Strategic Reasoning in Interdependence:
Logical and Game-theoretical Investigations

SIKS Dissertation Series No. 2011-34

The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Printed by Gildeprint Drukkerijen B.V., Enschede

Strategic Reasoning in Interdependence:
Logical and Game-theoretical Investigations

Strategische Redenering in Afhankelijkheid:
Logische en Speltheoretische Onderzoeken
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties in het openbaar
te verdedigen op maandag 17 oktober 2011 des middags te 12.45 uur

door

Paolo Turrini

geboren op 03 februari 1980 te Cagliari, Italië

Promotoren:	Prof. dr. J.-J. Ch. Meyer
	Prof. dr. A. Visser
Co-promotoren:	Dr. ir. J.M. Broersen
	Dr. R. Mastop

A Davide

"O frati," dissi "che per cento milia
perigli siete giunti a l'occidente
a questa tanto picciola vigilia

d'i nostri sensi ch'è del rimanente,
non vogliate negar l'esperïenza
di retro al sol, del mondo sanza gente.

Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza".

Dante Alighieri, *La Divina Commedia*, *Inferno*, XXVI 112-120.

Contents

Foreword

After graduating it took me a couple of years to understand that to model social interaction one needs some serious maths. Before that moment, coinciding with the start of my PhD studies in Cognitive Artificial Intelligence in Utrecht, I had been wandering around the corridors of the Institute of Cognitive Science and Technology in Rome, embarked upon a quest for an all-encompassing model of human behaviour that could be described in everyday language.

Two fellows must be held responsible for making that quest fail: John-Jules Meyer and, believe it or not, Cristiano Castelfranchi. The first made me love the depth of formal languages, their accuracy and flexibility in capturing intuitive notions; the latter convinced me that a proper science of social concepts, even though starting from the powerful intuitions coming from our experience, should be first of all rigorous and formal or, as he and Leibniz say, *more geometrico demonstrata*.

I took up the new challenge under the guidance of two inspiring and dedicated supervisors: Jan Broersen and, once again, John-Jules Meyer. Not only have they granted me the possibility of developing my PhD research but they have also been assisting me during its many turning points. In particular I thank John-Jules for instilling in me his optimism, his passion and his respect for science, and I thank Jan for his constant support, both as a great scientific advisor and as a great friend. Their huge work has been supported by Rosja Mastop and Albert Visser, who I also thank for joining my supervision.

To Utrecht I carried the cognitive scientist's burden, for which I owe much to the researchers at the Institute for Cognitive Science and Technology, starting from its director Cristiano Castelfranchi, a volcano of inspirations and challenges, and continuing with Giulia Andrighetto, Amedeo Cesta, Rosaria Conte, Rino Falcone, Francesca Giardini, Emiliano Lorini, Maria Miceli, Fabio Paglieri, Mario Paolucci, Gennaro di Tosto and Luca Tummolini.

In Utrecht - by the way my arrival would not even be conceivable without Frank Dignum and Rosaria Conte - I soon realized that my research enterprise was not to be conducted in loneliness. One person above all shared with me through endless discussions his idealistic and provocative standpoints, and finally joined my efforts in organizing a plot against classical game theory: Davide Grossi. Our strong bond has shaped my way of understanding science. To him, his enterprising and committed vision, I dedicate this thesis.

My colleagues have deilighted me with a cozy and stimulating environment. Thanks to Huib Aldewereld, Liz Black, Susan van den Braak, Mehdi Dastani, Gennaro di Tosto, Maaike Harbers, Max Knobbout, Eric Kok, Joost van Oijen, Marieke Peeters, Loris Penserini, Henry Prakken, Michal Sindlar, Bas Steunebrink, Nick Tinnemeier, Nieske Vergunst, Gerard Vreeswijk, Tom van der Weide and Joost Westra. Special mention to Henry Prakken, who besides being a brilliant researcher is also a brilliant chess player. It is thanks to him that I rediscovered my old passion.

My thesis well reflects my research work during my PhD. Many thanks to the members of the reading committee Marek Sergot, John Horty, Johan van Benthem, Eric Pacuit and Cristiano Castelfranchi, who have contributed to improving its quality and who have always been a source of inspiration for my research in the past years. Also thanks to the many colleagues who, although being neither coauthors nor supervisors nor reading committee members, have been actively commenting upon my research work: Rosaria Conte, Umberto Grandi, Maria Miceli, R. Ramanujam, Jaime Sichman, Leon van der Torre.

The university was not the only place in Utrecht that I have been visiting during the past few years. Café Bëlgie tops the list, not to mention the Sportcentrum Olympos. In this beautiful city I had the luck to come across marvellous friends. I thank Aline Duine, Henry Ervasti, Pierre-Olivier Guerra, Kiril Hristov, Eimear Murphy, Aline Pieterse and Nikolas Vaporis for sharing with me so many nice moments.

