subseteq X \times X$ is a total reflexive and transitive preference relation. The function V maps (nonempty) coalitions to a set the coalition is worth, and $\succsim_i$ is a preference relation over consequences for each agent. It is important to note that in this definition, the meaning of $V(A) = Y$ is that A can *get* or *achieve* Y. This differs from our definition of a coalition frame. $\dashv$

Remark 2.44 (Qualitative versus quantitative measure of coalitional power). Usually in game theory, a cooperative game with transferable payoff is a structure with $(\mathcal{A}, v)$ where $v : 2^{\mathcal{A}} \to \mathbb{R}$. The function v denotes the *value* or *worth* of a coalition A, and $v(A)$ is some real number meaning what A can obtain by cooperating. This is a quantitative measure of power, in which one can only say that $A \subseteq \mathcal{A}$ is not less powerful than $B \subseteq \mathcal{A}$ if $v(A) \geq v(B)$.

However, we are interested in a more *qualitative* measure of power, meaning that we can not necessarily say who is more powerful, but we can measure *what* a coalition is powerful *for*. This is the intention of our effectivity function E, and by writing for $X \subseteq \mathcal{S}$ that $X \in E(A)$, we mean that A can *force* X to be the case, or to force that the "outcome" would lie in X. The interested reader can consult [30] or even [49] for more on the distinction between games with and without transferable payoff. $\dashv$

2.7 Coalition logic

Coalition logic was first introduced by Marc Pauly in his thesis [52] and later papers, e.g. [53]. The logic formalises reasoning about *coalitional effectivity* in games, i.e. what a coalition is capable of forcing. However, as with the difference between strategic games and cooperative games, the semantic of coalition logic abstracts away strategies and outcomes. Given a set of agents $\mathcal{A}$ and a coalitional effectivity function $E : \mathcal{S} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$, this logic allows us to express that $A \subseteq \mathcal{A}$ can *force* some set X to be the case, if $X \in E_s(A)$. Since this is a logic, we will say that A can force φ to be the case if there is some set $X \in E_s(A)$ such that for all $x \in X$, φ is true in x. Recall that $[\![\varphi]\!]$ is the set of all states satisfying φ, we then can write that A is effective for φ, formally $\langle\!\langle A \rangle\!\rangle\varphi$ if, $[\![\varphi]\!] \in E_s(A)$. This notion will be made explicit in the following sections. The reader might observe that the syntax $\langle\![A]\!\rangle\varphi$, meaning A is effective for φ, differs somewhat from [53] (where $[A]\varphi$ is used), but is used several places (e.g. [20, 6]) since it matches the $\exists\forall$-pattern its truth definition has. In [21], the modalities in CL has the form $\langle\![A]\!\rangle \mathrm{X}\varphi$ to make it clear that the effectivity function talks about "one step computations".

The main result of Pauly's papers is, perhaps not the logic (being subsumed by ATL, see below), but Pauly's representation theorem. In the papers he gives an interesting representation theorem that sets a few very natural restrictions on how an effectivity function must look in order to have a corresponding strategic game. In addition, it is shown that every strategic game has a corresponding effectivity function (this is simply the α-effectivity function we defined in Definition 2.41).

2.7.1 Language

Formally, the language for CL, $\mathcal{L}_{\mathcal{CL}}$ is defined by the following rules

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\![A]\!\rangle\varphi,$$

where $p \in \Phi$ is a proposition letter and $A \subseteq \mathcal{A}$ is a coalition. The formula $\langle\![A]\!\rangle\varphi$ is read as "A can force φ to be the case in the next state". This is reminiscent of the ATL formula $\langle\!\langle A\rangle\!\rangle \mathrm{X}\varphi$, and as has been shown in amongst others, [30], ATL indeed subsumes CL by translating $\langle\![A]\!\rangle\varphi$ to $\langle\!\langle A\rangle\!\rangle \mathrm{X}\varphi$.[11]

2.7.2 Semantics

Coalition logic can be interpreted in both relational models, and in neighbourhood models. We will in this section see how to evaluate formulae from $\mathcal{L}_{\mathcal{CL}}$ in the latter.

Definition 2.45 (Coalition frames and models)**.** A coalition frame is a pair $\mathcal{C} = (\mathcal{S}, E)$, where $E : \mathcal{S} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ is a coalitional effectivity function. If $V : \Phi \to 2^{\mathcal{S}}$ is a valuation function, we call $\mathcal{N} = (\mathcal{C}, V)$ a coalition model. When $s \in \mathcal{S}$, we call the pair $(\mathcal{N}, s)$ a pointed coalition model. $\dashv$

[11] There is an extension of CL, ECL (Extended CL, not to be confused with E-CL which is Epistemic CL [21]), with the operator $\langle\![A*]\!\rangle\varphi$, where $A \subseteq \mathcal{A}$. $\langle\![A*]\!\rangle\varphi$ is read as "A has a strategy of maintaining φ". ATL also subsumes ECL by translating the aforementioned formula to $\langle\!\langle A\rangle\!\rangle \mathrm{G}\varphi$ [30].

Let $\varphi \in \mathcal{L}_{\mathcal{CL}}$ be a formula, $\mathcal{N} = (\mathcal{S}, E, V)$ a coalition model with $s \in \mathcal{S}$. We evaluate the truth of φ in $(\mathcal{N}, s)$ by the following rules:

$$
\begin{array}{ll}
(\mathcal{N}, s) \vDash_{\mathcal{CL}} p & \text{if and only if } s \in V(p) \\
(\mathcal{N}, s) \vDash_{\mathcal{CL}} \neg\varphi & \text{if and only if } (\mathcal{N}, s) \nvDash_{\mathcal{CL}} \varphi \\
(\mathcal{N}, s) \vDash_{\mathcal{CL}} \varphi \wedge \psi & \text{if and only if } (\mathcal{N}, s) \vDash_{\mathcal{CL}} \varphi \text{ and } (\mathcal{N}, s) \vDash_{\mathcal{CL}} \psi \\
(\mathcal{N}, s) \vDash_{\mathcal{CL}} \langle\!\langle A \rangle\!\rangle \varphi & \text{if and only if } [\![\varphi]\!]^{\mathcal{N}} \in E_s(A)
\end{array}
$$

2.7.3 Truly playable games

In this section we give a proof following the idea of Pauly [53] and Goranko, Jamroga and Turrini [31], but the construction we give is slightly modified. The reason we repeat it here is threefold. The first reason is that we will use the very same technique to prove a different theorem for the epistemic coalition frames. The second reason is that the construction is slightly altered, and even though the reader is familiar with the original construction, it might be fruitful to read this construction before reading the proof of our theorem for epistemic coalition frames. Last, but definitely not least, nowhere in the literature is this construction and proof written out in details. The first published result, by Pauly, was missing a property and the proof did not quite go through, and in the proof by Goranko et al., they prove the theorem and analyse Pauly's paper in parallel, which makes the proof harder to read if you are interested in the proof alone.

Before starting this section, the reader is highly encouraged to recall Definitions 2.13 and 2.15. The entire section is base on the notion of a truly playable effectivity function, and we will prove that ever truly playable effectivity function is the α-effectivity function (Definition 2.41) of a strategic game.

Theorem 2.46 (Pauly's representation theorem)**.** *A coalitional effectivity function E α-corresponds to a nonempty strategic game if and only if E is truly playable.*

The *only if* part is trivial and given a game $\mathcal{G}$, the natural α-effectivity function $E^{\mathcal{G}}$ is easily seen to be truly playable. So we will focus on the *if* part. Let $\mathcal{C} = (\mathcal{S}, E)$ be a coalition frame. We construct a set of strategies for each agent $i \in \mathcal{A}$ on the form

$$\sigma_i = (B_i, t_i, h_i),$$

where $B_i \subseteq \mathcal{S}$, is the outcome set i wants to have enforced, $t_i \in \mathcal{A}$ is an agent and $h_i : 2^{\mathcal{S}} \to \mathcal{S}$ with the restriction that $h_i(X) \in X$. Each action thus consists of three parts, and we will define them one by one, then define the outcome function, and finally we will prove that in a game with these actions and the outcome function, the game will α-correspond to $\mathcal{C}$.

But first, to justify part three of the construction, we start by introducing the concept of a crown.

Definition 2.47 (Crown)**.** An effectivity function $E : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ is a *crown* if and only if $X \in E(\mathcal{A})$ implies $\{x\} \in E(\mathcal{A})$ for some $x \in X$. ⊣

This can be read as the *reason* that $\mathcal{A}$ is effective for X is in fact that they are effective for an element x in X, and X is in the upset of $\{x\}$. Moreover, we have the following relationship between truly playable effectivity functions and crowns.

Proposition 2.48 (Goranko et al.). *The following are equivalent for every playable effectivity function* $E : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$.

1. *E is truly playable,*
2. *$E(\emptyset)$ has a non-empty nonmonotonic core,*
3. *$E^{nc}(\emptyset)$ is a singleton and $E(\emptyset)$ is a principal filter, generated by $E^{nc}(\emptyset)$ and*
4. *E is a crown.*

We will now go through each part step by step, and see how we construct the actions and outcome function. It is assumed that the agent set is of the form $\mathcal{A} = \{1, 2, \ldots, n\}$.

Part one is the Bs. Any agent can choose any subset of $\mathcal{S}$ she likes. However, if she chooses a too small set, it will "default" to $\mathcal{S}$ itself. Given $B_1, \ldots, B_n$, we partition the agents into coalitions of agreeing Bs. We let $A_1, \ldots, A_k$ be the partitioning of agents with the property that i and j are in the same partition if and only if $B_i = B_j$. Now, we define, for each partition $A_l \in \{A_1, \ldots, A_k\}$,

$$B_{A_l} = \begin{cases} B_i & \text{for some } i \in A_l \text{ if } B_i \in E(A_l) \\ \mathcal{S} & \text{otherwise.} \end{cases}$$

This means that each partition either votes for some set they are effective for according to E, or they vote for $\mathcal{S}$, which by safety, every partition is effective for. Since $A_1, \ldots, A_k$ are disjoint, and each A_l is effective for B_{A_l}, by finite superadditivity, $\mathcal{A}$ is effective for $\bigcap_{1 \leq l \leq k} B_{A_l} = B_{\mathcal{A}}$. The $B_{\mathcal{A}}$ is the resulting *outcome set*. The next two parts lets some chosen agent choose which of the elements in the outcome set will be the real outcome.

Remark 2.49. The role of the Bs are to give each coalition *enough* power to force a set they should be able to force. If we want to show that a coalition A has enough power, i.e. they are able to force X in $\mathcal{C}$ means they should be able to force X in the new game, we consider the collective A-action where each $i \in A$ has $B_i = X$. By this move, we are guaranteed that the outcome lies in X. Each (nonempty) coalition thus has enough power. ⊣

Part two is the *designated voter*. Each agent votes for an agent, meaning that each agent chooses some number $t_i \in \{1, \ldots, n\}$. Then we define the designated voter $i_0 = (\sum_{i \in \mathcal{A}} t_i \mod |\mathcal{A}|) + 1$. The important part of this way of picking an agent is that the only coalition that can force who the designated voter will be is the grand coalition. Each strict sub coalition has equally much power.

Part three is the selector function. Each agent picks some function $h \in \mathcal{H}$. This function will simply pick the element that is the final outcome. The designated voter gets to chose which selector function we use.

$$\mathcal{H} = \{h : 2^{\mathcal{S}} \setminus \emptyset \to \mathcal{S} \mid h : X \mapsto \{x\} \text{ for some } x \in X \text{ and } x \in E(\mathcal{A}) \text{ for all } X \subseteq \mathcal{S}\}$$

is the set of all selector functions, and each agent chooses one such function.

Remark 2.50. Where part one ensures that a coalition has enough power, part two and three serves the purpose of making sure a coalition does not have too much power. Assume a coalition $A \subset \mathcal{A}$ wanted to force a strict subset of what they actually were effective for. This can not be done by choosing a too small B_A (since it will default to $\mathcal{S}$), and since A is not in control of who the designated voter i_0 will be, we can assume $i_0 \notin A$. And since i_0 can pick any $h \in \mathcal{H}$ she want, she can force the outcome to be any $x \in B_A$ such that $\{x\} \in E(\mathcal{A})$. In addition, in the case $A = \mathcal{A}$, then A already must choose some x such that $\{x\} \in E(A)$. ⊣

Finally the outcome function takes a collective action $\sigma_{\mathcal{A}}$. First it calculates the outcome set, which is $B_{\mathcal{A}}$ as described in part one, then it calculates the designated voter i_0 as described in part two, i.e. $i_0 = (\sum_{i \in \mathcal{A}} t_i \mod |\mathcal{A}|) + 1$. Finally it returns the set $h_{i_0}(B_{\mathcal{A}})$.

Now we prove that the construction really works as intended.

Proof of Theorem 2.46. Let $\mathcal{C} = (\mathcal{S}, E)$ be given. Then the game $\mathcal{G}^{\mathcal{C}} = (\mathcal{S}, \Sigma^{\mathcal{C}}, o^{\mathcal{C}})$ α-corresponds to $\mathcal{C}$, where

- $\Sigma^{\mathcal{C}} = 2^{\mathcal{S}} \times \mathcal{A} \times \mathcal{H}$
- $o^{\mathcal{C}}$ is as described above: $o^{\mathcal{C}}(\sigma_{\mathcal{A}}) = h_{i_0}(B_{\mathcal{A}})$.

We will denote the α-effectivity function of $\mathcal{G}^{\mathcal{C}}$ by E^{α}, denote $\mathcal{G}^{\mathcal{C}}$ simply by $\mathcal{G}$, by Σ, we mean $\Sigma^{\mathcal{C}}$ and by o we mean $o^{\mathcal{C}}$. We prove this one direction at a time, making special cases for when we are dealing with coalitions A of the type $A = \emptyset$, $\emptyset \subset A \subset \mathcal{A}$ and $A = \mathcal{A}$. We start by showing that if $X \in E(A)$, then $X \in E^{\alpha}(A)$.

$\mathbf{X} \in \mathbf{E}(\mathbf{A}) \Rightarrow \mathbf{X} \in \mathbf{E}^{\alpha}(\mathbf{A})$. Assume $X \in E(A)$ for some $\emptyset \subset A \subseteq \mathcal{A}$. Then A are effective for X in $\mathcal{G}$ since A have the possibility of choosing $B_A = X$. A are then guaranteed to end up in the same partition, and since they are effective for X, by coalition monotonicity, any superset of A is effective for X, so $B_{\mathcal{A}} \subseteq X$. Hence, for any designated voter i_0, and any selector function h_{i_0}, $h_{i_0}(B_{\mathcal{A}}) \in X$.

Assume that $X \in E(\emptyset)$. We must prove that any outcome must lie in X. Since $X \in E(\emptyset)$ if and only if $X \supseteq Z$ for $E^{nc}(\emptyset) = \{Z\}$ (Proposition 2.48) we can instead prove that o hits only Z. This follows trivially since $Z = \{x \mid \{x\} \in E(\mathcal{A})\}$ (Proposition 2.42).

$\mathbf{X} \in \mathbf{E}(\mathbf{A}) \Leftarrow \mathbf{X} \in \mathbf{E}^{\alpha}(\mathbf{A})$. Suppose $A = \mathcal{A}$ and contrapositively that $X \notin E(A)$. Then, by $\mathcal{A}$-maximality, $\overline{X} \in E(\emptyset)$. By the above part of the proof, $\overline{X} \in E^{\alpha}(\emptyset)$, and hence $X \notin E^{\alpha}(A)$.

Let $\emptyset \subseteq A \subset \mathcal{A}$ and assume $X \notin E(A)$. We must show that $\overline{A}$ has a strategy such that $o(\sigma_A \sigma_{\overline{A}}) \notin X$. Let $\forall i \in \overline{A}$ have $B_i = \mathcal{S}$, furthermore, we can also assume $i_0 \in \overline{A}$. Now, since $X \notin E(A)$, $B_A \not\subseteq X$. What remains to show is that there is an $s \in B_A = B$ such that $s \notin X$ and $\{s\} \in E(\mathcal{A})$. Then we can have $h_{i_0} : Y \mapsto s$ for all $Y \subseteq \mathcal{S}$ such that $s \in Y$, and whatever otherwise. We can without loss of generality assume that all $j \in A$ had $B_j = B$, and furthermore we have that $B \in E(A)$.