Italians abroad are an institution. I thank Umberto Grandi, Davide Grossi, Daniele Porello and Pietro Galliani for keeping my love for Italy alive.

Thanks also to my chess team mates who have, possibly unconsciously, thaught me so much about strategic reasoning and, certainly consciously, a great deal of chess and friendship. I am sorry I am not (yet) as good in chess as they are. Thanks to Judith van Amerongen, Ed van Eden, Olivier Huizer, Michel Kerkhof, Marijke Kok and Marcel van Os.

I am proud to have such a supportive family at my side. Huge thanks to my parents Luigi and Maria Assunta and to my brother Antonello. Besides, as an added member, thanks to Yvonne, especially for the chats with beer and nuts.

Finally, thanks to Anouk, for her love, patience and support during all these years.

Paolo Turrini

Luxembourg Ville, August 30th 2011

Chapter 1

Introduction

> *The language of game theory — coalitions, payoffs, markets, votes — suggests that it is not a branch of abstract mathematics; that is motivated by and related to the world around us; and that it should be able to tell us something about that world.*
>
> Robert J. Aumann, *What is game theory trying to accomplish?* [6]

1.1 Motivation

The present work analyzes various aspects of coalitional rationality in strategic interaction, i.e. how a group of rational individuals, each endowed with their own preferences and strategies, takes a joint decision.

Nowadays game theory, the discipline studying interactive decision making [51], is subdivided into two branches: *cooperative* game theory, that studies how abstract groups of individuals — called coalitions — act together, and *non-cooperative* game theory, that studies instead how rational individuals strive to realize their goals.

However a variety of relations can be drawn between the two faces of the game-theoretical coin. The link between the behaviour of rational individuals and that of coalitions is already emphasized in von Neumann and Morgenstern's seminal account of games [70, p.221]:

> As soon as there is a possibility of choosing with whom to establish parallel interests, this becomes a case of choosing an ally. When alliances are formed, it is to be expected that some kind of mutual understanding between the two players involved will be necessary. One can also state it this way: A parallelism of interests makes a cooperation desirable, and therefore will probably lead to an agreement between the players involved.

In the spirit of [70] we will study the rationality of coalitions that arise from an underlying *parallelism of interests* among their members focusing on *how individual decisions are reflected in the decisions of the coalitions in which they take part*. In this

purpose, the notion of rationality available in game theory for individual players will be lifted to coalitions and studied in a cooperative twist. We will also analyze the reasons for rational individuals to work together, proposing a theory of coalitions that takes these reasons into account. Their formation, stability and disruption will ultimately be judged against the preferences and the strategic possibilities of the individuals composing them.

1.2 Research Questions

1.2.1 Coalitions and rationality

In non-cooperative interaction self-interested individuals strive to achieve their own goals. The point of view taken in this work is to view how the competitive perspective associated to such situations can be turned into a cooperative one, by modelling how individuals are to act should they decide to join their forces.

To illustrate the research questions cropping up from our standpoint let us reconsider a classic of game theory, the prisoner's dilemma. Its story runs as follows [45]:

> Tanya and Cinque have been arrested for robbing the Hibernia Savings Bank and placed in separate isolation cells. Both care much more about their personal freedom than about the welfare of their accomplice. A clever prosecutor makes the following offer to each. - You may choose to confess or remain silent. If you confess and your accomplice remains silent I will drop all charges against you and use your testimony to ensure that your accomplice does serious time. Likewise, if your accomplice confesses while you remain silent, he will go free while you do the time. If you both confess I get two convictions, but I'll see to it that you both get early parole. If you both remain silent, I'll have to settle for token sentences on firearms possession charges. If you wish to confess, you must leave a note with the jailer before my return tomorrow morning.

The dilemma lies in the fact that both prisoners would profit from remaining silent. Nevertheless, considering what the other does, each prisoner is better off confessing than remaining silent. Being incapable of coordinating, each prisoner is faced with an individual choice that needs to be rational whatever the accomplice is going to choose. The scene is set in such a way that each prisoner sees his accomplice as his opponent, and reasons accordingly.

However, suppose that during the night Tanya manages to escape from her isolation cell, reach Cinque, and return to her cell before the prosecutor is back. Completely different possibilities are now available. During their secret meeting, Tanya and Cinque might have been able to reach a binding agreement and might now be able to reason as a coalition, taking their decisions as part of a larger entity.

The newly formed coalition of Tanya and Cinque has a variety of choices at its disposal, combining the choices of each individual. But how to order them? And how are the individual preferences to be reflected in the coalitional decisions? Applying the analysis of rational decision making to coalitions poses a number of interesting challenges.