Claim 2.51. $B \cap Z \in E(A)$ *for* $Z = \{x \mid \{x\} \in E(\mathcal{A})\}$

Since $X \notin E(A)$, $B \cap Z \not\subseteq Z$ (since $E(A)$ is upset), hence there is an $s \in (B \cap Z) \setminus X$ such that $\{s\} \in E(\mathcal{A})$. This concludes the proof. □

Proof of Claim 2.51. We prove that for any truly playable effectivity function E that if $X \in E(A)$ and $Z = \{x \mid \{x\} \in E(\mathcal{A})\}$ then $X \cap Z \in E(A)$. But this holds trivially by superadditivity. $A \cap \emptyset = \emptyset$ and since $Z \in E(\emptyset)$ and $X \in E(A)$, $X \cap Z = E(A \cup \emptyset) = E(A)$. □

Part II

Group knowledge in neighbourhood semantics

We have seen in Section 2.4 the main ideas for epistemic logic, and also how to interpret epistemic logic in relational models. This is by far the most used semantics for epistemic logic. We will in the following two chapters show how we can use neighbourhood semantics to model the same logic, i.e. we will give a new semantic definition that is sound and complete with respect to the standard axiomatization for group knowledge, just as relational structures are. This is no surprise, since we have already mentioned that neighbourhood semantics is a proper generalisation of relational semantics. However, showing in details how we can interpret epistemic logics in neighbourhood systems will help us define epistemic effectivity functions later, as we do in Chapter 5.

Since distributed knowledge is our main interest throughout this thesis, we will in the next chapter analyse, in detail, distributed knowledge in neighbourhood semantics. Finally, before diving into the discussion on epistemic effectivity, we will also analyse *mutual* and *common knowledge* in neighbourhood systems.

Chapter 3

Distributed knowledge in neighbourhood semantics

Recall that we have fixed $\mathcal{A}$ to be a finite set of agents and Φ to be an at most countable set of atomic propositions. We defined an S5 relational model as a triple $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$, where $\mathcal{S}$ is any set of states, $\sim_i \subseteq \mathcal{S}\times\mathcal{S}$ is an equivalence relation for all $i \in \mathcal{A}$ and $V : \Phi \to 2^{\mathcal{S}}$ is an assignment function. We will now discuss distributed knowledge, and as such, we define the "coalitional" indistinguishability relation for $A \subseteq \mathcal{A}$ as

$$\sim_A = \bigcap_{i\in A} \sim_i$$

and note that this is an equivalence relation as well. For $s \in \mathcal{S}$ and $A \subseteq \mathcal{A}$, by $[s]_{\sim_A}$ we mean A's equivalence class around s, i.e. $[s]_{\sim_A} = \{t \in \mathcal{S} \mid t \sim_A s\}$. Furthermore, we observe that $[s]_{\sim_A} = \bigcap_{i\in A}[s]_{\sim_i}$.

We define the language $\mathcal{L}_{\mathcal{D}}$:

$$\varphi := p \mid \neg\varphi \mid \varphi\wedge\varphi \mid \mathrm{D}_A\varphi$$

where $p \in \Phi$ and $A \subseteq \mathcal{A}$. The formula $\mathrm{D}_A\varphi$ is read as "φ is distributed knowledge for A". It is standard to also add the K_i modality per agent, but we will omit it here and simply note that $\mathrm{D}_{\{i\}}\varphi$ is the same, in every way, as $\mathrm{K}_i\varphi$, and can therefore define K_i as such.

The definition of truth of $\mathcal{L}_{\mathcal{D}}$-formulae is defined in *pointed* relational models, i.e. if $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$ is a relational model with $s \in \mathcal{S}$, then $(\mathcal{M}, s)$ is a pointed relational model. The truth definition is completely standard:

$(\mathcal{M}, s) \vDash_{\mathcal{RS}} p$	if and only if $s \in V(p)$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\varphi$	if and only if not $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi\wedge\psi$	if and only if both $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$ and $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \psi$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{D}_A\varphi$	if and only if for all $t \in \mathcal{S}$ such that $t \sim_A s$ $(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$

Recall Definition 2.19, that for $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$ we defined $[\![\varphi]\!]^{\mathcal{M}} = \{s \in \mathcal{S} \mid (\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi\}$ to be the truth set of φ.

Proposition 3.1. *Let $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$ be given. Then $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{D}_A\varphi$ if and only if $[s]_{\sim_A} \subseteq [\![\varphi]\!]^{\mathcal{M}}$*

Proof.

$$
\begin{aligned}
&(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{D}_A\varphi \\
&\qquad\Leftrightarrow \forall t \sim_A s\ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi \\
&\qquad\Leftrightarrow \forall t \in [s]_{\sim_A}\ (\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi \\
&\qquad\Leftrightarrow \forall t \in [s]_{\sim_A}\ t \in [\![\varphi]\!]^{\mathcal{M}} \\
&\qquad\Leftrightarrow [s]_{\sim_A} \subseteq [\![\varphi]\!]^{\mathcal{M}}
\end{aligned}
$$

□

Corollary 3.2. *If $\mathcal{F}$ is the principal filter on $2^{\mathcal{S}}$ generated by $[s]_{\sim_A}$ then $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{D}_c\varphi$ if and only if $[\![\varphi]\!]^{\mathcal{M}} \in \mathcal{F}$.*

Proof. $\mathcal{F}$ is a filter generated by $[s]_{\sim_A}$ means that $\mathcal{F}$ consists of *exactly* the supersets of $[s]_{\sim_A}$, hence $[s]_{\sim_A} \subseteq [\![\varphi]\!]^{\mathcal{M}}$ if and only if $[\![\varphi]\!]^{\mathcal{M}} \in \mathcal{F}$. □

Recall furthermore that we defined a neighbourhood model as a triple $\mathcal{N} = (\mathcal{S}, (\nu_s)_{s\in\mathcal{S}}, V)$, where $\nu_s : 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ for all $s \in \mathcal{S}$. The following definition puts necessary and sufficient restrictions on the class of neighbourhood frames for them to have modally equivalent relational frames.

Definition 3.3 (Distributed knowledge effectivity function)**.** $(\nu_s)_{s\in\mathcal{S}}$ is a family of distributed knowledge effectivity functions given that the following conditions hold:

- For all $i \in \mathcal{A}$ and $s \in \mathcal{S}$, $\nu_s(\{i\}) = Z \uparrow^{\mathcal{S}}$ for some $Z \subseteq \mathcal{S}$ with
  - $s \in Z$ and
  - for all $t \in Z$ we have that $\nu_s(\{i\}) = \nu_t(\{i\})$.
- For all $s \in \mathcal{S}$ and for all A such that $\emptyset \subset A \subseteq \mathcal{A}$, $\nu_s(A) = (\bigcap_{i\in A} Z_s^i) \uparrow$,

where $Z_s^i \subseteq \mathcal{S}$ is the set generating $\nu_s(\{i\})$ (this is by definition of ν_s being a distributed knowledge effectivity function, well defined). We also define Z_s^A as the intersection of all the Z_s^i for $i \in A$. ⊣

In Section 2.3.3, we defined the truth of formulae φ from $\mathcal{L}_{\mathcal{D}}$ in pointed neighbourhood models. We will now repeat the definition of truth in *pointed* *distributed knowledge neighbourhood models* for clarity. Let $(\mathcal{N}, s)$ be a pointed distributed knowledge neighbourhood model.

$(\mathcal{N}, s) \vDash_{\mathcal{NS}} p$	if and only if $s \in V(p)$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \neg\varphi$	if and only if not $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \wedge \psi$	if and only if both $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$ and $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \psi$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \mathrm{D}_A\varphi$	if and only if $[\![\varphi]\!]^{\mathcal{N}} \in \nu_s(A)$,

Similar to what we observed for the relational models, we now observe that $[\![\varphi]\!]^{\mathcal{N}} \in \nu_s(A)$ if and only if $Z_s^A \subseteq [\![\varphi]\!]^{\mathcal{N}}$. This will be a very nice and helpful observation when we prove the equivalence of the two classes of models.

Proposition 3.4. *If $(\nu_s)_{s\in\mathcal{S}}$ is a family of distributed knowledge effectivity functions, then for all $s \in \mathcal{S}$, ν_s satisfies the liveness property, safety property, outcome monotonicity, and superadditivity.*

Proof. Let ν_s be given and $A \subseteq \mathcal{A}$. Liveness says that $\emptyset \notin \nu_s(A)$. Since $\nu_s(A)$ is the intersection of all Z_s^i for $i \in A$ and since every Z_s^i contains s, if $X \in \nu_s(A)$ then $s \in X$ (one might say that $\nu_s(A)$ is *reflexive*). The safety property says that $\mathcal{S} \in \nu_s(A)$, but this follows immediately from $\nu_s(A)$ being a generated filter. The outcome monotonicity is also immediate by the same reason. Finally for the superadditivity, assume $X \in \nu_s(A)$ and that $Y \in \nu_s(B)$. Then $Z_s^A \subseteq X$ and $Z_s^B \subseteq Y$. Let $Z_s^{A\cup B} = Z_s^A \cap Z_s^B$, then $X \cap Y \supseteq Z_s^A \cap Z_s^B$, so $X \cap Y \in \nu_s(A \cup B)$. $\square$

Notice that we indeed have a stronger version of the superadditivity than is usual; We do not need our coalitions to be disjoint.

Let $\mathcal{N} = (\mathcal{S}, (\nu_s)_{s\in\mathcal{S}}, V)$ be a neighbourhood model with $(\nu_s)_{s\in\mathcal{S}}$ a family of distributed knowledge effectivity functions. We define the corresponding equivalence relation for $i \in \mathcal{A}$ as $\sim_i^{\mathcal{N}} = \{(s,t) \in \mathcal{S}^2 \mid t \in Z_s^i\}$.

Proposition 3.5. *For all $i \in \mathcal{A}$, $\sim_i^{\mathcal{N}}$ is an equivalence relation.*

Proof. This follows immediately from the definition of Z_s^i; Reflexivity follows from the fact that $s \in Z_s^i$, and symmetricity and transitivity follows from the fact that if $t \in Z_s^i$, then $u \in Z_t^i$, then $Z_s^i = Z_t^i = Z_u^i$. $\square$

For the other direction, given a relational model $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$, we define the corresponding family of effectivity functions as follows: $\nu_s^{\mathcal{M}}(A) = [s]_{\sim_A} \uparrow$ for $s \in \mathcal{S}$ and $A \subseteq \mathcal{A}$.

Proposition 3.6. *For all $A \subseteq \mathcal{A}$ and $s \in \mathcal{S}$, $\nu_s^{\mathcal{M}}(A)$ is indeed a family of distributed knowledge effectivity functions.*

Proof. For singleton coalitions, it follows immediately by the fact that $[s]_{\sim_i}$ is an equivalence relation. The other coalitions follows by $[s]_{\sim_A} = \bigcap_{i\in A}[s]_{\sim_i}$. $\square$

Finally, given a relational model $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$, we define the corresponding neighbourhood model as $\mathcal{N}^{\mathcal{M}} = (\mathcal{S}, (\nu_s^{\mathcal{M}})_{s\in\mathcal{S}}, V)$, where $\mathcal{S}$ and V are the same in both models, and likewise for a distributed knowledge neighbourhood model $\mathcal{N} = (\mathcal{S}, (\nu_s)_{s\in\mathcal{S}}, V)$, we define the corresponding relational model as $\mathcal{M}^{\mathcal{N}} = (\mathcal{S}, (\sim_i^{\mathcal{N}})_{i\in\mathcal{A}}, V)$.

3.1 Representation theorem

Theorem 3.7. *Let $\varphi \in \mathcal{L}_{\mathcal{D}}$, $(\mathcal{M}, s)$ a pointed relational model and $(\mathcal{N}, t)$ a pointed distributed knowledge neighbourhood model be given. Then*

1. $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$ *if and only if* $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \varphi$ *and*
2. $(\mathcal{N}, t) \vDash_{\mathcal{NS}} \varphi$ *if and only if* $(\mathcal{M}^{\mathcal{N}}, t) \vDash_{\mathcal{RS}} \varphi$.

The proof of this theorem is actually quite simple after all our observations.

Proof. We prove both cases by induction on φ, and skipping the base case and the propositional cases; they follow trivially from the fact that the valuation function and state space are the same in both models.
1. Let $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$. So assume $\varphi = \mathrm{D}_A\psi$. We observed that the truth of the formula is equivalent to saying $[s]_{\sim_A} \subseteq [\![\psi]\!]^{\mathcal{M}}$. But by induction hypothesis, we have that $[\![\psi]\!]^{\mathcal{M}} = [\![\psi]\!]^{\mathcal{N}^{\mathcal{M}}}$. We also know that by definition, $Z_s^A = [s]_{\sim_A}$, hence $Z_s^A \subseteq [\![\psi]\!]^{\mathcal{N}^{\mathcal{M}}}$. But then, it follows that $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \mathrm{D}_A\psi$ and hence $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \varphi$.
2. Let $(\mathcal{N}, t) \vDash_{\mathcal{NS}} \varphi$ and $\varphi = \mathrm{D}_A\psi$. We know that this is the same as $Z_t^A \subseteq [\![\psi]\!]^{\mathcal{N}}$. By i.h., we have $[\![\psi]\!]^{\mathcal{N}} = [\![\psi]\!]^{\mathcal{M}^{\mathcal{N}}}$, and hence since $[t]_{\sim_A^{\mathcal{N}}} = Z_t^A$, $[t]_{\sim_A^{\mathcal{N}}} \subseteq [\![\psi]\!]^{\mathcal{M}^{\mathcal{N}}}$, hence $(\mathcal{M}^{\mathcal{N}}, t) \vDash_{\mathcal{RS}} \mathrm{D}_A\psi$ so $(\mathcal{M}^{\mathcal{N}}, t) \vDash_{\mathcal{RS}} \varphi$. □

3.2 Soundness & Completeness

3.2.1 Axiomatization

Axioms are the standard S5 axioms for distributed knowledge together with the D-axiom:

P	Instances of propositional tautologies
K	$\mathrm{D}_A(\varphi \to \psi) \to (\mathrm{D}_A\varphi \to \mathrm{D}_A\psi)$
T	$\mathrm{D}_A\varphi \to \varphi$
B	$\neg\mathrm{D}_A\varphi \to \mathrm{D}_A\neg\mathrm{D}_A\varphi$
4	$\mathrm{D}_A\varphi \to \mathrm{D}_A\mathrm{D}_A\varphi$
D	$\mathrm{D}_A\varphi \to \mathrm{D}_{A'}\varphi$ when $A \subseteq A'$
MP	From φ and $\varphi \to \psi$ infer ψ
Nec	From φ infer $\mathrm{D}_A\varphi$

Theorem 3.8 (Soundness and completeness)**.** *The axiomatization above is sound and complete with respect to distributed knowledge neighbourhood models.*

The axiomatization is known to be sound and complete with respect to relational S5 models, hence by Theorem 3.7, soundness and completeness for distributed knowledge neighbourhood models follows from part 2 and 1, respectively. The proof below is essentially the same as the proof for Theorem 4.7.

Proof of Theorem 3.8. Soundness: Let φ be a theorem of the above axiomatization. By the soundness of relational S5 models, φ is valid on the class of all relational S5 models. So assume φ was not valid on the class of distributed knowledge neighbourhood models. Then there is such a model $\mathcal{N}$ with a point s such that $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \neg\varphi$. But by Theorem 3.7, there is a pointed relational S5 model $(\mathcal{M}^{\mathcal{N}}, s)$ such that $(\mathcal{M}^{\mathcal{N}}, s) \vDash_{\mathcal{RS}} \neg\varphi$, contradicting the fact that φ is valid in $\mathcal{M}$.
Completeness: Contrapositively, assume φ is not a theorem. Then we must show that there is a pointed distributed knowledge neighbourhood model that satisfies $\neg\varphi$. But by the completeness of relational S5 models, there is a pointed relational S5 model $(\mathcal{M}, s)$ such that $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\varphi$, but then, again by Theorem 3.7, we have a pointed distributed neighbourhood model $(\mathcal{N}^{\mathcal{M}}, s)$ such that $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \neg\varphi$. $\square$

See [72, 1, 27, 32] for completeness proofs for distributed knowledge.

Chapter 4

Intermezzo: Group knowledge in neighbourhood semantic

In the previous chapter, we presented a representation theorem for distributed knowledge in neighbourhood semantics. The strategy we used was to use the equivalence classes for each agent as a basis for the agent's effectivity function, and then we constructed coalitional effectivity in the same way the coalitional indistinguishability relation is constructed, namely by taking the intersection of each agent's equivalence class. But this strategy is easily seen to work with other concepts of group knowledge. We will in this chapter give a representation theorem for the remaining two most used concepts of group knowledge.

Let $\mathcal{L}_{\mathcal{EDC}}$ be the language now defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathrm{E}_A\varphi \mid \mathrm{D}_A\varphi \mid \mathrm{C}_A\varphi$$

for $p \in \Phi$ and $A \subseteq \mathcal{A}$. We will observe that despite the fact that $\mathcal{L}_{\mathcal{EDC}}$ has a modal similarity type of three different group modalities, our relational frames will only have one relation per agent. However, in the neighbourhood semantics, we will use three different neighbourhood functions, one per modality type.