Research Question 1. In strategic interaction, coalitional choices result from the choices of the individuals composing them. But how can individuals endowed with their own preferences merge in a larger coalition and decide there what choice to take?

Chapter 3 proposes a model of coalitional rationality in strategic interaction, where a group of individuals, each holding preferences and strategic possibilities, chooses among possible alternatives. Issues are addressed as how to aggregate individual preferences and how to order the choices at the disposal of a coalition.

1.2.2 Cooperation and competition

Game theory textbooks draw a clear distinction between models of cooperative behaviour, called *cooperative* or *coalitional* games, and models of non-cooperative behaviour, called *non-cooperative* or *strategic* games. [1]

> We may distinguish between two types of models: those in which the sets of possible actions of *individual* players are primitive and those in which the sets of possible joint actions of *groups* of players are primitive.

[51, p.2]

Much effort has been devoted to the understanding of their relation. A variety of formal results [49, 56] have established that a certain class of cooperative games fully describe strategic games, in the sense that for each strategic game the cooperative possibilities of individuals involved can be described by an element of that class, and that for each element of that class there is a strategic game which has an equivalent description of its cooperative possibilities.

These results make use of similar formal apparatus, representing coalitional power via the so-called effectivity functions, that associate to each group of individuals the properties of an interaction that they can achieve together. For instance, in the prisoner's dilemma Tanya and Cinque can achieve together that they both remain silent or they can achieve together that Tanya remains silent while Cinque doesn't.

> To describe the power distribution over a set of outcomes A between coalitions we consider a very simple object which assigns to every coalition its effectiveness power, intuitively understood as follows: if a coalition is effective for some feasible set, then the coalition has the power to

[1] The term *strategic game* is often used to describe interactions among rational individuals that do not consider the sequential structure of the decisions, as opposed to *extensive games*, that instead do [51]. The present work is not concerned with extensive games and the term *strategic game* is used as a synonym of non-cooperative game.

> obtain some alternative in this set; it might not have the power to choose this alternative from the set, but at least it can guarantee that something from the set will be chosen.

[1, p.30]

Generally speaking, each strategic game consists of a description of what outcomes players can achieve by executing their strategies. Effectivity functions describe how groups of these players can coordinate their strategies to achieve joint goals. The interplay between effectivity functions and strategic games shows that the models of strategic and of cooperative behaviour are two faces of the same coin, confirming the standpoint taken in [51, p.3]:

> In particular we do not share the view of some authors that noncooperative models are more 'basic' than cooperative ones; in our opinion, neither group of models is more 'basic' than the other.

The relation between the two modes of interaction can be investigated further, reasoning on the assumptions behind the results in the literature and inserting it into a larger theory of coalitional rationality.

Research Question 2. How can a model of non-cooperative interaction be described with a model of cooperative interaction? And how can a model of non-cooperative interaction be retrieved from a model of cooperative interaction?

Chapter 3 answers the question for a more general case than the one treated in [49, 56] and moreover corrects a believed correspondence given in [54].

1.2.3 Coalitions and interdependence

The standard approach to describing the cooperative possibilities of individuals in strategic interaction consists of attributing to a coalition the capacity of fully coordinating its members. In other words, in the classical account exemplified by the effectivity functions, a coalition disposes of each combination of actions that can be performed by its members. In a cooperative version of the prisoner's dilemma for instance Tanya and Cinque can achieve together that Tanya remains silent while Cinque doesn't. It is though difficult to believe that Tanya will be willing to accept such an agreement, as (i) the consequence of it would be Cinque to be released and Tanya to do serious time in prison; (ii) there is a better alternative that she can reach on her own, namely cooperate with the prosecutor; (iii) there is a different strategy that Tanya would like Cinque to play, namely remaining silent. This observation sheds light on a more general fact: coalitions do not form unless there is a reason for it, and the reason lies in the possibility for players to take advantage of each other.

In strategic interaction — but a similar point can be made for social interaction in general— what each individual does affects the other individuals involved. As put in [20, p. 161–162],

> Sociality obviously presupposes two or more players in a common, shared world. A "Common World" implies that there is *interference* among the actions and goals of the players: the effects of the action of one player are relevant for the goals of another: i.e., they either favour the achievement or maintenance of some goals of the other's (*positive interference*), or threat some of them (*negative interference*).

The classical account of coalition formation in games only focuses on cooperative possibilities and does not take players' interdependence into account. Tanya and Cinque depend on each other for not staying long in prison and in this sense the agreement that Tanya remains silent and Cinque talks is patently unfair. Generally speaking, the process of coalition formation seems to require a certain grade of reciprocity among the individual involved and to result in a mutual resolution of what in [20, p. 161–162] are called *positive interferences*. It is then desirable to weaken the classical theory of coalition considering more structured versions of cooperative interaction that incorporate notions such as interdependence and reciprocity.

Research Question 3. Can coalitions be seen as resulting from an exchange of favours between interdependent individuals?

This question is answered in Chapter 5, that extends the classical theory of coalitions in strategic interaction with the concept of interdependence, giving rise to a specific class of cooperative games, i.e. the dependence games.

1.2.4 Rationality and logic

In the opening of his book *Logic and Structure* Dirk van Dalen writes [67, p.5]:

> Traditionally, logic is said to be the art (or study) of reasoning; so in order to describe logic in this tradition, we have to know what 'reasoning' is. According to some traditional view reasoning consists of the building of chains of linguistic entities by means of a certain relation '. . . follows from . . . ', a view which is good enough for our present purposes.

When studying the choices of individuals in strategic interaction we are undoubtedly dealing with a form of reasoning. In the prisoner's dilemma the choice of each individual not to remain silent *follows from* the judgment of the situations resulting from the possible choices of the other prisoner: if Cinque talks, I had better talk; if Cinque does not talk, I had better talk; In conclusion, I had better talk.

If logic is the study of reasoning then a specific class of logical languages can certainly be developed to study strategic reasoning, the type of reasoning associated with strategic interaction.

Traditionally the class of logical languages that have been devoted to describe strategic reasoning are *modal* logics, mathematical languages that talk about graph-like structures.

> Over the years modal logic has been applied in many different ways. It has been used as a tool for reasoning about time, beliefs, computational systems, necessity and possibility, and much else besides. These applications, though diverse, have something important in common: the key ideas they employ (flows of time, relations between epistemic states, transitions between computational states, networks of possible worlds) can all be represented as simple graph-like structures. And as we shall see, modal logic is an interesting tool for talking about such structures: it provides an internal perspective on the information they contain.

[11, p.2]

In models of strategic interaction, graph-like structures are ubiquitous. Think for instance of how in our example Tanya orders the possibilities resulting from the interaction with Cinque. She prefers a situation in which they both remain silent to a situation where Cinque talks and she remains silent, and she prefers even more a situation where she talks and Cinque remains silent, and so on. But outcomes can not only be preferred, but also achieved. Tanya can for instance achieve that she remains silent, but she cannot achieve that Cinque talks. Tanya and Cinque can instead achieve this. Each individual, and arguably each coalition, can order the outcomes according to his preferences or according to his capacities, giving rise to mathematical structures that are well-suited for a modal account.

Hereby, both in the case of classical coalitional games and in the case of dependence games a logical analysis of coalitional reasoning can be systematically carried out by means of modal languages.

Research Question 4. What is the logical structure of strategic reasoning in coalitional games and dependence games? How can interesting properties of coalitional rationality, such as presence of reciprocity among the individuals, be characterized by means of a logical language?

Chapter 4 and Chapter 6 develop modal languages to describe the fundamental properties of coalitional rationality studied in Chapter 3 for standard coalitional games and Chapter 5 for dependence games.

1.2.5 Rationality and norms

When a group of individuals is confronted with a number of possible choices, often the question arises of what the individuals *should* do. Traditionally, the formal study of terms as should, must, ought to, may etc. has been dealt with by deontic logic, a branch of modal logic that analyzes the structure of normative concepts. Recently John Horty's seminal contribution [41] has brought deontic logic into the realm of strategic interaction, establishing a parallel between coalitional rationality and normative concepts.

> In the past, the task of mapping the relations between deontic logic and act utilitarianism has resulted in surprising difficulties, leading some

> writers to suggest the possibility of a conflict in the fundamental principles underlying the two theories. One source of these difficulties, I believe, is the gap between the subjects of normative evaluation involved in the two areas: while deontic logic has been most successfully developed as a theory of what ought or ought not to be, utilitarianism is concerned with classifying actions, rather than states of affairs, as right or wrong. The present account closes this gap, developing a deontic logic designed to represent what agents ought to do within a framework that allows, also, for the formulation of a particular variant of act utilitarianism, the dominance theory.

[41, p.70]

In a nutshell, Horty's proposal consists of viewing choices that should be performed as carrying a meaning in terms of rationality, i.e. they are the rational choices at a coalition's disposal.

A different, arguably more abstract, view has been proposed within computer science, where John-Jules Meyer's account [48] relates deontic notions to dynamic logic, a mathematical language to reason about computer programs. Building upon previous philosophical work, Meyer studies a labelling of outcomes as violations against which individual actions get evaluated in a deontic sense. The two approaches look at norms from complementary perspectives: while Horty emphasizes the *internal* side of norms (norms are related to choices and preferences of the individuals involved in the interaction), Meyer emphasizes the *external* side (obligations are not related to choices and preferences but depend on a preestablished label on the states). The two views on norms, apparently at odds with each other, offer a comprehensive perspective on the analysis of rationality and can as such be studied together.