4.1 Epistemic relational models

Let $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$ be a relational model and $A \subseteq \mathcal{A}$. We define the coalitional indistinguishability relations as follows:

$$\sim^{\mathrm{E}}_A = \bigcup_{i\in A} \sim_i$$
$$\sim^{\mathrm{D}}_A = \bigcap_{i\in A} \sim_i$$
$$\sim^{\mathrm{C}}_A = (\bigcup_{i\in A} \sim_i)^*,$$

where R^* is the transitive closure. Furthermore, we use the notation $[s]_i$ to denote i's equivalence class around s, i.e. $[s]_i = \{t \in \mathcal{S} \mid s \sim_i t\}$. We also write $[s]^{\mathcal{K}}_A$ for the set $\{t \in \mathcal{S} \mid s \sim^{\mathcal{K}}_A t\}$ for $\mathcal{K} \in \{\mathrm{E}, \mathrm{D}, \mathrm{C}\}$. It is important to notice that R^{E}s and $[\cdot]^{\mathrm{E}}$s are not equivalence relations and classes, respectively; They are tolerance relations (reflexive and symmetric), but the notation makes the proofs more intuitive.

$(\mathcal{M}, s) \vDash_{\mathcal{RS}} p$	iff	$s \in V(p)$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\varphi$	iff not	$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi \wedge \psi$	iff for both $\chi = \varphi$ and $\chi = \psi$	$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \chi$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{E}_A\varphi$	iff for all $t \in \mathcal{S}$ such that $s \sim^{\mathrm{E}}_A t$	$(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{D}_A\varphi$	iff for all $t \in \mathcal{S}$ such that $s \sim^{\mathrm{D}}_A t$	$(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$
$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathrm{C}_A\varphi$	iff for all $t \in \mathcal{S}$ such that $s \sim^{\mathrm{C}}_A t$	$(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$

The definition of truth in pointed models above is the standard text book definition. We observe:

Proposition 4.1. *For all $\mathcal{K} \in \{\mathrm{E}, \mathrm{D}, \mathrm{C}\}$, the following are equivalent:*

- $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathcal{K}_A\varphi$
- $[s]^{\mathcal{K}}_A \subseteq [\![\varphi]\!]^{\mathcal{M}}$
- $[\![\varphi]\!]^{\mathcal{M}} \in [s]^{\mathcal{K}}_A \uparrow$.

Proof. Let $(\mathcal{M}, s)$ be a pointed epistemic relational model and assume that $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathcal{K}_A\varphi$. This means that for all $t \in \mathcal{S}$, if $sR^{\mathcal{K}}_A t$, then $(\mathcal{M}, t) \vDash_{\mathcal{RS}} \varphi$, which means that all $t \in \mathcal{S}$, if $sR^{\mathrm{E}}_A t$, then $t \in [\![\varphi]\!]^{\mathcal{M}}$. This set of ts is exactly $[s]^{\mathcal{K}}_A$, thus $[s]^{\mathcal{K}}_A \subseteq [\![\varphi]\!]^{\mathcal{M}}$. By the definition of upset, $[s]^{\mathcal{K}}_A \uparrow$ is the set of all supersets of $[s]^{\mathcal{K}}_A$ and clearly it follows that $[\![\varphi]\!]^{\mathcal{M}}$ is a member of that set. For the last to the first, assume $[\![\varphi]\!]^{\mathcal{M}} \in [s]^{\mathcal{K}}_A \uparrow$. This means with the same reasoning as above that $[s]^{\mathcal{K}}_A \subseteq [\![\varphi]\!]^{\mathcal{M}}$ which means that for all $t \in \mathcal{S}$ such that $sR^{\mathcal{K}}_A t$, $t \in [\![\varphi]\!]^{\mathcal{M}}$, which means that $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathcal{K}_A\varphi$. □

4.2 Epistemic neighbourhood models

As the epistemic language $\mathcal{L}_{\mathcal{EDC}}$ in the previous sections had three modal constructs, contrary to the relational model, we equip our neighbourhood frame with three effectivity functions, or neighbourhood functions, ν^{E}, ν^{D} and ν^{C}.

We call $(\mathcal{S}, \nu^{\mathrm{E}}, \nu^{\mathrm{D}}, \nu^{\mathrm{C}})$ an *epistemic neighbourhood frame* if the following conditions hold:

- $\nu_s^{nc}(\{i\}) = \{Z_s^i\}$ with
  - $s \in Z_s^i$ and
  - for all $t \in Z_s^i$ we have that $\nu_s^{nc}(\{i\}) = \nu_t^{nc}(\{i\})$
- $\nu_s^{\mathrm{E}}(A) = \bigcup_{i \in A} \nu_s^{nc}(\{i\}) \uparrow$
- $\nu_s^{\mathrm{D}}(A) = \bigcap_{i \in A} \nu_s^{nc}(\{i\}) \uparrow$
- $\nu_s^{\mathrm{C}}(A) = (\bigcup_{i \in A} \nu_s^{nc}(\{i\}))^* \uparrow$

First we note that all the $\nu_s^{\mathcal{K}}(A)$ contains $\mathcal{S}$ for all A and s. This follows by the fact that $\nu_s^{\mathcal{K}}(A)$ is a non-empty upper set defined from some base set. This is safety. In addition, it does not contain $\emptyset$, which follows from reflexivity ($s \in Z_s^i$), and which is liveness. Finally we note that it is upwards closed by definition, that is, it is outcome monotonic, and it has a complete non-monotonic core (since it is in fact defined by it). Also observe that the first item forces each $\nu_s^{nc}(\{i\})$ to induce equivalence classes over $\mathcal{S} \times \mathcal{S}$.

An *epistemic neighbourhood model*, $\mathcal{N} = (\mathcal{S}, \nu^{\mathrm{E}}, \nu^{\mathrm{D}}, \nu^{\mathrm{C}}, V)$ is an epistemic neighbourhood frame with a valuation function, and a pointed epistemic neighbourhood model is as above, $(\mathcal{N}, s)$ with the truth of formulae as follows:

$(\mathcal{N}, s) \vDash_{\mathcal{NS}} p$	iff	$s \in V(p)$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \neg\varphi$	iff not	$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi \wedge \psi$	iff for both $\chi = \varphi$ and $\chi = \psi$	$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \chi$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \mathrm{E}_A\varphi$	iff	$[\![\varphi]\!]^{\mathcal{N}} \in \nu_s^{\mathrm{E}}(A)$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \mathrm{D}_A\varphi$	iff	$[\![\varphi]\!]^{\mathcal{N}} \in \nu_s^{\mathrm{D}}(A)$
$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \mathrm{C}_A\varphi$	iff	$[\![\varphi]\!]^{\mathcal{N}} \in \nu_s^{\mathrm{C}}(A)$

4.3 Representation theorem

4.3.1 Relational models to neighbourhood models

Let $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i \in \mathcal{A}}, V)$ be an epistemic relational model. We define the effectivity functions as follows:

- $\nu^{\mathcal{M},\mathrm{E}} = \bigcup_{A\subseteq\mathcal{A}, s\in\mathcal{S}} (s, A, [s]^{\mathrm{E}}_A)\uparrow$
- $\nu^{\mathcal{M},\mathrm{D}} = \bigcap_{A\subseteq\mathcal{A}, s\in\mathcal{S}} (s, A, [s]^{\mathrm{D}}_A)\uparrow$
- $\nu^{\mathcal{M},\mathrm{C}} = \bigcup_{A\subseteq\mathcal{A}, s\in\mathcal{S}} (s, A, [s]^{\mathrm{C}}_A)^*\uparrow$

and define the corresponding epistemic neighbourhood model as

$$\mathcal{N}^{\mathcal{M}} = (\mathcal{S}, \nu^{\mathcal{M},\mathrm{E}}, \nu^{\mathcal{M},\mathrm{D}}, \nu^{\mathcal{M},\mathrm{C}}, V).$$

The functions listed above are slightly unreadable, but the triples simply say that $\nu^{\mathcal{M},\mathcal{K}}_s(A) \mapsto [s]^{\mathcal{K}}_A\uparrow$.

Proposition 4.2. *Given $\mathcal{M} = (\mathcal{S}, (\sim_i)_{i\in\mathcal{A}}, V)$, the corresponding neighbourhood model $\mathcal{N}^{\mathcal{M}} = (\mathcal{S}, \nu^{\mathcal{M},\mathrm{E}}, \nu^{\mathcal{M},\mathrm{D}}, \nu^{\mathcal{M},\mathrm{C}}, V)$ is indeed an epistemic neighbourhood model.*

Proof. We need simply to check that the properties above hold, i.e. for each of the functions, for single agents, the functions behave equivalent, and that the grouping of coalitions behave according to the rules. For $\nu^{\mathcal{K}}_s(\{i\})$, we note that since $\sim_i$ is an equivalence relation, $[s]_i$ is an equivalence class around s, and hence $\nu^{\mathcal{K}}_s(\{i\})$ is the upper set of an equivalence class and $[s]_i = [t]_i$ for all $s \sim_i t$. Furthermore, the set $[s]^{\mathcal{K}}_A$ is exactly the union, intersection and transitively closed union of its members' equivalence classes around s for $\mathcal{K} = \mathrm{E}$, D and C respectively. □

Proposition 4.3. *$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$ if and only if $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \varphi$.*

Proof. We will prove that $[\![\varphi]\!]^{\mathcal{M}} = [\![\varphi]\!]^{\mathcal{N}^{\mathcal{M}}}$ for all φ by induction on φ.

Base case $\varphi = p$

By definition, since $\mathcal{N}^{\mathcal{M}}$ has the same valuation function. $[\![p]\!]^{\mathcal{M}} = V(p) = [\![p]\!]^{\mathcal{N}^{\mathcal{M}}}$.

Case $\varphi = \neg\psi$

We have that $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\psi$ if and only if we do not have $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \psi$ if and only if we do not have (i.h.) $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \psi$ if and only if $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \neg\psi$.

Case $\varphi = \psi \wedge \chi$

$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \varphi$ iff $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \psi$ and $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \chi$ iff (i.h.) $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \psi$ and $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \chi$ iff $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \varphi$.

Case $\varphi = \mathcal{K}_A\psi$

$(\mathcal{M}, s) \vDash_{\mathcal{RS}} \mathcal{K}_A\psi$. This is true by above proposition if and only if $[\![\psi]\!]^{\mathcal{M}} \in [s]^{\mathcal{K}}_A\uparrow$ which by i.h. means $[\![\psi]\!]^{\mathcal{N}^{\mathcal{M}}} \in [s]^{\mathcal{K}}_A\uparrow$. But $[s]^{\mathcal{K}}_A\uparrow = \nu^{\mathcal{M},\mathcal{K}}_s(A)$, hence we have $[\![\psi]\!]^{\mathcal{N}^{\mathcal{M}}} \in [s]^{\mathcal{K}}_A\uparrow$ if and only if $[\![\psi]\!]^{\mathcal{N}^{\mathcal{M}}} \in \nu^{\mathcal{M},\mathcal{K}}_s(A)$ if and only if $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \mathcal{K}_A\psi$.

The last case is equivalent for all $\mathcal{K} \in \{\mathrm{E}, \mathrm{D}, \mathrm{C}\}$. $\square$

4.3.2 Neighbourhood models to relational models

Let $\mathcal{N} = (\mathcal{S}, \nu^{\mathrm{E}}, \nu^{\mathrm{D}}, \nu^{\mathrm{C}}, V)$ be an epistemic neighbourhood model. We generate $\mathcal{N}$'s corresponding epistemic relational model as $\mathcal{M}^{\mathcal{N}} = (\mathcal{S}, (\sim_i^{\mathcal{N}})_{i \in \mathcal{A}}, V)$ where $s \sim_i t$ if and only if $t \in X$ for $\{X\} = \nu_s^{\mathrm{E},nc}(\{i\})$.

Fact 4.4. *$\sim_i$ as above is an equivalence relation.*

Proof. We need to check the cases for reflexivity, symmetricity and transitivity. The reflexivity case holds by the fact that $s \in X$ for $\{X\} = \nu_s^{nc}(\{i\})$. Symmetricity and transitivity follows from the fact that $\nu_s(\{i\}) = \{X\} = \nu_t(\{i\})$ for all $t \in X$. $\square$

Proposition 4.5. *$(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$ if and only if $(\mathcal{M}^{\mathcal{N}}, s) \vDash_{\mathcal{RS}} \varphi$.*

Proof. We prove that $[\![\varphi]\!]^{\mathcal{N}} = [\![\varphi]\!]^{\mathcal{M}^{\mathcal{N}}}$ by induction on φ, but this time we skip the base case and propositional cases. The cases are equivalently for all $\mathcal{K} \in \{\mathrm{E}, \mathrm{C}, \mathrm{D}\}$.

Case $\varphi = \mathcal{K}_A \psi$
: $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \mathcal{K}_A\psi$ if and only if $[\![\psi]\!]^{\mathcal{N}} \in \nu_s^{\mathcal{K}}(A)$ if and only if (i.h.) $[\![\psi]\!]^{\mathcal{M}^{\mathcal{N}}} \in \nu_s^{\mathcal{K}}(A)$. But $\nu_s^{\mathcal{K}}(A) = [s]_A^{\mathcal{K}} \uparrow$, hence $[\![\psi]\!]^{\mathcal{M}^{\mathcal{N}}} \in \nu_s^{\mathcal{K}}(A)$ if and only if $[\![\psi]\!]^{\mathcal{M}^{\mathcal{N}}} \in [s]_A^{\mathcal{K}} \uparrow$ if and only if $\mathcal{M}^{\mathcal{N}}, s \vDash_{\mathcal{RS}} \mathcal{K}_A\psi$.

$\square$

We conclude by writing down our theorem, for the sake of convenience.

Theorem 4.6 (Epistemic relational and neighbourhood models)**.** *Given an epistemic relational model there exists an epistemic neighbourhood model which satisfies exactly the same formulae, and given an epistemic neighbourhood model there exists an epistemic relational model which satisfies exactly the same formulae.*

Proof. This follows from Proposition 4.5 and Proposition 4.3. $\square$

4.4 Soundness & completeness

4.4.1 Axiomatization

We will in the axiomatization below use $\mathrm{K}_i\varphi$, and define $\mathrm{E}_A\varphi = \bigwedge_{i \in A} \mathrm{K}_i\varphi$. Since we have finitely many agents, this is a finite conjunction.

P	Instances of propositional tautologies
K	$\mathrm{K}_i(\varphi \to \psi) \to (\mathrm{K}_i\varphi \to \mathrm{K}_i\psi)$
T	$\mathrm{K}_i\varphi \to \varphi$
KB	$\neg\mathrm{K}_i\varphi \to \mathrm{K}_i\neg\mathrm{K}_i\varphi$
$K4$	$\mathrm{K}_i\varphi \to \mathrm{K}_i\mathrm{K}_i\varphi$
CK	$\mathrm{C}_A(\varphi \to \psi) \to (\mathrm{C}_A\varphi \to \mathrm{C}_A\psi)$
CM	$\mathrm{C}_A\varphi \to (\varphi \wedge \mathrm{E}_A\mathrm{C}_A\varphi)$
CI	$\mathrm{C}_A(\varphi \to \mathrm{E}_A\varphi) \to (\varphi \to \mathrm{C}_A\varphi)$
DK	$\mathrm{D}_A(\varphi \to \psi) \to (\mathrm{D}_A\varphi \to \mathrm{D}_A\psi)$
DT	$\mathrm{D}_A\varphi \to \varphi$
DB	$\neg\mathrm{D}_A\varphi \to \mathrm{D}_A\neg\mathrm{D}_A\varphi$
$D4$	$\mathrm{D}_A\varphi \to \mathrm{D}_A\mathrm{D}_A\varphi$
DD	$\mathrm{D}_A\varphi \to \mathrm{D}_{A'}\varphi$ when $A \subseteq A'$
KD	$\mathrm{D}_{\{i\}}\varphi \leftrightarrow \mathrm{K}_i\varphi$
KC	$\mathrm{C}_{\{i\}}\varphi \leftrightarrow \mathrm{K}_i\varphi$
MP	From φ and $\varphi \to \psi$ infer ψ
Nec	From φ infer $\mathrm{K}_i\varphi$
Nec	From φ infer $\mathrm{C}_A\varphi$
Nec	From φ infer $\mathrm{D}_A\varphi$

The given axiomatization is sound and complete with respect to mutual, common and distributed knowledge in relational S5 models. Hence, by Theorem 4.6, soundness and completeness follows. The axiomatization is the combination of the axiomatization given in the previous chapter and in [67].

Theorem 4.7 (Soundness and completeness)**.** *The axiomatization above is sound and complete with respect to epistemic neighbourhood models.*

Proof. Soundness: Let φ be a theorem of the above axiomatization. By the soundness of relational S5 models, φ is valid on the class of all relational S5 models. So assume φ was not valid on the class of epistemic neighbourhood models. Then there is an epistemic neighbourhood model $\mathcal{N}$ with a point s such that $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \neg\varphi$. From Theorem 4.6, we know that there is a pointed relational S5 model $(\mathcal{M}^{\mathcal{N}}, s)$ such that $(\mathcal{M}^{\mathcal{N}}, s) \vDash_{\mathcal{RS}} \neg\varphi$, contradicting the fact that φ is valid in $\mathcal{M}$.
Completeness: Contrapositively, assume φ is not a theorem. Then we must show that there is a pointed epistemic neighbourhood model that satisfies $\neg\varphi$. But by the completeness of relational S5 models, there is a pointed relational S5 model $(\mathcal{M}, s)$ such that $(\mathcal{M}, s) \vDash_{\mathcal{RS}} \neg\varphi$, but then, again by Theorem 4.6, we have a pointed epistemic neighbourhood model $(\mathcal{N}^{\mathcal{M}}, s)$ such that $(\mathcal{N}^{\mathcal{M}}, s) \vDash_{\mathcal{NS}} \neg\varphi$. □

Part III

Epistemic effectivity functions

Chapter 5

Epistemic coalition logic

In their paper in 1997, Alur, Henzinger and Kupferman published the seminal paper (or at least start of a seminal series) on the logic they named *Alternating-Time Temporal Logic*, or ATL for short. This series of papers, continued in 1998 and 2002 [9, 10], developed a multi-agent system based on *computation tree logic*, or CTL. For an introduction to CTL, introduced in [24, 25], consult e.g. [38], chapter on branching time logic. The introduction of this logic opened a new line of research, a fusion of game theory and modal logic; using modal logic to express strategic abilities.

In these papers, and in the entire line of research, it is assumed that every agent has full knowledge of all there is to know; amongst some of the problems are that agents' action set is common knowledge and the outcome function is common knowledge. But by introducing the concept of knowledge, we can maybe get rid of these problems. Suppose that you have two games, and one of the agents does not know which of the games they actually play. This means she is not necessarily aware of what the actions the other agents have available, nor how her action will be evaluated, since she does not necessarily know which of the two outcome functions are "real".

Before we dive into the literature, we highlight some important issues that must be dealt with.

5.1 Relevant issues

In this section we briefly mention some issues that should be thought about when trying to model actions and beliefs.

5.1.1 Execute strategies or identify strategies?

When an agent is ignorant of which state she currently is situated, it is apparent that her actions she considers possible must be actually possible to execute in all states she consider herself to be in. This means that in our logic, we need a form of uniformity of actions over indistinguishable states for each agent. Formally, whenever agent i has $s \sim_i t$, i.e. she cannot discern between s and t, we need that $\delta_i(s) = \delta_i(t)$, that is, her available actions in s is the same as her available actions in t. This type of uniformity must also hold for coalitions when we discuss group knowledge (see Chapter 6). If the type of group knowledge we consider is distributed knowledge, this uniformity already holds for coalitions if it hold for single agents. However, as the reader maybe already can identify, there is a problem. When group knowledge makes the indistinguishability sets larger when the group grows, e.g. when discussing common knowledge, this uniformity is broken if we only demand uniformity on single agent level. We therefore need that for each agent, $\delta_a(s) = \delta_a(t)$ for all $t \sim^{\mathrm{C}}_{\mathcal{A}} s$ where $\sim^{\mathrm{C}}_{\mathcal{A}}$ is the common knowledge relation for the grand coalition.

Remark 5.1 (Collective uniformity over available actions). If we wish to discuss uniform actions over mutual knowledge, and we have $\mathcal{A} = \{a, b\}$ with the following model:

$$\boxed{s_1} \text{ - - - } b \text{ - - - } s_2$$

we can observe that since agent a can distinguish between s_1 and s_2, she can have different possible actions in both states, e.g. $\delta_a(s_1) = \{\alpha_1\}$ and $\delta_a(s_2) = \{\alpha_2\}$. Say at the same time that agent b has $\delta_b(s_1) = \{\beta_1, \beta_2\}$ and $\delta_a(s_2) = \{\beta_1, \beta_2\}$ (agent b needs to have the same actions available in both states, since they are indistinguishable for her). On the other hand, a problem arises if we would like to consider a collective action under mutual knowledge. Under mutual knowledge, the model looks like this:

$$\boxed{s_1} \text{ - - } \{a, b\} \text{ - - } s_2$$

For agents a and b to make a mutual collective action, agent b must assume that the actions a can do in s_1, she can also do in s_2, for otherwise she might wrongfully believe what actions a might do.

It should therefore be a criterion when modelling situations like this, that for each pair $(s, t) \in (\bigcup_{i \in \mathcal{A}} \sim_i)*$, $\delta_i(s) = \delta_i(t)$, that is, the actions must be commonly known to be applicable in every state. This might seem like a too hard restriction, and as we mention in Chapter 7, we must leave this for future work. $\dashv$

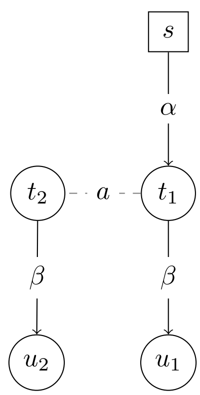
A different aspect of the uniformity of actions is the distinction between *de re* strategies and *de dicto* strategies, or *knowing how* and *knowing that*, [6, 2, 7, 36]. The problem has been pointed out in our section on ATEL, Section 2.5.4 in Example 2.38. In this example, we saw that an agent had a strategy to enforce some goal, and even knew that he had such a strategy. However, this strategy was not uniform in the sense that he did not have one strategy that worked in every possible world he considered. We call this a de dicto strategy. Consider the following example:

Example 5.2 (Robber in front of the vault)**.** A robber stands in front of a bank vault. To open the vault, a ten digit password needs to be typed. Now, the robber actually has a strategy to open this vault, and what is the strategy? To type the correct password. She has 10^{10} many strategies, and this is one of them, hence she really is in possession of such a strategy. On the other side, she does not know which strategy that makes her rich. $\dashv$

The example shows a de dicto strategy, knowing *that* she has such a strategy, whereas we often want to model de re strategies, knowing *how*, or more, knowing which strategy that leads to the wanted outcome. It should be noted that *uniform strategies* are often referred to as *conformant strategies*, as it is used in the planning community [56, 20].

5.1.2 Recollection and strategies in the long run

But identifying and executing actions are not necessarily enough to obtain a reasonable semantic definition of strategies. Recall from Section 2.5.2 that a strategy maps states to actions. A *uniform* strategy is a strategy that respects knowledge, i.e. if $s \sim_a s'$, then if f_a is a strategy for agent a, $f_a(s) = f_a(s')$. But this simple definition of uniform strategies brings up a new problem. Consider the following example:



Example 5.3. Agent a is in state s and knows that the current state indeed is s. She has one action α that she can perform, which she knows leads to the state t_1. However, there is a third state t_2 which is indistinguishable for agent a, i.e. $t_1 \sim_a t_2$. From t_1 and t_2 she can get to u_1 and u_2 respectively. Does she have a strategy in s that ensures she ends up in u_1? $\dashv$

When we said that a strategy f_i for agent $i \in \mathcal{A}$ must be uniform, we meant it in the sense that when $s \sim_i t$ then $f_i(s) = f_i(t)$. This is only when the strategy is memoryfree, i.e. the strategy is on the form $f_i : \mathcal{S} \to \Sigma$, f_i maps a state to an action. If, on the other hand, we allow memory, either bounded or unbounded, the strategies will be on the form $f_i : \mathcal{S}^* \to \Sigma$; It maps *sequences of states* to an action. If we want restrict the agents to bounded memory, we can simply choose a constant c such that f_i maps sequences of length at most c to an action.[1]

The concept of *recalling* (perfect versus imperfect) is known and has been studied in great detail, see [49, 57, 8] and is trying to deal with the difference between remembering the actions you did to get to the current state, versus simply being in a state without a clue of how you got there. In Example 5.3 above, we observe that it seems plausible that agent i has a strategy to get to u, but this seems harder to model. A way to model this might be to use dynamic models [67, 3] to remove indistinguishability relations between two states when an action has been performed, and evaluate truth of a formula in the new updated model.

5.2 A brief history

5.2.1 The start of ATEL

ATEL was introduced in the 2002 paper by van der Hoek and Wooldridge, [65] and then later in the 2003 paper by the same authors [66]. Both these two papers uses the Alternating-time Epistemic Transition Systems (AETS) as semantics, a semantic that has been under critique, e.g. by Jamroga in the 2003 paper [39], for its lack of actions, lack of a general agent set (the agent set is simply modelled by $k \in \omega$, denoting the number of agents) and problems with the transition function simply being too hard to work with. The concurrent epistemic game system (CEGS) was a way of repairing AETS in the same way the CGS improved on ATS. In the same paper, Jamroga investigates in detail the de dicto/de re distinction and uniform strategies, arguing that indeed ATEL as introduced in the two aforementioned papers suffers under highly unintuitive truths in models (recall Examples 2.38 and 5.2).

5.2.2 Imperfect recall and imperfect knowledge

Schobbens discusses in a paper from 2004 [57], four different logics, ATL$_{ir}$, ATL$_{Ir}$, ATL$_{iR}$ and ATL$_{IR}$ (the latter is really the standard ATL; In it we assume perfect information and perfect recall)[2]. The letter i versus I denotes imperfect versus perfect Information, respectively, whereas r versus R denotes imperfect versus perfect Recall. First, he draws the distinction between *complete information* and *perfect information*, saying that the former assumes full knowledge of the state space in each state, whereas the latter assumes in addition full knowledge of all choices that has happened in the past. He further explains that in the logics discussed there, the difference is inessential. The

[1]There is a problem with this simple definition of bounded memory; Why could an agent not choose to remember some longer and more important sequences instead of, say, some sequences that can never even occur?

[2]In fact, as is shown in e.g. [10, 57], ATL$_{Ir}$ is also equivalent to ATL, just the same way as ATL$_{IR}$ is. Informally, this means that ATL cannot distinguish between memorybased and memoryfree strategies.

main contribution of the paper is the implementation of explicit memory of each agent; We no longer assume each agent can base her choices on the entire past. As we mentioned in Footnote 1, Page 70, there are several aspects we must consider when incorporating bounded memory to agents. He defines four types of strategies, where $f_a : P \to \Sigma$ (where P contains some kind of information about the past and the present):

1. In IR, perfect information and perfect recall, the *view of the past* is on the form $P = \mathcal{S}^+$, where $\mathcal{S}^+$ is the sequence encoding the entire past,
2. in Ir, perfect information and imperfect recall, the view of the past $P = \mathcal{S}$ is limited to only the current state (completely memoryfree),
3. in iR, imperfect information and perfect recall, $P = \mathcal{S}_a^+$ where $\mathcal{S}_a^+$ is based on $\sim_a$ and
4. in ir, imperfect information and imperfect recall, $P = \mathcal{S}_a$; an agent can base her choice only on her current view of the world with respect to $\sim_a$.

It might be worth mentioning here that ATL_{iR} (imperfect information, perfect recall) is closely related to what we do in this thesis, but differing from our approach in Chapter 6, ATL_{iR} does not consider group knowledge. An iR-strategy, as it is called in this logic is purely individual; The uniformity is only with respect to single agents.

In their 2004 paper [41], Jamroga and van der Hoek further pushes the boundaries of ATL-like logics, by extending the language of ATEL. They introduce two new logics based on ATEL, namely Alternating-time Temporal Observational Logic (ATOL) and Alternating-time Temporal Epistemic Logic with Recall (ATEL-R*) where they discuss perfect and imperfect recall (recall Section 5.1.2). The first logic, ATOL, is a generalisation of ATL_{ir}. In this logic, agents have imperfect recall, that is they have *bounded memory.* They introduce a new modality $\langle\langle A\rangle\rangle^{\bullet}$ with the intended meaning that A "can enforce a property while their ability to remember is bounded", however, they argue that "for every model in which the agents can remember a limited number of past events, an equivalent model can be constructed in which they can recall no past at all" [41]. Hence, they consider only memoryfree strategies. The *observational* part of the name "refers to features that agents can *observe* on the spot" [41]. The following operators, or modalities, are introduced to the language to reason about observations:

$$\text{Obs}_a\varphi \quad | \quad \text{CO}_A\varphi \quad | \quad \text{DO}_A\varphi \quad | \quad \text{EO}_A\varphi,$$

where the first is an a-observation for an agent $a \in \mathcal{A}$, and the latter three are "common observation", "everybody sees" and "distributed observation" modalities for each coalition $A \subseteq \mathcal{A}$. The *observations* can then be used to make memoryfree strategies depending on, in each state, what an agent or a coalition can observe.

The logic we give in Chapter 6 is a fragment of ATOL. A straightforward translation of the DCL modality $\langle\!\![A]\!\!\rangle\varphi$ is the ATOL formula $\langle\langle A\rangle\rangle^{\bullet}_{DO(A)}\bigcirc\varphi$. The former says what A is d-effective for (effective under distributed knowledge), whereas the latter says what A can force in the next state if they are allowed to make distributed observations.

In ATEL-R*, they extend ATOL with the perfect recall strategies, and *past time operators* are introduced, i.e. modalities that allows one to talk about what the world was like before the current state; The language adds $\mathrm{X}^{-1}\varphi$ ("previously φ") and $\varphi \mathrm{S} \psi$ ("φ since ψ"). From this one can derive $\Diamond^{-1}\varphi = \top\mathrm{S}\varphi$ and $\Box^{-1}\varphi = \neg\Diamond^{-1}\neg\varphi$. The semantics of the past time fragment is evaluated in (path, number)-pairs, i.e. if $\lambda : \omega \to \mathcal{S}$ is a path and $n \in \omega$, then

$$\begin{array}{lll}
\lambda, n \vDash_{\mathcal{ATEL}} \mathrm{X}^{-1}\varphi & \text{if and only if } n > 0 \text{ and } \lambda, (n-1) \vDash_{\mathcal{ATEL}} \varphi \\
\lambda, n \vDash_{\mathcal{ATEL}} \varphi\mathrm{S}\psi & \text{if and only if there is a } k \leq n \text{ such that} \\
& \qquad \lambda, k \vDash_{\mathcal{ATEL}} \psi \text{ and } \lambda, l \vDash_{\mathcal{ATEL}} \varphi \text{ for all } k < l \leq n.
\end{array}$$

In that same paper, they proceed to explain they believe adding the past time fragment does not change ATEL-R*s expressive power. The paper contains little analysis of the logics ATOL and ATEL-R*, meaning that there is little to say about them yet. Their main result (besides introducing the logics themselves), seems to be that ATOL model checking is NP-hard (and in Δ_2). This is in contrast to ATEL, which is in P. However, this is not a surprising result as ATL_{ir} is already shown NP-hard in [57]. (Consult [59] for an introduction to complexity theory).

Otterloo and Jonker [68, 69] introduces in 2004 and 2005 papers, *strategic knowledge*. They argue that earlier attempts on "repairing" the unintuitive properties of ATEL have made logics being too strict on the restrictions of uniform strategies and indeed knowing which strategies should be used etc. They argue that "any coalition of sensible agents will choose [...] the best strategy they will come up with"[68], meaning that it should not be required that they *know* it will succeed, however, they will be effective, if the best strategy they can think of will achieve the goal. Their interpretation uses dominated strategies over extensive game forms [49]. They also discuss how recall relates to this framework.

In a 2007 paper, Jamroga and Ågotnes [40] defines the notion of *constructive knowledge*. They introduce a new logic, the Constructive Strategic Logic (CSL), in which the following language is used:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle A \rangle\rangle \bigcirc \varphi \mid \langle\langle A \rangle\rangle \Box \varphi \mid \langle\langle A \rangle\rangle \varphi \mathrm{U} \varphi \mid \mathbb{C}_A \varphi \mid \mathbb{E}_A \varphi \mid \mathbb{D}_A \varphi,$$

where $A \subseteq \mathcal{A}$ and $p \in \Phi$. The three last modalities are *constructive knowledge modalities* for coalition A (the three standard group knowledge modalities are derivable from their respective constructive group knowledge modalities; the following is a strong validity $\mathrm{C}_A\varphi \leftrightarrow \mathbb{C}_A\langle\langle\emptyset\rangle\rangle\varphi\mathrm{U}\varphi$, and likewise for the two other). The semantic for CSL is the regular CEGS, but the truth definition differs from the standard definitions; Truth of formulae is evaluated in (nonempty) *sets of states*, rather than just a single point. The standard truth definition corresponds to $(\mathcal{M}, \{s\}) \vDash \varphi$ for some $s \in S(\mathcal{M})$, however, when interpreting, recursively, the truth of the constructive knowledge modalities, we must evaluate truth of subformulae in larger sets of states. CSL is a generalisation of several strategic epistemic logics, like ATL$_{ir}$, ATOL and "Feasible ATEL"[3]. Indeed, not only does CSL in a reasonable way capture de re strategies, but it also includes the possibility of talking about de dicto strategies, and they give the following examples:

1. $\mathbb{K}_a\langle\langle a\rangle\rangle\varphi$: a has a strategy de re to enforce φ,
2. $\mathrm{K}_a\langle\langle a\rangle\rangle\varphi$: a has a strategy de dicto to enforce φ and
3. $\langle\langle a\rangle\rangle\varphi$: a has a strategy, possibly without even knowing it, to enforce φ,

where $\mathbb{K}_a = \mathbb{C}_{\{a\}}$ and K_a is the derived usual knowledge modality for a. They show that model checking CSL formulae is $\Delta_2^{\mathbf{P}}$-complete [40], meaning that they are not making big sacrifices of complexity by this general approach, as ATL$_{ir}$ is as hard. In addition, the logic is an improvement of ATOL in the way that, as we saw above, ATOL *binds* the observation modality to the strategy modality, e.g. $\langle\langle A\rangle\rangle^{\bullet}_{DO(A)}\bigcirc\varphi$, whereas in CSL, these two modalities are indeed separate modalities, and can be handled as such.