Research Question 5. What are the obligations, permissions and prohibitions to be applied to a coalition in order for it to act rationally? What are instead its obligations in order to fulfill external requirements? How are these two views related to each other?

Chapter 4 and Chapter 6 studies the two perspectives in coalitional and dependence games. In line with the internal view analyzed in [41], we study norms as linked with rational action; and in line with the external view analyzed in [48], we also study them as originating from *a priori* established violations.

1.3 Thesis Outline

This thesis analyzes the properties of rationality in strategic interaction combining several perspectives, which can be articulated into two main branches:

- Coalitional games and dependence games analysis;
- Structural and logical analysis.

	Coalitional Games	Dependence Games
Structural analysis	Chapter 3	Chapter 5
Logical analysis	Chapter 4	Chapter 6

Table 1.1: Core chapters outline. The analysis of coalitional games and dependence games is carried out at the structural level, analyzing the properties of the models, and at the logical level, analyzing the properties of the logical languages interpreted on the models.

More specifically, the thesis presents an analysis of both coalitional games, i.e. models of group rational behaviour under the assumptions of perfect coordination, and dependence games, i.e. models of group rational behaviour under the assumption of reciprocity. The analysis is carried out at the structural level, i.e. analyzing the properties of the models, and at the logical level, i.e. studying the properties of a logical language interpreted on the models.

The two branches are reflected on the thesis structure, sketched in Table 1.1, which consists of two parts and each part of two chapters:

- Part I looks at the structure (Chapter 3) and the logic (Chapter 4) of coalitional games;
- Part II looks at the structure (Chapter 5) and the logic (Chapter 6) of dependence games.

Part I and Part II are preceded by Chapter 2 that introduces the mathematical preliminaries used throughout the thesis. The chapter mainly recalls well-known concepts in set theory, logic and game theory, providing some basic results.

1.3.1 Part I: Strategic reasoning and coalitional games

The first part consists of two chapters analyzing coalitional games from a structural and logical perspective.

Chapter 3 studies a notion of coalitional rationality analogue to that used in non-cooperative games for individual players and obtained by introducing orders on effectivity functions that take into account preference relations and opponents' possibilities. It moreover establishes a correspondence between a class of effectivity functions and strategic games.

Chapter 4 introduces a logic language to reason on coalitional rationality and able to characterize the structural notions studied in Chapter 3. Part of the chapter is devoted to studying the regulation of coalitional rationality, inspired by classical work on deontic logic.

1.3.2 Part II: Strategic reasoning and dependence games

The second part consists of two chapters analyzing dependence games from a structural and logical perspective.

Chapter 5 provides a formal definition of dependence relation between players in a strategic games, obtained by generalizing the classical ones of best response and dominant strategy. The notion of dependence relation allows to study the reciprocity cycles arising among the players, i.e. what they can do for each other, and allows for the formulation of the notion of agreement. Agreements allow in turn to view strategic games as dependence games, the class of cooperative games where coalitional choices are determined by agreements.

Chapter 6 provides a logical analysis of dependence games. The theory of dependence elaborated in Chapter 5 is equipped with the tools defined in Chapter 3 to build up a semantics of coalitional rationality in undertaking agreements. A part of the chapter revisits the classical deontic operators and redefines them in terms of agreements, proposing a deontic logic for coalition formation.

The concluding Chapter 7 wraps up the work and briefly summarizes how the research questions have been answered.

Chapter 2

Preliminaries

> *We strive to make statements that, while perhaps not falsifiable, do have some universality, do express some insight of a general nature; we discipline our minds through the medium of mathematical model; and at their best, our disciplines do have beauty, simplicity, force and relevance.*
>
> Robert J. Aumann, *What is game theory trying to accomplish?* [6]

This chapter introduces the notational conventions and the basic facts on the mathematical notions to be used henceforth. Even though most of them are well-established in the literature, it is convenient, to avoid ambiguities, to list them once more.

Set theoretical notions will be introduced in Section 2.1, game-theoretical notions in Section 2.2 and logical notions in Section 2.3. Main references for Section 2.1 are [37, 3, 28], for Section 2.2 [70, 51, 50, 1, 9, 30] and for Section 2.3 [23, 10, 46, 54].