5.2.3 Irrevocable strategies

There are some pointed CGS models that satisfy formulae in ATL that seems counter intuitive. One example is given by Ågotnes, Goranko and Jamroga in the 2007 paper [3], and the example has the form $\langle\langle A\rangle\rangle\mathrm{G}\langle\langle A\rangle\rangle\mathrm{X}\varphi$. The formula is read as A has a strategy such that they always in the next step can force φ. And this is where the problem of intuition arises. The reason is that the second coalition modality asks for a *new* A-strategy, not necessarily conforming to the strategy that makes the first part true. They therefore propose a new logic, a *dynamic* logic, in which a strategy is binding; If you nest coalition modalities, you *fix* the strategy. This is closer to our intuition about extensive games, and indeed the notion of a strategy; One cannot have a strategy to ensure some property, and then change the strategy and still believe we ensure the property. They propose the logics IATL and MIATL, the former being a memoryfree and the second being a *memorybased* irrevocable ATL. The satisfiability of the nexttime fragment of IATL is based on CGSs and is defined as follows:

$$\begin{aligned}(\mathcal{M}, s) \models_{\mathcal{ATL}} \langle\langle A\rangle\rangle\mathrm{X}\varphi \text{ if and only if } & A \text{ has a collective strategy } F_A \text{ such that}\\ & \text{for all collective strategies } F_{\overline{A}},\ (\mathcal{M} \dagger F_A, o_s(F_A F_{\overline{A}})) \models_{\mathcal{ATL}} \varphi,\end{aligned}$$

where $\mathcal{M} \dagger F_A$ is a dynamic updating of $\mathcal{M}$ where we remove all actions not conforming to F_A. In this definition, memory makes a crucial difference, but we will not go into those details here.

Herzig and Troquard in 2006 [36] further elaborates on the counter intuitive properties of ATEL, before they go on arguing for how the Stit framework handles uniform strategies more gracefully. STIT is a logic expressing the statement "agent i sees to it that φ".

[3]Due to space requirements, we do not investigate feasible ATEL in this survey.

5.2.4 Seeing to it that

Stit theory was introduced in the late eighties, early nineties by Belnap, [14] and further explored by Belnap et al. [15] in 2001. In Stit, they introduce the modalities [α *stit:* Q], with the intended meaning "α *sees to it that* Q". We will stick to the more conventional way of writing it, as is done in e.g. [21], by writing $[i]\varphi$ for some formula φ (i sees to it that φ). Belnap introduces the concept of a BT model (*branching time model*), which is a tuple $\mathcal{M} = (\mathcal{T}, R, V)$, where $(\mathcal{T}, R)$ is a tree and V is a valuation function mapping $\mathcal{T}$ to sets of propositions that are true in each point. A *history* is a path h in the tree respecting the R-relation, i.e. $h : \omega \to \mathcal{T}$ such that for all $i \in \omega$, $R(h(i), h(i+1))$. A history contains $m \in \mathcal{T}$ if there is an $i \in \omega$ such that $h(i) = m$. We evaluate the truth of formulae from the STIT language in pointed BT models, where a pointed model is the model together with a (history, moment)-pair:[4]

$$(\mathcal{M}, m/h) \vDash_{\mathcal{STIT}} p \text{ iff } m/h \in V(p)$$

boolean combinations are obvious

$$(\mathcal{M}, m/h) \vDash_{\mathcal{STIT}} \mathrm{F}\varphi \text{ iff } h \text{ contains an } m',\ R(m, m') \text{ and } (\mathcal{M}, m'/h) \vDash_{\mathcal{STIT}} \varphi$$

That logic, however is a simple CTL looking logic interpreted in trees. Belnap goes on to now implement the concept of *acting*. An agent i is able to *act* in moment m from history h, and by doing so, changing the course of the future, or, more technical, agent i can in moment m change the history (remember that a history contains the future). This is implemented in what is called a BT+AC model: When $\mathcal{A}$ is a set of agents, and *Choice* is a choice function on the form $Choice : \mathcal{A} \times \mathcal{T} \to 2^{2^H}$ where H is the set of histories, we call $\mathcal{F} = (\mathcal{T}, R, \mathcal{A}, Choice)$ a BT+AC structure. Finally, with a valuation function V, we obtain $\mathcal{M} = (\mathcal{F}, V)$, a BT+AC model. We will not explain how the choice function works or behaves in detail, consult [15, 21] for a thorough introduction and analysis of the BT+AC semantics. The truth of the Stit operator, which is maybe most important, follows immediately:

$$(\mathcal{M}, m/h) \vDash_{\mathcal{STIT}} [i]\varphi \text{ iff for all } h' \in Choice_i^m(h),\ (\mathcal{M}, m/h') \vDash_{\mathcal{STIT}} \varphi,$$

where $Choice_i^m(h)$ gives the partition of $Choice_i^m$ containing h.

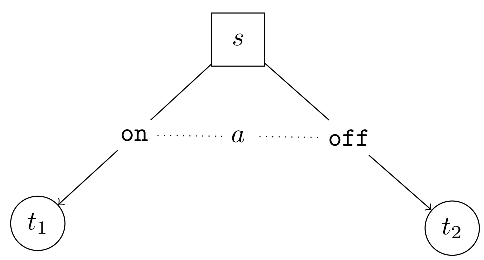
Broersen et al. introduces knowledge to Stit in 2007 [20]. In the same paper they continue the trend of exploring the relationship between coalition logic and Stit, which they started the year before in the Broersen et al., 2006 paper [18] In epistemic relations to the frames used in [19]. Broersen continues in 2009 to further expand on the notion of knowledge in Stit, by examining what an agent can "(un)knowingly do". [17].

In [20, 21], Broersen et al. argues that ATEL has inferior models to STIT logics, or mainly that CEGS (recall Definition 2.37) and MEGM are inferior to the

[4]The base case is seemingly odd, especially with respect to what is usual in modal logic: How can a *proposition* rely on a history, and not only in a moment? Belnap argues for why this could be the case but we will not explore that in this short survey.

models for E-X-Ldm^G and ENCL. They argue that having an indistinguishability relation over (state, action)-pairs has advantages over indistinguishability relations over just states.

Example 5.4 (STIT)**.** The following model models a situation where an agent a has the pairs $(s, \texttt{on})$ and $(s, \texttt{off})$ indistinguishable, that is, a knows what the world really is like, but cannot distinguish between doing the action `on` and `off`, and thus does not know which of the states t_1 or t_2 she will end up in. To model the same situation with a CEGS, we would need one extra copy of s, that is, we would need s' with actions `toggle` and `toggle`$'$ available in each of the states, and `toggle` would in each of the states lead to different outcome, as would `toggle`$'$.



⊣

We do not consider either view as more superior, the frameworks simply model knowledge and actions differently. Furthermore, each of the two approaches reflects different notions of epistemic relations and indistinguishability over actions. In this thesis, we follow the ATEL way of introducing indistinguishability, and argue that it seems strange for an agent to have actions `activate` in one state and `deactivate` when the two states are indistinguishable for her. We argue that the only thing an agent really has to hold on to is the name of the action (that is after all how she picks her choice), and the meaning of it. If the states are indistinguishable, the actions should be indistinguishable, and named e.g. `toggle-active` or simply `press-button`.

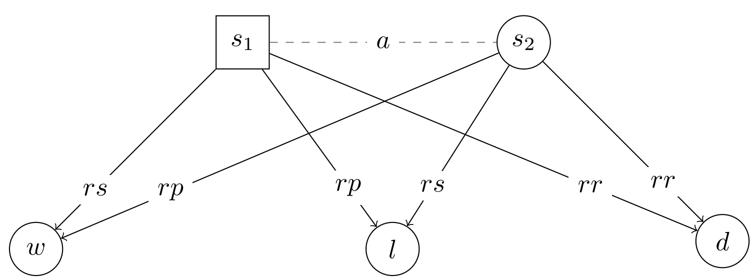
Chapter 6

Coalition logic with distributed knowledge

6.1 Motivation

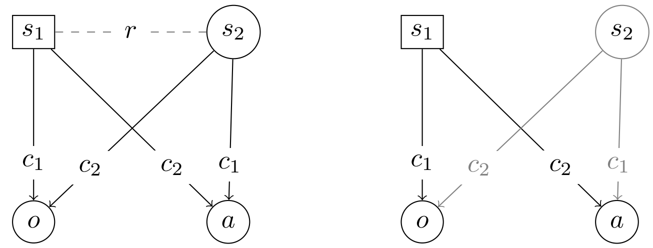
As we observed in Section 2.7 on coalition logic, and more general in Section 2.6 on game theory, in the standard strategic and coalitional games, we are using the assumption that everything is common knowledge to every agent. This means for instance that an agent knows all actions available to all the other players, what the outcome will be if the agent knows what the rest will play, and also what another agent would like to be the case (when we have a preference relation). However, this is far from always being the case. Consider two players playing Rock-Scissors-Paper, but one of the agents in fact doesn't know the rules exactly. She might be uncertain of whether `rock` beats `paper` or if it was the other way around.

Example 6.1. Consider two players, a and b, playing Rock-Scissors-Paper, but agent a cannot discern between the the two states u and v. We draw only what happens if a plays `rock`.



⊣

The above example may seem strange and artificial; Why would an agent play a game in which he does not know the rules, and in that case, why not just



(a) Individual knowledge (b) Distributed knowledge

Figure 6.1: Individual and distributed knowledge of the code that opens the vault. In the first picture, r clearly has no strategy ensuring o as outcome, and in the second picture, $\{r, b\}$ has a strategy ensuring o.

ask? The example is not as artificial as it appears, just replace the word "game" with "situation", and "rule" with "what happens when an action is performed", and we obtain situations like the one in the next example.

Example 6.2 (Robber and banker)**.** Recall Example 5.2, where a robber is standing in front of a vault. The vault has a code which is unknown to the robber. Assume the robber has actions c_i: type code i at the vault. Only the correct code c will open the vault. An incorrect attempt will trigger the alarm. Suppose that there is a banker with her who is tied to his chair. The banker knows the code, but, being tied up, has no actions available but the `noop`-action. If r is the robber and b is the banker, then we would say that $\{b, r\}$ has a strategy of opening the vault, however, neither $\{b\}$ nor $\{r\}$ has.

Figure 6.1 shows a model of a situation where there are only two codes, c_1 and c_2. The robber does not know that c_1 opens the safe and that c_2 triggers the alarm, but the banker does. It is distributed knowledge in s_1 among $\{r, b\}$ that c_1 is the action that opens the vault. $\dashv$

6.2 Preliminaries

The logic we construct will be evaluated in neighbourhood models. These models are referred to as coalition models (or frames) and when we deal with knowledge, we will evaluate truth in *epistemic* coalition models (or frames).

Definition 6.3 (Coalition Frame)**.** A coalition frame, or CF for short, is a tuple $\mathcal{C} = (\mathcal{S}, E)$ where $E : \mathcal{S} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ is a coalitional effectivity function. When $V : \Phi \to 2^{\mathcal{S}}$ is a valuation function, we refer to $\mathcal{N} = (\mathcal{S}, E, V)$ as a coalition model, or CM for short. $\dashv$

Our result is that some restricted class of the latter game frames corresponds to the former frames in respect to what coalitions can *force*.

We will stick to the assumptions we have had throughout this thesis that we have a finite set of agents $\mathcal{A}$, and we assume that it is on the form $\{1, \dots, n\}$, a nonempty set of states and outcomes $\mathcal{S}$, possibly infinite and a set of equivalence relations $\sim_i \subseteq \mathcal{S} \times \mathcal{S}$, one per agent $i \in \mathcal{A}$.

We will in this chapter only discuss *distributed knowledge*, so we define our coalitional equivalence relations as usual, i.e. $\sim_A = \bigcap_{i \in A} \sim_i$.

Remark 6.4 (Why distributed knowledge?). Distributed knowledge is closely related to our intuition of what agents know if they *share knowledge* or information. Since coalition logic is designed for reasoning about cooperation, this seems a suiting concept of knowledge to use. $\dashv$

Definition 6.5 (Coalitional knowledge)**.** Let $\mathcal{A}$ be a set of agents, $\mathcal{S}$ a set of states and $(\sim_i)_{i \in \mathcal{A}}$ be an equivalence relation for each agent. We define the coalitional equivalence relation as follows. Let $A \subseteq \mathcal{A}$, then

$$\sim_A = \begin{cases} \bigcap_{i \in A} \sim_i & \text{when } A \neq \emptyset \text{ and} \\ (\bigcup_{i \in \mathcal{A}} \sim_i)^* & \text{otherwise,} \end{cases}$$

where $\sim^*$ is the transitive closure of $\sim$. $\dashv$

This is an unconventional definition of the knowledge of the empty coalition; The empty coalition knows exactly whatever is common knowledge of $\mathcal{A}$.

6.2.1 Properties of effectivity functions

In addition to the agents, the state space (or equivalently, outcome space) and the indistinguishability relations, we also need the *coalitional effectivity functions*, which tells which sets of outcomes a coalition is effective for. Recall that given $\mathcal{A}$, $\mathcal{S}$ and $(\sim_i)_{i \in \mathcal{A}}$ as above, a function $E : \mathcal{S} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ is called a coalitional effectivity function. We will, for ease of reading (and admittedly, writing), write $E_s(A) \mapsto X$ instead of $E(s, A) \mapsto X$. Recall furthermore that in coalition logic and in the representation theorem of [53, 31], we get some restrictions on an effectivity function in order for it to have a corresponding strategic game. In this chapter, we will put some other restrictions on these effectivity functions.

Definition 6.6 (Quasi-spike)**.** If $X \in E_s(A)$ and $|X| \leq |[s]_{\sim_A}|$, we call X an *A-s-quasi-spike*. We denote by Q^s_A the set of all A-s-quasi-spikes. $\dashv$

Definition 6.7 (Quasi-crown)**.** E is a *quasi-crown* for $\mathcal{A}$ if for all $s \in \mathcal{S}$ and for all $X \in E_s(\mathcal{A})$, there is a $Y \subseteq X$ such that $Y \in E_s(\mathcal{A})$, Y is minimal under set inclusion and that Y is an $\mathcal{A}$-s-quasi-spike. $\dashv$

Definition 6.8 (Epistemically playable effectivity functions)**.** Given a set of agents $\mathcal{A}$, a set of outcomes $\mathcal{S}$, a family of indistinguishability relations $\sim_i \subseteq \mathcal{S} \times \mathcal{S}$ for $i \in \mathcal{A}$ and a coalitional effectivity function $E : \mathcal{S} \times 2^{\mathcal{A}} \to 2^{2^{\mathcal{S}}}$ we say that E is *epistemically playable* if:

- $E_s(A) = E_t(A)$ whenever $s \sim_A t$ (E respects coalitional knowledge),

- $E(A)$ is *alive*, *safe* and *upset*,
- E is superadditive,
- for all $s \in \mathcal{S}$, $E_s(\emptyset) = Q^s_\emptyset \uparrow$ and
- E is a quasi-crown for $\mathcal{A}$.

$\dashv$

We denote by Q^s the set $Q^s_\emptyset$.