2.1 Sets and Relations

Sets are denoted $X, Y, Z, \ldots$, possibly with subscripts and superscripts. The fact that the object x is an element of the set X is denoted $x \in X$, while the fact that it is not an element of the set X is denoted $x \notin X$. The fact that a set X is included in a set Y is denoted $X \subseteq Y$, while the fact that it is not included in Y is denoted $X \nsubseteq Y$. Strict inclusion is denoted $\subset$, i.e. $X \subset Y$ means that $X \subseteq Y$ and $Y \nsubseteq X$. The powerset of a set X, i.e. the set of all its subsets, is denoted 2^X. The empty set is denoted $\emptyset$. Union and intersection are denoted $X \cup Y$ and $X \cap Y$ respectively. If $\mathcal{X}$ is a set of sets, then $\bigcup \mathcal{X}$ and $\bigcap \mathcal{X}$ denote respectively the union and the intersection of all sets in $\mathcal{X}$, i.e. $\bigcup \mathcal{X} = \{x \mid x \in X \text{ for some } X \in \mathcal{X}\}$ and $\bigcap \mathcal{X} = \{x \mid x \in X \text{ for all } X \in \mathcal{X}\}$. The complement of X in Y is the set of all elements of Y that are not in X and is denoted $Y \setminus X$. When Y is understood $\overline{X}$ is used. The cartesian product of sets X and Y is the set of ordered pairs $X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}$. More generally, if X_α is an indexed family of sets, for $\alpha \in I$, the cartesian product of the sets X_α is the set $\bigtimes_{\alpha \in I} X_\alpha$ consisting of all I-tuples whose α^{th} component is in X_α for all $\alpha \in I$. The set of natural numbers will be denoted $\mathbb{N}$, that of reals $\mathbb{R}$.

A binary relation R defined on a set X is a subset of $X \times X$. R is called:

- *reflexive* if $(x, x) \in R$ for all $x \in X$;
- *irreflexive* if $(x, x) \notin R$ for all $x \in X$;
- *symmetric* if $(x, y) \in R$ whenever $(y, x) \in R$;
- *transitive* if $(x, z) \in R$ whenever both $(x, y) \in R$ and $(y, z) \in R$;
- *antisymmetric* if $x = y$ whenever $(x, y) \in R$ and $(y, x) \in R$;
- *complete* if for all $x, y \in X$ either $(x, y) \in R$ or $(y, x) \in R$;
- *trichotomous* if for all $x, y \in X$ either $(x, y) \in R$ or $(y, x) \in R$ or $x = y$.

Notice that a relation is complete if and only if it is reflexive and trichotomous. A binary relation is moreover called a:

- *preorder* if it is reflexive and transitive;
- *partial order* if it is an antisymmetric preorder;
- *strict partial order* if it is irreflexive and transitive;
- *total order* if it is a trichotomous strict partial order;
- *total preorder* if it is complete preorder;
- *equivalence relation* if it is reflexive, transitive and symmetric.

When no confusion can arise the fact that $(x, y) \in R$ can be denoted in infix notation, i.e. xRy. The set on which a relation is defined is referred to as *domain* and omitted when clear.

Given a relation R defined on a domain W, a *cycle* $\mathfrak{c}$ of length $k - 1$ in R is a tuple $(x_1, \ldots, x_k)$ such that:

1. $x_1, \ldots, x_k \in W$;
2. $x_1 = x_k$;
3. for all x_i, x_j with $1 \leq i \neq j < k$, $x_i \neq x_j$;
4. $x_1 R x_2 R \ldots R x_{k-1} R x_k$.

Given a cycle $\mathfrak{c} = (x_1, \ldots, x_k)$, its orbit $O(\mathfrak{c}) = \{x_1, \ldots, x_{k-1}\}$ denotes the set of its elements.

Consider a set $X \subseteq W$ and a set of sets $\mathcal{X}$ such that each $Y \in \mathcal{X}$ is such that $Y \subseteq W$. The operation of superset closure $(\mathcal{X})^{\sup_W}$ and that of set restriction $\mathcal{X} \sqcap X$ are defined as follows:

$$(\mathcal{X})^{\sup_W} = \{Z \subseteq W \mid \text{ there is } Y \in \mathcal{X} \text{ such that } Y \subseteq Z \subseteq W\}$$

$$\mathcal{X} \sqcap X = \{Y \cap X \mid Y \in \mathcal{X}\}$$

$\mathcal{X}$ is called:

- *closed under finite unions*, if $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ implies $X \cup Y \in \mathcal{X}$;
- *closed under supersets*, if $X \in \mathcal{X}$ and $X \subseteq Y$ implies $Y \in \mathcal{X}$; or alternatively if $\mathcal{X} = (\mathcal{X})^{\sup}$;
- *closed under subsets*, if $Y \in \mathcal{X}$ and $X \subseteq Y$ implies $X \in \mathcal{X}$;
- *closed under finite intersections*, if $X \in \mathcal{X}$ and $Y \in \mathcal{X}$ implies $X \cap Y \in \mathcal{X}$;
- *containing the unit*, if $W \in \mathcal{X}$.
- *filter*, if it is closed under finite intersections, contains the unit and it is closed under supersets;
  - *principal filter*, if it is a filter and $\bigcap \mathcal{X} \in \mathcal{X}$;
  - *nonprincipal filter*, if it is a filter but not a principal filter;
  - *proper filter*, if it is a filter different from 2^W;
- *partition*, if $\bigcup \mathcal{X} = W$ and $X, Y \in \mathcal{X}$ implies that $X \cap Y = \emptyset$. The partition of a set X is abbreviated $P(X)$.