Definition 6.9 (Epistemic coalition frame)**.** When E is an epistemically playable effectivity function over $(\sim_i)_{i\in\mathcal{A}}$, and $\mathcal{C} = (\mathcal{S}, E)$ is a coalition frame, we call $\mathcal{F} = (\mathcal{S}, E, (\sim_i)_{i\in\mathcal{A}})$ an epistemic coalition frame, or ECF for short. $\dashv$

Lemma 6.10. *Let $s \in \mathcal{S}$ be a point in an outcome set, $A \subseteq \mathcal{A}$ a coalition and E an epistemically playable effectivity function. If $Y \in E_s(A)$, then $(Y \cap Q^s) \in E_s(A)$.*

Proof. By definition of epistemic playability, $Q^s \in E_s(\emptyset)$ and by superadditivity, $Y \cap Q^s \in E_s(\emptyset \cup A) = E_s(A)$. $\square$

6.2.2 Game frames

We will now introduce the two different kinds of game frames we are working with, one is a strategic game frame (also referred to in the literature as a *non-cooperative* game frame) whereas the other is a coalitional game frame (referred to in the literature as *cooperative* game frames).

Recall that a CEGF, a concurrent epistemic game frame, was a tuple $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}})$. We treat each state in s as a strategic game, and in s, an agent does not know if we are playing s or t when $s \sim_i t$. The outcome function o maps collective actions to a new state $u \in \mathcal{S}$. Observe that $\mathcal{S}$ acts both as a set of outcomes, and as a set of "possible worlds"; there can be states s and t both in $\mathcal{S}$, but agent i can be unable to distinguish them, disabling her from deciding on her best available strategies in s. It is also assumed that δ_i always respects $\sim_i$, i.e. if an agent cannot distinguish said states, she will always have the same available actions in those states.[1]

Given $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}})$, we define the extended version of δ for coalitions; If $A \subseteq \mathcal{A}$ and $s \in \mathcal{S}$, $\delta_A(t) = \times_{i\in A}\delta_i(t)$. We can now say that a coalition A have a collective action σ_A in a state s if $\sigma_A \in \delta_A(s)$.

Given such a game frame, we can create a corresponding effectivity function, and we can analyse which of the aforementioned properties it has. There are many different concepts of corresponding effectivity functions in the literature, amongst them α-*effectivity* and β-*effectivity*. However, neither of them handles knowledge, and to the authors knowledge, no such effectivity function

[1] In fact, δ^s_i should always respect $\sim_{A_i}$ for all $A_i \subseteq \mathcal{A}$ with $i \in A_i$. This is only an issue that needs to be taken care of when the indistinguishability classes grow when coalitions grow, which does not happen when treating $\sim_A$ as intersection.

has been discussed in the literature. We therefore introduce a new concept of corresponding effectivity functions now, and we name it *epistemic d-effectivity*, where the d is an abbreviation for *distributed*. As is obvious below, it does not take much imagination to make epistemic e- and c-effectivity functions, if we treat coalitional knowledge as mutual and common knowledge, respectively.

Definition 6.11 (Epistemic d-effectivity)**.** Given a CEGF $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i \in \mathcal{A}})$, we say that a coalition $A \subseteq \mathcal{A}$ is *epistemic d-effective*, or simply *d-effective* in $s \in \mathcal{S}$ for a set of outcomes $X \subseteq \mathcal{S}$ if A have a collective action $\sigma_A \in \delta_A(s)$ such that for all $t \in \mathcal{S}$ such that $s \sim_A t$, for all collective actions $\sigma_{\overline{A}} \in \delta_{\overline{A}}(t)$, we have that $o_t(\sigma_A \sigma_{\overline{A}}) \in X$. ⊣

The notion of d-effectivity says that if A is a coalition d-effective for $X \subseteq \mathcal{S}$ in the state s, then they can, if they choose to cooperate, *force* the outcome of the game to be in X. This despite the fact that they are not really certain what the real world is like, nor what actions the other players might play, or even have available. So this is really a notion of *forcing* or *being effective for* which corresponds to the intuition of their ability.

It might be interesting to note that $X \in E_s(\emptyset)$ if and only if for all $t \in \mathcal{S}$ such that $s \sim_\emptyset t$, then for all $\sigma_\mathcal{A} \in \delta_\mathcal{A}(t)$ we have that $o_t(\sigma_\mathcal{A}) \in X$. Intuitively $E_s(\emptyset)$ can be read as "what is common knowledge that the outcome state must be in".

Definition 6.12 (d-effectivity function of a CEGF)**.** Given a CEGF $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i \in \mathcal{A}})$, we define the d-effectivity function $E^\mathcal{G} : \mathcal{S} \times 2^\mathcal{A} \to 2^{2^\mathcal{S}}$ as follows:

$$X \in E_s^\mathcal{G}(A), \text{ whenever } A \text{ is } d\text{-effective for } X \text{ in } s.$$

⊣

This means that $X \in E_s^\mathcal{G}(A)$ if and only if A have a collective action σ_A such that no matter what collective action the other agents choose, $\sigma_{\overline{A}}$, $o_t(\sigma_A \sigma_{\overline{A}}) \in X$ for all $t \in \mathcal{S}$ such that $s \sim_A t$.

We prove that any such effectivity function $E^\mathcal{G}$ is indeed epistemically playable.

Lemma 6.13 (d-effective functions are epistemically playable)**.** *Given a CEGF $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i \in \mathcal{A}})$, the corresponding effectivity function $E^\mathcal{G} : \mathcal{S} \times 2^\mathcal{A} \to 2^{2^\mathcal{S}}$ is epistemically playable, i.e. it satisfies the properties* liveness*,* safety*,* superadditivity*,* outcome monotonic*,* completeness of the nonmonotonic core of the empty coalition *and* quasi-crown over $\mathcal{A}$*.*

Proof. We prove this case by case.

Liveness Assume $\emptyset \in E_s^\mathcal{G}(A)$. This means that there is a collective action $\sigma_\mathcal{A}$ and a state t s.t. $o_t(\sigma_\mathcal{A}) \in \emptyset$ which is a contradiction.

Safety Since the codomain of o is $\mathcal{S}$, $\mathcal{S} \in E_s^\mathcal{G}(A)$ for all states s and all coalitions $A \subseteq \mathcal{A}$.

Superadditivity Assume that $X \in E_s^{\mathcal{G}}(A)$, $Y \in E_s^{\mathcal{G}}(B)$ and that A and B are disjoint. This means exactly that there are (disjoint) strategies σ_A and σ_B such that for all $t \sim_A s$ and all $u \sim_B s$ and for all σ_E for $E = \mathcal{A} \smallsetminus (A \cup B)$ we have that $o_t(\sigma_A \sigma_B \sigma_E) \in X$ and $o_u(\sigma_A \sigma_B \sigma_E) \in Y$. But this means that, by the definition of coalitional knowledge, for all $v \sim_{A \cup B} s$, $o_v(\sigma_A \sigma_B \sigma_E) \in X \cap Y$. It is crucial here that $\sim_{A \cup B}$ is at least as fine as each of $\sim_I$ for $I \in \{A, B\}$.

Outcome monotonicity Assume $X \in E_s^{\mathcal{G}}(A)$ and let $Y \supseteq X$. There is a collective action σ_A such that for all $t \sim_A s$ and for all $\sigma_{\overline{A}}$, $o_s(\sigma_A \sigma_{\overline{A}}) \in X$. But since $X \subseteq Y$, clearly then $o_s(\sigma_A \sigma_{\overline{A}}) \in Y$.

Completeness of nonmonotonic core We consider $E_s^{\mathcal{G}}(\emptyset)$ for some $s \in \mathcal{S}$. Since the empty coalition only has one action, namely the empty action ϵ, it is easy to see that $E_s^{\mathcal{G}}(\emptyset) = \{X \subseteq \mathcal{S} \mid \exists \sigma_\emptyset \text{ s.t. } \forall \sigma_{\mathcal{A}}. o_s(\sigma_\emptyset \sigma_{\mathcal{A}}) \in X\} = \{X \subseteq \mathcal{S} \mid \forall \sigma_{\mathcal{A}}. o_s(\sigma_{\mathcal{A}}) \in X\}$. We can also observe that $E_s^{\mathcal{G}}(\emptyset)$ is the principal filter over $\mathcal{S}$ generated by $Z_s = \bigcup_{t \sim_\emptyset s} E_t^{\mathcal{G}nc}(\mathcal{A})$.

Quasi-crown We must show that if $X \in E_s^{\mathcal{G}}(\mathcal{A})$ for some s, then there is a $Y \subseteq X$ such that $Y \in E_s^{\mathcal{G}}(\mathcal{A})$, Y is minimal under set inclusion and that $|Y| \leq |[s]_{\sim_{\mathcal{A}}}|$. Since X is in $E_s^{\mathcal{G}}(\mathcal{A})$, this means there is a $\sigma_{\mathcal{A}}$ such that for all $t \sim_{\mathcal{A}} s$, $o_t(\sigma_{\mathcal{A}}) \in X$. But since the actual outcomes is exactly $Z = \bigcup_{t \sim_{\mathcal{A}} s} o_t(\sigma_{\mathcal{A}})$, we obtain, by letting $Y = Z$, exactly the set we need.

□

In (Goranko et al.) they use the concept of a *crown* (recall Definition 2.47) instead of what we here call a quasi-crown to prove the equivalence between truly playable effectivity functions and α-effectivity in strategic games. However, as we will see very shortly, the d-effectivity of a CEGF does *not* constitute a crown.

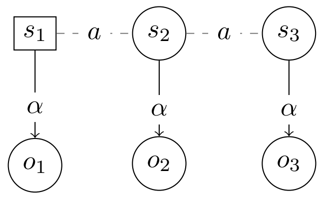
Proposition 6.14 (d-effectivity in CEGF is not a crown)**.** *Given a CEGF $\mathcal{G}$, $E_s^{\mathcal{G}}$ is* not *a crown.*

Proof. Consider the case where we have some set of states $\mathcal{S} = \{s_1, s_2, s_3, o_1, o_2, o_3\}$, with $s_1 \sim_a s_2 \sim_a s_3$, one agent, $\mathcal{A} = \{a\}$ with one action, $\Sigma = \{\sigma\}$. Assume furthermore that $o_{s_i}(\sigma) = o_i$. Now, $\{o_1, o_2, o_3\} \in E_{s_1}(\{a\})$, but no strict subset is.

Now, the crown property says that whenever it is the case that $X \in E_s^{\mathcal{G}}(\mathcal{A})$, then the grand coalition $\mathcal{A}$ can indeed force some $x \in X$ to be the case. However, $\mathcal{A} = \{a\}$ cannot force o_1, o_2 nor o_3, simply because she does not know which outcome playing σ leads to. □

In the proof above we constructed an CEGF with a "minimal" choice set of size three, but this construction is easily modified to create arbitrarily large choice sets, and when an equivalence class is infinite, we can make infinitely large minimal choice sets. We give here the model described above of a minimal choice set of size three, which shows how to force such sets:

Example 6.15. Below is depicted an example with $\mathcal{A} = \{a\}$ being effective for $X = \{o_1, o_2, o_3\}$ and no smaller set.



$\dashv$

The reason why having a crown is a nice thing for each choice set, is that we can construct strategies in a way that they pick only the spikes in the crown. We can however not do it in this way. If we have that A has X as a minimal choice set in s, with X not a singleton, we need to make sure that there are equivalent states t where A pick a different spike of the X-quasi-crown.

Proposition 6.16 (Maximum size of minimum choice set for $\mathcal{A}$)**.** *Given a CEGF $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}})$, and a state $s \in \mathcal{S}$, the size of a minimal choice set in $E_s^{\mathcal{G}}(\mathcal{A})$ is at most $|[s]_{\sim_\mathcal{A}}|$, the size of the $\mathcal{A}$-equivalence class around s.*

Proof. Let $X \in E_s^{\mathcal{G}}(\mathcal{A})$ be minimal (under set inclusion). Then there is an action $\sigma_\mathcal{A}$ such that for all $t \sim_\mathcal{A} s$ we have $o_t(\sigma_\mathcal{A}) \in X$. Since for each t, $o_t(\sigma_\mathcal{A})$ is a single element $X = \bigcup_{t\sim_\mathcal{A} s} o_t(\sigma_\mathcal{A})$, there is a surjective function mapping $[s]_{\sim_\mathcal{A}}|$ onto X. The result follows. $\square$

6.3 Representation theorem

6.3.1 Strategic to coalitional

Let $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}})$ be a CEGF. We define the epistemic coalition game frame $\mathcal{F}^\mathcal{G} = (\mathcal{S}, E, (\sim_i)_{i\in\mathcal{A}})$ where $\mathcal{S}$ and $\sim_i$ are as usual, and $\mathcal{S} \supseteq X \in E_s(A)$ for $A \subseteq \mathcal{A}$ and $s \in \mathcal{S}$ if and only if there is a collective action σ_A such that for all collective actions of $\overline{A}$, $\sigma_{\overline{A}}$, and for all t such that $s \sim_A t$, $o_t(\sigma_A\sigma_{\overline{A}}) \in X$, i.e. if and only if A is d-effective for X in s.

It follows from the above lemma that E is epistemically playable, and by construction, the following lemma is trivial.

Lemma 6.17. *Let $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i\in\mathcal{A}})$ be a CEGF. The corresponding frame $\mathcal{F}^\mathcal{G} = (\mathcal{S}, E, (\sim_i)_{i\in\mathcal{A}})$ is an epistemic coalition frame where E respects distributed knowledge. Furthermore for all $s \in \mathcal{S}$, $A \subseteq \mathcal{A}$ and all $X \subseteq \mathcal{S}$, A is d-effective for X in $\mathcal{G}$ and s if and only if $X \in E_s(A)$.*

Proof. We have proved earlier that E indeed is an epistemically playable effectivity function, hence $\mathcal{F}^\mathcal{G}$ is an ECF. What remains is to show that for all $s \in \mathcal{S}$, $A \subseteq \mathcal{A}$ and all $X \subseteq \mathcal{S}$, A is d-effective in $\mathcal{G}$ and s if and only if $X \in E_s(A)$. But this follows immediately from Definition 6.11. $\square$

6.3.2 Coalitional to strategic

Let $\mathcal{F} = (\mathcal{S}, E, (\sim_i)_{i\in\mathcal{A}})$ be any epistemic coalition frame. We will construct an epistemic strategic game frame in which the d-effectivity corresponds strictly to E. We will follow the proof from Pauly and Goranko et al. to some extent, but the lack of $\mathcal{A}$-maximality and the crown property, and the knowledge needs to be taken care of, that is we must for every state $s \in \mathcal{S}$ create strategies, instead of just for one state, and we need to make sure that if $s \sim_A t$ and there is some set X with more than one element, with $X \in E_s(A) = E_t(A)$, we create A-actions that in s sends A to one of the elements of X and in t sends A to another.

We will create the CEGF $\mathcal{G}^{\mathcal{F}} = (\mathcal{S}^{\mathcal{F}}, \Sigma^{\mathcal{F}}, \delta, o, (\sim_i^{\mathcal{F}})_{i\in\mathcal{A}})$, where

- $\Sigma^{\mathcal{F}}$ will be described in detail below,
- $\mathcal{S}^{\mathcal{F}} = \mathcal{S} \cup \bigcup_{s\in\mathcal{S}} \{s\} \times Q^s$,
- $\sim_i^{\mathcal{F}} = (\sim_i \cup \bigcup_{q\in Q^s}(s,q))^*$,
- δ_i maps states to a set of actions for each $i \in \mathcal{A}$ and
- o is the outcome function, mapping states and collective actions to an outcome state

To give an intuition of the extension of the state space, every point $s \in \mathcal{S}$ has attached $|Q^s|$ many states that nobody can distinguish between, i.e. $s \sim_{\mathcal{A}}^{\mathcal{F}} q_1^s \sim_{\mathcal{A}}^{\mathcal{F}} q_2^s \sim_{\mathcal{A}}^{\mathcal{F}} \dots$ for $q_l^s \in Q^s$ for all $l \leq |Q^s|$. Informally, we need this construction to make a the coalitions dumb enough to only being able force sufficiently large sets. Note that since no agent can distinguish between s and the Q^s many attached states, we can talk about only s, just making sure that the actions are the same in indistinguishable states.

Construction. We assume we have fixed an ECF $\mathcal{F} = (\mathcal{S}, E, (\sim_i)_{i\in\mathcal{A}})$. We will now construct a family of actions, per agent and per state, and finally an outcome function. The constructed game, $\mathcal{G}^{\mathcal{F}}$ will have a corresponding d-effectivity function $E^{\mathcal{G}^{\mathcal{F}}}$ that corresponds to E. The actions available for player i in state s will be on the form

$$\sigma_i^s = (B_i^s, t_i^s, e_i^s).$$

The first component of the action, B_i^s is simply a set of states that agent i wants to force. However, she should not pick any arbitrary set; She has to make sure that the agents in the coalition she wants to cooperate with all chooses a set B_A^s that they are effective for. If they choose a set they are not effective for, their actual coalitional choice will be $\mathcal{S}$, that is, not constraining the outcome at all.

Given a family of such sets, $B_{\mathcal{A}}^s = (B_i^s)_{i\in\mathcal{A}}$, we will collect all coalitions of agents that votes for the same set, i.e. we partition the agents of $\mathcal{A}$ into partitions $A_1, \dots, A_k$ such that two agents are in the same coalition if and only if they chose the same set. Now, we simply define the coalitional choice of such a partition to be

$$B^s_{A_l} = \begin{cases} B^s_i & \text{for some } i \in A_l \text{ whenever } B^s_i \in E_s(A_l) \text{ and} \\ \mathcal{S} & \text{otherwise,} \end{cases}$$

and we let $B^s = \bigcap_{A_1,\ldots,A_k} B^s_{A_l}$.

Proposition 6.18. *For any collective action for A, the resulting coalitional choice set of A, $B^s_A \in E_s(A)$.*

Corollary 6.19. *By superadditivity, if $A_1, \ldots, A_k$ are partitions of $\mathcal{A}$ (not necessarily covering the entire $\mathcal{A}$), $B^s_{A_1,\ldots,A_k} = \bigcap_{i\leq k} B^s_{A_i} \in E_s(\bigcup_{i\leq k} B_{A_i})$. By liveness, we get that $B^s \neq \emptyset$.*

Before we go on with the construction, it might be fruitful to stop for a while and see how a coalition can use these B^ss to force an outcome in the generated CEGF.

Example 6.20 ($A \subseteq \mathcal{A}$ forcing outcome X in s)**.** Let $A \subseteq \mathcal{A}$ be a coalition. To show how A will force X for some $X \in E_s(A)$, we let all agents $i \in A$ choose $B_i = X$. Since A all vote for the same set, they all end up in the same partition $A' \supseteq A$. By definition of B^s, and coalition monotonicity, $X = B^s_{A'} \subseteq B^s$. What will be evident later is that A has now effectively forced the outcome to be in X. ⊣

The rest of the construction is to make sure a coalition doesn't get too powerful. Recall that for each agent i, the action available in s will be on the form $\sigma^s_i = (B^s_i, t^s_i, e^s_i)$. We are done with the first component.

The second component of the action is simply the number of an agent, i.e. $t^s_i \in \mathcal{A}$, or, wlog a number in $[1, |\mathcal{A}|]$. If $t_{\mathcal{A}} \in \mathcal{A}^n$ is an n-tuple of agents, we define its *value* to be $t = (\Sigma^n_{i=1} t_i \bmod n) + 1$. Observe that this value will be a number in $[1, |\mathcal{A}|]$, which means we can again associate a tuple $t_{\mathcal{A}}$ with an agent in $\mathcal{A}$. We will let this agent be our designated voter, and we will always call her i_0; she will be allowed to choose how the outcome will be in the different states, as long as she keeps it in B^s. It is important to note that *any* agent can swing this designated voter to any other voter. I.e. let $t_{\mathcal{A}\setminus\{i\}}$ be an $n-1$-tuple. Then, for any $j \in \mathcal{A}$, i has a t_i to choose so that $t_{\mathcal{A}} = j$. This makes sure that no coalition (except the grand coalition) can choose who the designated voter will be. This concludes the second component.

The third component of the action is a function. Let $e^s_i : \mathcal{S} \times 2^{\mathcal{S}} \to \mathcal{S}$, and $B \in E_s(\mathcal{A})$. We consider all functions of that form with the restriction that $e^s_i(t, B) \in B$, and that

$$\bigcup_{t \sim_{\mathcal{A}} s} e^s_i(t, B) \supseteq X \in E^{nc}_s(\mathcal{A}).$$

Since $B \in E_s(\mathcal{A})$, by E being a quasi-crown, $B \subseteq \mathcal{S}$ and $|[s]_{\sim_{\mathcal{A}}}| \geq |Q^s|$, we know that such an X exists, hence we can make such a function. Let E^s_i be the set of all such functions. Note that this is the point where we make use of our large set of indistinguishable states, which we have guaranteed to be as large as the set of all of Q^s.

The outcome function will now be designed with the wanted features. Let $s \in S$ be a state and $\sigma_{\mathcal{A}} = (B_i^s, t_i^s, e_i^s)_{i \in \mathcal{A}}$ be a collective action. i_0 is the designated voter, as described above, and we define the outcome of $\sigma_{\mathcal{A}}$ to be

$$o_s(\sigma_{\mathcal{A}}) = e_{i_0}^s(s, B^s)$$

This concludes the construction of the CEGF. To summarise, given $\mathcal{F}$, we let the set of players be the same, we add some outcomes to the outcome set, fix the indistinguishability relations so that they are as in the original game, except for the ignorance of this new set of states, and define the set of actions as described above. Finally the outcome function picks a designated voter, creates the B^s set and use i_0s e-function to select the final outcome.

Fact 6.21. *Let $\sigma_{\mathcal{A}} = (B_i, t_i, e_i)_{i \in \mathcal{A}}$ be a collective action. Let $B_{\mathcal{A}}^s = (B_i^s)_{i \in \mathcal{A}}$ and B^s be as described. Not only is $B^s \in E_s(\mathcal{A})$, but there is an $X \subseteq B^s$ such that $X \in E_s^{nc}(\mathcal{A})$. Since X is a quasi-spike, we know that $|[s]_{\sim_{\mathcal{A}}}| \geq |X|$, and hence there is an e that maps B over $t \sim_{\mathcal{A}} s$ onto X.*

Lemma 6.22. *Let $\mathcal{F} = (\mathcal{S}, E, (\sim_i)_{i \in \mathcal{A}})$ be an ECF. Define $\mathcal{G}^{\mathcal{F}} = (\mathcal{S}^{\mathcal{F}}, \Sigma, \delta, o, (\sim_i^{\mathcal{F}})_{i \in \mathcal{A}})$ where $\Sigma_i^s = 2^{\mathcal{S}} \times \mathcal{A} \times \mathrm{E}_i^s$, o defined as above and $\delta_i(s) = \Sigma_i^s$. Then $\mathcal{G}^{\mathcal{F}}$ is an CEGF and for all $s \in \mathcal{S}$, all $X \subseteq \mathcal{S}$ and all $A \subseteq \mathcal{A}$, $X \in E_s^{\mathcal{G}^{\mathcal{F}}}(A)$ if and only if $X \in E_s(A)$.*

Proof. We will now prove that if $X \in E_s(A)$ then $X \in E_s^{\mathcal{F}}(A)$ and if $Y \in E_t^{\mathcal{F}}(B)$ then $Y \in E_t(B)$ for all $s, t \in \mathcal{S}$, $X, Y \subseteq \mathcal{S}$ and $A, B \subseteq \mathcal{A}$.

($\subseteq$) Let $s \in \mathcal{S}$ and $A \subseteq \mathcal{A}$ be arbitrary, $A \neq \emptyset$. Assume $X \in E_s(A)$ (this is safe, by safety). We choose any strategy $\sigma_A^s = (B_i^s, t_i^s, e_i^s)_{i \in A}$ such that for all $i \in A$, $B_i^s = X$. By the partitioning criterion, and coalition monotonicity, we know that $A \subseteq A_l$ for some A_l in the partitioning, $B_{A_l}^s = X$ and hence $B^s \subseteq X$. But then, regardless of the value i_0, $e_{i_0}^s(t, B^s) \in X$ for all $t \sim_A s$. Since the strategies are uniform in the strategic game, they can for all $t \sim_A s$ choose to play σ_A, which means exactly that σ_A is a witness for X, so $X \in E_s^{\mathcal{G}^{\mathcal{F}}}(A)$.

What remains for this part is to show that if $X \in E_s(\emptyset)$, then $X \in E_s^{\mathcal{F}}(\emptyset)$. Since we know that $E_s(\emptyset)$ is a principal filter generated by Q^s, we know that $Q^s \subseteq X \subseteq \mathcal{S}$. Let $\sigma_{\mathcal{A}}$ be any collective action. Then for any partitioning $B_{\mathcal{A}}^s$, $B^s \subseteq B_{A_l}^s \subseteq Q^s$, and hence the outcome must be in X.

($\supseteq$) Let $s \in \mathcal{S}$ and $A \subset \mathcal{A}$, be an arbitrary, non-empty coalition, i.e. $A \neq \emptyset$ and $A \neq \mathcal{A}$. Assume $X \in E_s^{\mathcal{F}}(A)$ and that $X \notin E_s(A)$. We know then that $B_A^s = Y \nsubseteq X$. Since this holds for any strategy of the opponents, consider the opponents voting $B_{\overline{A}}^s = \mathcal{S}$. Then we know that $Y \in E_s(A)$, and that there is a $y \in (Y \setminus X)$. But consider then that when $i_0 \in \overline{A}$, she can let $e_{i_0}^s(s, Y) = y$, `.

It remains to prove the case for $A = \emptyset$ and $A = \mathcal{A}$. Let $X \in E_s^{\mathcal{F}}(\mathcal{A})$. Then there is a $Y \subseteq X$ such that Y is a quasi-spike and so $Y \in E_s(\mathcal{A})$. Done.

Assume, contra-positively, that $X \notin E_s(\emptyset)$. Then $Q^s \not\subseteq X$, ergo $\exists q \in (Q^s \setminus X)$. Let $Y \in E_s^{nc}(\mathcal{A})$ and let $B^s = Y \cup \{q\}$. Since Y is a quasi-spike, $[s]_{\sim_\mathcal{A}}$ is more than big enough to cover Y *and* q, so let $e_{i_0}^s$ be such that $e_{i_0}^s(s, B^s) = q$ and that the rest of the equivalence class covers Y. This shows that $X \notin E_s^{\mathcal{F}}(\emptyset)$. $\square$

We now wrap up the correspondence proof by the following theorem.

Theorem 6.23. *An effectivity function is epistemically playable if and only if it is the d-effectivity function of a CEGF.*

What might be important to notice is that we in the construction of the strategic game created a great deal of new states (we keep the general cardinality, however; if $\mathcal{S}$ was finite, $\mathcal{S}'$ will be, and if $|\mathcal{S}| = \kappa$ is infinite, then $|\mathcal{S}'| = \kappa$). However, despite the fact that there are many states in the newly created model that was not in the original, none of these states will be of any significant importance to us, not even in extensive games, i.e. when we consider games that keep moving. The reason being that the strategies we created can only hit states in the original state space $\mathcal{S}$, i.e. the newly created states are not reachable by any strategy; they are merely a tool to make the agents confused.

6.4 Knowledge is Power

In the literature, the concept of a *dummy player* is often associated with an agent who has no actions that can change the outcome, i.e. no matter what the other players chooses, the result is determined by *their* chosen actions, and his action does not affect the outcome. In coalitional games, we often measure how many coalitions an agent can "swing" from one outcome to another (e.g. Penrose-Banzhaf power index [13]). Often, the two notions above coincide, but in the framework of epistemic game frames, this is not the case. We can indeed have an agent who has only one action, maybe the `void` action, and still be a great swing player. This is because of the knowledge, and in particular, the distributed knowledge of the coalition. Despite the fact that an agent cannot *add power*, she can *add knowledge*, and thereby making the coalition more powerful. See also [5] for a treatise of distributed knowledge and the Penrose-Banzhaf power index; They model the *power of an agent with respect to a goal formula* $\gamma \in \mathcal{L}_{\mathcal{CL}}$ and a pointed model $(\mathcal{M}, s)$. Consider $\gamma \in \mathcal{L}_{\mathcal{CL}}$. If we for all coalitions $A \subseteq \mathcal{A}$ consider whether or not $(\mathcal{M}, s) \vDash_{\mathcal{DCL}} \langle\!\langle A \rangle\!\rangle \gamma$, we can see which agents *swing* most coalitions, i.e. we call $i \in \mathcal{A}$ a *swing player* for A with respect to γ if $(\mathcal{M}, s) \nvDash_{\mathcal{DCL}} \langle\!\langle A \rangle\!\rangle \gamma$, but $(\mathcal{M}, s) \vDash_{\mathcal{DCL}} \langle\!\langle A \cup \{i\} \rangle\!\rangle \gamma$. In the figure for the next example, Figure 6.2, b is in s_1 a swing player for $\{a\}$ and for $\emptyset$ for the goal t_1 (if that were a proposition true in t_1 and only there). However, a is not a swing player for any coalition, hence we could argue b was more powerful with respect to the Penrose-Banzhaf power index.

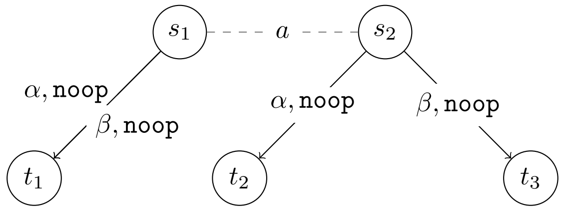


Figure 6.2: Dummy player effective for outcome

Seen in terms of the effectivity of coalitions, a consequence of the importance of knowledge is that agents and coalitions without actual power, but with great knowledge of the world are in fact modelled to be *effective* for more than dumber, yet more powerful agents and coalitions are. Consider the simple model in Figure 6.2 with two agents, $\mathcal{A} = \{a, b\}$. Consider also that b is a *dummy player* in the sense that b has only one action, namely the `noop`-action. In s_1, however, player b alone is effective for $\{t_1\}$, whereas player a is not. Player a is effective for $\{t_1, t_2, t_3\}$ only.

6.5 Logic

Let Φ again be our countable set of propositional letters and $\mathcal{A}$ the finite set of agents, then $\mathcal{L}_{\mathcal{CL}}$ (coalition logic) is the language generated by the following syntax:

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

where $p \in \Phi$ and $A \subseteq \mathcal{A}$.

Let $\mathcal{G} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i \in \mathcal{A}})$ be a CEGF. We called (Definition 2.37) $\mathcal{M} = (\mathcal{G}, V)$ a *concurrent epistemic game model* or *system*, CEGS for short, when $V : \Phi \to 2^{\mathcal{S}}$. Given $s \in \mathcal{S}$, we call $(\mathcal{M}, s)$ a *pointed CEGS* and given $\varphi, \psi \in \mathcal{L}_{\mathcal{CL}}$ and $p \in \Phi$, we evaluate the *truth* of φ, ψ in a pointed CEGS as follows:

$\mathcal{M}, s \models_{\mathcal{DCL}} p$	if and only if $s \in V(p)$
$\mathcal{M}, s \models_{\mathcal{DCL}} \neg\varphi$	if and only if not $\mathcal{M}, s \models_{\mathcal{DCL}} \varphi$
$\mathcal{M}, s \models_{\mathcal{DCL}} \varphi \wedge \psi$	if and only if $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models_{\mathcal{DCL}} \psi$
$\mathcal{M}, s \models_{\mathcal{DCL}} \langle\!\langle A \rangle\!\rangle \varphi$	if and only if $[\![\varphi]\!]^{\mathcal{M}} \in E_s^{\mathcal{G}}(A)$

As a consequence of the properties of the effectivity function, the following formulae are valid, and correspond to liveness, safety, outcome monotonicity, coalition monotonicity and superadditivity, respectively:

- $\neg\langle\!\langle A \rangle\!\rangle \bot$,
- $\langle\!\langle A \rangle\!\rangle \top$,
- $\langle\!\langle A \rangle\!\rangle (\varphi \wedge \psi) \to \langle\!\langle A \rangle\!\rangle \varphi$,

- $\langle\!\langle A \cap B \rangle\!\rangle\varphi \to \langle\!\langle A \rangle\!\rangle\varphi$ and
- $\langle\!\langle A \rangle\!\rangle\varphi \wedge \langle\!\langle B \rangle\!\rangle\psi \to \langle\!\langle A \cup B \rangle\!\rangle(\varphi \wedge \psi)$, whenever $A \cap B = \emptyset$.

We can notice that the axiom for $\mathcal{A}$-maximality, $\neg\langle\!\langle \emptyset \rangle\!\rangle\neg\varphi \to \langle\!\langle \mathcal{A} \rangle\!\rangle\varphi$ is *not* a validity in DCL (recall Figure 6.1 (a), Page 78 in which, when o is the propositional letter being true in only the state where the vault is open, $s_1 \models_{\mathcal{DCL}} \neg\langle\!\langle \emptyset \rangle\!\rangle\neg o$ and yet $s_1 \models_{\mathcal{DCL}} \neg\langle\!\langle \mathcal{A} \rangle\!\rangle o$). It should be clear, however, that in models where each agent's knowledge is perfect, i.e. $\sim_i = \{(s,s) \in \mathcal{S} \times \mathcal{S}\}$, CL coincides with DCL.