A principal filter $\mathcal{F}$ is also said to be *generated* by the intersection of all its members, formally $X \in \mathcal{X}$ if and only if $\bigcap \mathcal{X} \subseteq X$.

Functions, denoted $f, g, h, \ldots$, are binary relations such that no two distinct elements have the same first component, i.e. $(x, y) \in f$ and $(x, y') \in f$ implies that $y = y'$, which is also denoted $f(x) = y$. $f : Y \to Z$ indicates that function f has domain Y and range Z. If $X \subseteq Y$ and $f : Y \to Z$ then $f(X) = \{f(x) | x \in X\} \subseteq Z$ is the image of set X under function f. If $X \subseteq Y$ and $f : Z \to Y$ then $f^{-1}(X) = \{x | f(x) \in X\} \subseteq Z$ is the preimage of set X under function f. A function $f : A \to B$ is *injective* if, for $x, y \in X$, $x \neq y$ implies that $f(x) \neq f(y)$, *surjective* if for all $y \in Y$ there exists $x \in X$ such that $f(x) = y$; *bijective* if both surjective and injective. If $f : X \to Y$ is a bijective function then its *inverse function* f^{-1} associates to each element y of Y a unique x such that $f(x) = y$. Given two functions $f : X \to Y$ and $g : Y \to Z$ their *composition* $f \circ g$ is defined as follows: $(f \circ g)(x) = y$ if and only if $g(f(x)) = y$.

Given a set X, a *permutation* on X is a bijection $\mu : X \to X$. The set $PERM_X$ of all permutations on a set X consists of $|X|!$ different elements. When the set X is finite, each of these permutations $\mu : X \to X$ induces a set of cycles on X, that naturally partition X [3].

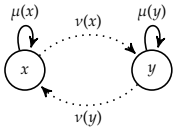

Figure 2.1: Permutations, cycles and partitions on a finite set X. Permutations μ and ν induce a set of cycles on X, whose orbits form a partition.

Proposition 1 (Permutations and cycles) *Let μ be a permutation on a finite set X and consider a relation $R \subseteq X \times X$ such that xRy if and only if $y = \mu(x)$. Then R consists of a set of cycles $\mathfrak{C}$ such that $\{O(\mathfrak{c}) \mid \mathfrak{c} \in \mathfrak{C}\}$ is a partition of X.*

Proof *As the set X is finite we can enumerate all its elements with the first $|X|$ natural numbers. Now we can construct the cycles as follows: open a pair of brackets, write down the element associated to 1 followed by its R-successor and so on, closing the brackets when the first element is repeated. Open a new pair of brackets, list element associated to the smallest number which has so far not been mentioned and repeat the procedure until all the elements of X are mentioned. Notice that being R corresponding to a permutation, for each element of X there is a unique R-successor and a unique R-predecessor in X. Suppose that the procedure outputs a tuple $(x_1, \ldots, x_k)$ that is not a cycle. Then we have that for some x_i, x_j with $1 \leq i \neq j < k$, $x_i = x_j$. Suppose without of loss of generality that $i < j$. But then for each $m \in \mathbb{N}$ and for $x_i R x_{i+1} \ldots R x_{i+m}$ and $x_j R x_{j+1} \ldots R x_{j+m}$ we have that $x_{j+m} = x_{i+m}$. Pick now an $n \in \mathbb{N}$ such that $x_{j+n} = x_k$. We must have that $x_{i+n} = x_k$ and $1 < i + n < j + n$. Contradiction. As X is finite the procedure guarantees that all $x \in X$ are member of some cycle. In conclusion we obtain a set of cycles $\mathfrak{C}$ such that $\{O(\mathfrak{c}) \mid \mathfrak{c} \in \mathfrak{C}\}$ is a partition of X.*

Example 1 (Permutations and cycles) *Consider the set $X = \{x, y\}$ and the permutations μ, ν such that $\mu(x) = x, \mu(y) = y$ and $\nu(x) = y, \nu(y) = x$, all possible different permutations on X as, notice, $|X|! = 2$. μ induces the cycles $(x, x)(y, y)$, connecting x and y with themselves, while ν induces the cycle (x, y, x), connecting x with y and y with x. In both cases the set of orbits of the cycles partition the set X. Figure 2.1 illustrates this.*

Let $P(\mu)$ be the partition induced by μ on X, and $\mathcal{P}(\mu)$ the nonempty powerset of this partition. The fact that a set $Y \subseteq X$ is the union of some members of the partition induced by μ will be denoted with $Y \in \mathcal{P}_X(\mu)$. Whenever X is understood the notation $\mathcal{P}(\mu)$ will be adopted. Permutations form a group under the operation of function composition [3] and are therefore closed under composition and inverse, i.e. for $\mu', \mu'' \in PERM_X$, we have that $\mu' \circ \mu'' \in PERM_X$ and that $\mu'^{-1} \in PERM_X$.