Proposition 6.24. *Let $\mathcal{M} = (\mathcal{S}, \Sigma, \delta, o, (\sim_i)_{i \in \mathcal{A}}, V)$ be a CEGS. If for all $i \in \mathcal{A}$, $\sim_i = \{(s,s) \in \mathcal{S} \times \mathcal{S}\}$, then $\mathcal{M}' = (\mathcal{S}, \Sigma, \delta, o, V)$ is a CGS with the property that for all $s \in \mathcal{S}$, $(\mathcal{M}, s) \models_{\mathcal{DCL}} \varphi$ if and only if $(\mathcal{M}', s') \models_{\mathcal{CL}} \varphi$.*

Proof. We prove this by induction on the complexity of φ, skipping the base case and boolean combinations. Since $[s]_A = \{s\}$, we simply observe that for $\langle\!\langle A \rangle\!\rangle\varphi$, the definition of truth in DCL coincides with the truth definition of CL, namely $s \models_{\mathcal{DCL}} \langle\!\langle A \rangle\!\rangle\varphi$ if and only if there is a joint action σ_A for A such that for all $\sigma_{\overline{A}}$ and all states $t \sim_A s$, $o_t(\sigma_A\sigma_{\overline{A}}) \models_{\mathcal{DCL}} \varphi$. But since s is the only state in $\sim_A$-relation to itself, this breaks down to the truth definition for CL, namely $s \models_{\mathcal{DCL}} \langle\!\langle A \rangle\!\rangle\varphi$ if and only if there is a joint action σ_A for A such that for all $\sigma_{\overline{A}}$ $o_s(\sigma_A\sigma_{\overline{A}}) \models_{\mathcal{DCL}} \varphi$. $\square$

The rules *modus ponens* and *equivalence* has the following form (respectively)

$$\frac{\varphi \qquad \varphi \to \psi}{\psi} \qquad\qquad \frac{\varphi \leftrightarrow \psi}{\langle\!\langle A \rangle\!\rangle\varphi \leftrightarrow \langle\!\langle A \rangle\!\rangle\psi}$$

for all $A \subseteq \mathcal{A}$. Together with $\mathcal{A}$-maximality, and the two rules, the validities listed above make the logic CL sound and complete with respect to coalition models [53]. A natural question is if the validities above with the two rules make the logic DCL and are sound and complete with respect to epistemic coalition models.

Part IV

Discussion and conclusion

Chapter 7

Discussion and conclusion

7.1 Comparison to earlier work

This thesis has been based heavily on the work of Pauly [53] and Goranko et al. [31] for the neighbourhood semantic part. We also based notions of effectivity on ideas occurring in both [40] and [21], furthermore we tried to model a de re form of effectivity, contrary to what ATEL does [65]. Much of the intuition is based on experiences gained from modelling knowledge in ATL-like systems, e.g. ATEL, ATL_{ir} [57], ATOL [41] and STIT [21].

7.2 Conclusion

This thesis introduced three main results the author has been unable to find in the literature, in two different, but closely related fields of modal logic. The first two results are part of an analysis of *distributed knowledge* in neighbourhood semantics. We prove a representation theorem between the relational semantics of distributed knowledge and the neighbourhood semantics. We then proceed to give a sound and complete axiomatisation for distributed knowledge in neighbourhood semantics.

The final contribution was the notion of an *effectivity function* which respects knowledge. We focused on distributed knowledge, and called this d-effectivity and left the two other standard concepts of group knowledge as possible topics future research, that is, mutual knowledge and common knowledge.

We have given a new semantic for coalition logic, a semantic that incorporates knowledge to the models used usually in CL. The logic obtained, DCL, from the validities of these new models lack $\mathcal{A}$-maximality, which is valid in CL. We also made a relational semantic for DCL, but used the models CEGS that are known and have been studied in several papers. We gave a representation theorem linking CEGSs and epistemically playable coalition models.

7.3 Future work

Complexity Complexity of a logic is a highly interesting and important property of the logic. Metalogically, it tells a lot about the expressive properties of the logic, and practically, it tells us what, if anything, we can use the logic to, and how large instances of the logic are feasible, or tractable for a computer to verify or construct. However, the author will not conjecture the complexity. ATL_{iR} model checking is undecidable, but this logic contains global and until operators that increases the complexity. Model checking ATL_{ir} is $\Delta_2^{\mathbf{P}}$-complete, even for single agents [57, 22], this hints at a problem towards the memorybased strategies (the undecidable had perfect recall, the latter is based on memoryfree strategies). The satisfiability problem of coalition logic is **PSPACE**-complete [53].

Completeness Soundness and completeness for a semantic system with respect to a syntactic system is also important. Metalogically, it gives great insight in the logic to have a sound and complete axiomatic system, practically, it paves route for automated theorem proving. We have essentially defined the logic for distributed coalition logic semantically, i.e. the logic is the set of the validities on the game frames, but we did not construct a proof system (we only know that some of the validities of CL are validities in DCL, and that there are validities in CL that are not validities in DCL). These two issues are possibly the most important properties to know about a logic, we did however not investigate them fully.

Group knowledge We left open effectivity functions that complies with mutual knowledge and common knowledge. If we take the e-effectivity function E of a CEGS with mutual knowledge instead of distributed knowledge, we will note that E lacks both superadditivity and coalition monotonicity. It will however still be alive, safe and outcome monotonic. The reason we lose coalition monotonicity is that a supercoalition $B \supseteq A$ can be weaker than A because they know less. Meaning that they are indeed at least as powerful *locally*, but less powerful in their indistinguishability set. Another interesting thing about mutual knowledge which separates it from common and distributed knowledge is that the coalitional indistinguishability relation is not an equivalence relation; It lacks transitivity. This means that we do not have the strong property that if $s \sim_A b$ then $E_s(A) = E_t(A)$.

Public announcement logic Another interesting route could be to incorporate *announcements* or *actions* to the logic (see e.g. public announcement logic [67]) and ask; What happens with coalitional power when coalitions truthfully announce information?

Coalition model In the logic DCL, we keep the language $\mathcal{L_{CL}}$ and to evaluate truth of formulae, we need only coalition models, i.e. models on the form $\mathcal{N} =$

$(\mathcal{S}, E, V)$. However, when checking that E indeed is an epistemically playable effectivity function, we need a family of equivalence relations. The results in this paper would be much nicer, could we work in neighbourhood semantics without ever referring to equivalence relations. One way to look at this is to replace the quasi-crown and Q^s properties with the properties

- if $Y \in E_s(\mathcal{A})$ with $|Y| = \kappa$ is minimal under set inclusion, then there are κ many t such that $E_s = E_t$ and
- $E_s^{nc}(\emptyset) = \bigcup E_s^{nc}(\mathcal{A})$.

Then the representation theorem still holds if we can solve the problem does there exist $(\sim_i)_{i \in \mathcal{A}}$ such that E respects $(\sim_i)_{i \in \mathcal{A}}$ and that E also has the quasi-crown property for $\mathcal{A}$ over $\sim$.

Filtration It seems obvious that the logic DCL admits filtration, and thus have the strong finite model property.

Definition 7.1 (Strong Finite Model Property)**.** A logic admits the strong finite model property if for all satisfiable $\varphi \in \mathcal{L}_{\mathcal{CL}}$, there is a pointed model $(\mathcal{N}, s)$ such that $(\mathcal{N}, s) \vDash_{\mathcal{NS}} \varphi$ and the size of $\mathcal{N}$ is bounded by $f(|\varphi|)$, for a computable function f. $\dashv$

Hansen and Pauly proves in [34] that CL admits filtration and thus has the strong finite model property with the function $f(n) = 2^{2^n}$. DCL seems to have the same property. This together with a completeness proof would also give decidability, a property that is extremely preferable.

Completeness of distributed knowledge All completeness proofs in the literature of distributed knowledge uses the canonical model approach. The proofs go on to say that the canonical model is not well behaved. Next step is to unravel the model and do some tweaking (see [72] for a completely self-contained completeness proof) before trying to map the new model onto the canonical model. According to [55], the canonical model is a full communication model. This property could possibly be used to obtain a nicer and much simpler completeness proof. Additionally, it would be nice to also see a full completeness proof made directly for the neighbourhood semantics.

Bibliography

[1] T. Ågotnes and N. Alechina. Completeness of epistemic coalition logic with group knowledge. *CoRR*, abs/1107.0278, 2011.

[2] T. Ågotnes, P. Balbiani, H. van Ditmarsch, and P. Seban. Group announcement logic. *Journal of Applied Logic*, 8(1):62–81, 2010.

[3] T. Ågotnes, V. Goranko, and W. Jamroga. Alternating-time temporal logics with irrevocable strategies. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 15–24. ACM, 2007.

[4] T. Ågotnes, W. van der Hoek, M. Tennenholtz, and M. Wooldridge. Power in normative systems. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 145–152. International Foundation for Autonomous Agents and Multiagent Systems, 2009.

[5] T. Ågotnes, W. van der Hoek, and M. Wooldridge. Scientia potentia est. In *Proceedings of the Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Taipei, Taiwan, May 2011. To appear.

[6] T. Ågotnes and H. van Ditmarsch. Coalitions and announcements. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 673–680. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[7] T. Ågotnes and H. van Ditmarsch. What will they say?—public announcement games. *Synthese*, pages 1–29, 2010.

[8] T. Ågotnes and D. Walther. A logic of strategic ability under bounded memory. *Journal of Logic, Language and Information*, 18(1):55–77, 2009.

[9] R. Alur, T. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Compositionality: The Significant Difference*, pages 23–60, 1998.

[10] R. Alur, T. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM (JACM)*, 49(5):672–713, 2002.

[11] C. Areces and D. Figueira. Which semantics for neighbourhood semantics? In *Proceedings of the 21st international jont conference on Artifical intelligence*, pages 671–676. Morgan Kaufmann Publishers Inc., 2009.

[12] R. Aumann. Agreeing to disagree. *The annals of statistics*, pages 1236–1239, 1976.

[13] J. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers L. Rev.*, 19:317, 1964.

[14] N. Belnap and M. Perloff. Seeing to it that: a canonical form for agentives. *Theoria*, 54(3):175–199, 1988.

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

[16] P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic*, volume 53. Cambridge Univ Pr, 2002.

[17] J. Broersen. A complete stit logic for knowledge and action, and some of its applications. *Declarative Agent Languages and Technologies VI*, pages 47–59, 2009.

[18] J. Broersen, A. Herzig, and N. Troquard. From coalition logic to stit. *Electronic Notes in Theoretical Computer Science*, 157(4):23–35, 2006.

[19] J. Broersen, A. Herzig, and N. Troquard. A stit-extension of atl. *Logics in Artificial Intelligence*, pages 69–81, 2006.

[20] J. Broersen, A. Herzig, and N. Troquard. A normal simulation of coalition logic and an epistemic extension. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 92–101. ACM, 2007.

[21] J. Broersen, A. Herzig, and N. Troquard. What groups do, can do, and know they can do: an analysis in normal modal logics. *Journal of Applied Non-Classical Logics*, 19(4), 2009.

[22] N. Bulling and W. Jamroga. Alternating epistemic mu-calculus. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[23] A. Chagrov and M. Zakharyaschev. *Modal logic*, volume 35. Oxford University Press, USA, 1997.

[24] E. Emerson and E. Clarke. Using branching time temporal logic to synthesize synchronization skeletons. *Science of Computer programming*, 2(3):241–266, 1982.

[25] E. Emerson and J. Halpern. Decision procedures and expressiveness in the temporal logic of branching time* 1. *Journal of computer and system sciences*, 30(1):1–24, 1985.

[26] R. Fagin. *Reasoning about knowledge*. The MIT Press, 2003.

[27] R. Fagin, J. Halpern, and M. Vardi. What can machines know?: On the properties of knowledge in distributed systems. *Journal of the ACM (JACM)*, 39(2):328–376, 1992.

[28] J. Gerbrandy. Distributed knowledge. 1998.

[29] J. Gerbrandy. *Bisimulations on planet Kripke*. ILLC Dissertation Series, 1999.

[30] V. Goranko and W. Jamroga. Comparing semantics of logics for multi-agent systems. *Synthese*, 139(2):241–280, 2004.

[31] V. Goranko, W. Jamroga, and P. Turrini. Strategic games and truly playable effectivity functions. In *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, 2011.

[32] J. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial intelligence*, 54(3):319–379, 1992.

[33] J. Y. Halpern and Y. 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*, pages 480–490. Morgan Kaufmann Publishers Inc., 1985.

[34] H. H. Hansen and M. Pauly. Axiomatising nash-consistent coalition logic. In S. Flesca, S. Greco, N. Leone, and G. Ianni, editors, *JELIA*, volume 2424 of *Lecture Notes in Computer Science*, pages 394–406. Springer, 2002.

[35] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic logic*. Citeseer, 1982.

[36] A. Herzig and N. Troquard. Knowing how to play: uniform choices in logics of agency. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 209–216. ACM, 2006.

[37] J. Hintikka. *Knowledge and belief: an introduction to the logic of the two notions*. Cornell University Press, 1962.

[38] M. Huth and M. Ryan. *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press, 2006.

[39] W. Jamroga. Some remarks on alternating temporal epistemic logic. In *FAMAS'03 – Formal Approaches to Multi-Agent Systems, Proceedings*, pages 133–140, Warsaw, Poland, 2003.

[40] W. Jamroga and T. Ågotnes. Constructive knowledge: What agents can achieve under incomplete information. In *Journal of Applied Non-Classical Logics*. Citeseer, 2005.

[41] W. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63:185–219, 2004.

[42] S. Kripke. Semantical considerations on modal logic. *Acta philosophica fennica*, 16(1963):83–94, 1963.

[43] H. Kuhn. Extensive games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(10):570, 1950.

[44] E. Lemmon. Algebraic semantics for modal logics i. *The Journal of Symbolic Logic*, 31(1):46–65, 1966.

[45] E. Lemmon. Algebraic semantics for modal logics ii. *Journal of Symbolic Logic*, pages 191–218, 1966.

[46] H. Levesque and S. L. for Artificial Intelligence Research. *A logic of implicit and explicit belief*. Laboratory for Artificial Intelligence Research, Fairchild, Schlumberger, 1984.

[47] R. Montague. Universal grammar. *Theoria*, 36(3):373–398, 1970.

[48] J. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, 1950.

[49] M. Osborne and A. Rubinstein. *A course in game theory*. The MIT press, 1994.

[50] E. Pacuit. Neighborhood semantics for modal logic. *Lecture Notes, ESSLLI*, 2007.

[51] E. Parmann. Expressing properties of social choice functions using modal logic. Master's thesis, University of Bergen, Norway, November 2011. Preprint.

[52] M. Pauly. Logic for social software. *Illc dissertation series*, 10, 2001.

[53] M. Pauly. A modal logic for coalitional power in games. *Journal of logic and computation*, 12(1):149, 2002.

[54] M. Reynolds. An axiomatization of full computation tree logic. *Journal of Symbolic Logic*, pages 1011–1057, 2001.

[55] F. Roelofsen. Distributed knowledge. *Journal of Applied Non-Classical Logics*, 17(2):255–273, 2007.

[56] S. Russell, P. Norvig, J. Canny, J. Malik, and D. Edwards. *Artificial intelligence: a modern approach*, volume 74. Prentice hall Englewood Cliffs, NJ, 1995.

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

[58] D. Scott. Advice on modal logic. *Philosophical problems in Logic*, 143:74, 1970.

[59] M. Sipser. *Introduction to the Theory of Computation*. Course Technology, 1996.

[60] F. Thompson. Equivalence of games in extensive form. *Classics in game theory*, pages 36–45, 1952.

[61] J. van Benthem. Open problems in logical dynamics. *Mathematical Problems from Applied Logic I*, pages 137–192, 2006.

[62] J. van Benthem and G. Bezhanishvili. Modal logics of space. *Handbook of spatial logics*, pages 217–298, 2007.

[63] W. van der Hoek, B. van Linder, and J. Meyer. Group knowledge is not always distributed (neither is it always implicit). *Mathematical social sciences*, 38(2):215–240, 1999.

[64] W. van der Hoek, R. Verbrugge, et al. Epistemic logic: A survey. *Game theory and applications*, 8:53–94, 2002.

[65] W. van der Hoek and M. 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. ACM, 2002.

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

[67] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic epistemic logic*, volume 337. Springer Verlag, 2007.

[68] S. van Otterloo and G. Jonker. On Epistemic Temporal Strategic Logic. *Electronic Notes in Theoretical Computer Science*, XX:35–45, 2004. Proceedings of LCMAS'04.

[69] S. van Otterloo and G. Jonker. On epistemic temporal strategic logic. *Electronic Notes in Theoretical Computer Science*, 126:77–92, 2005.

[70] P. Vanderschraaf and G. Sillari. Common knowledge. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Stanford Encyclopedia of Philosophy, spring 2009 edition, 2009.

[71] J. von Neumann, O. Morgenstern, A. Rubinstein, and H. Kuhn. *Theory of games and economic behavior*. Princeton Univ Pr, 2007.

[72] Y. Wáng and T. Ågotnes. Public announcement logic with distributed knowledge. *Logic, Rationality, and Interaction*, pages 328–341, 2011.