2.2 Games

The present work deals with strategic interaction. Therefore the basic ingredients we will be working with are a finite set N, to be understood as a set of *players*, and a

	L	R
U	2, 2	0, 3
D	3, 0	1, 1

Prisoner's dilemma

	L	R
U	3, 3	2, 2
D	2, 2	1, 1

Full convergence

	L	R
U	1, 1	0, 0
D	0, 0	1, 1

Coordination

	L	R
U	3, 3	2, 2
D	2, 5	1, 1

Partial convergence

Figure 2.2: Examples of two players strategic games. The strategies are labelled according to their position in the matrix: L stands for left, U for up and so on. Players, that henceforth will get the anonymous names of *Row* and *Column*, have a preference order over the outcomes as described by the matrix entries, with *Row* being associated to the first component and *Column* to the second one.

set W to be understood as a set of *alternatives*. Players are denoted $i, j, k, \ldots$ while sets of players, i.e. elements of 2^N, are denoted $C, C', C'', \ldots$ and are henceforth called *coalitions*. The coalition made by all players, i.e. the set N, will be referred to as the *grand coalition*. Alternatives are denoted $u, v, w, \ldots$ and are also called *outcomes*, *states* or *worlds*. Players are assumed to have preferences over the alternatives. Therefore, each player i is endowed with a preference order $(\succeq_i)_{i \in N}$, a total preorder on the set of alternatives, where $v \succeq_i w$ has the intuitive reading that outcome v is *at least as nice* as outcome w for player i. The corresponding strict partial order is defined as expected: $v \succ_i w$ if, and only if, $v \succeq_i w$ and not $w \succeq_i v$, to mean that for player i outcome w is *strictly better* than outcome v. The notation $\prec_i, \preceq_i$ for the reverse relations will be used as well when no confusion can arise.

2.2.1 Strategic games

As informally illustrated in Chapter 1 strategic games are models of interactive decision making that relate players' preferences to their strategic possibilities. A *strategy* σ_i for a player i is a specification of i's moves at each of i's decision points, i.e. the move he makes at each turn of his. The intuition behind strategic games, first introduced by von Neumann and Morgenstern, is that "players begin a game by making a firm preplay commitment to a particular strategy" [9]. In other words, strategic games abstract away from the sequential structure of the decision problems. Their formal definition goes as follows.

Definition 1 (Strategic game) *A strategic game is a tuple* $\mathbb{G} = (N, W, \Sigma_i, \succeq_i, o)$ *where:*

- N *is a set of players;*
- W *is a set of outcomes;*
- Σ_i *is a set of strategies for player* $i \in N$*;*

- $\succeq_i$ *is a total preorder on* W *for player* $i \in N$*;*
- $o : \bigtimes_{i \in N} \Sigma_i \to W$ *is the outcome function, relating tuples of individual strategies, also called* strategy profiles*, to elements of* W*.*

Games will be denoted $\mathbb{G}, \mathbb{G}', \mathbb{G}'', \ldots$. $\sigma_i \in \Sigma_i$ will denote an individual strategy for player i in his set of strategies Σ_i, while σ_C will denote an element of $\bigtimes_{j \in C} \Sigma_j$, a tuple of individual strategies for each member of the coalition C; as a convention $\sigma_{\overline{\{i\}}}$ will be denoted σ_{-i}.

Examples of strategic games are given in Figure 2.2. Following the usual convention, the row player and the column player obtain a payoff as established in the entries of the matrix corresponding to the intersection of the respective choices. In other words, vectors representing payoffs in a bimatrix are of the form (payoff(row),payoff(column)). Numerical entries carry a preference order, and in this sense the games given in Figure 2.2 fall under the definition of strategic game given in Definition 1. Not all preference relations though can be univocally associated to a numerical entry [1], so not all strategic games can be represented in a matrix form. However for illustrative purposes matrices will often be used to represent the general case.

The outcome function, that associates strategy profiles to outcomes, is usually required to be a bijection (see for instance [51]) and can be consequently dispensed with. Here the outcome function will not be assigned any particular property, unless otherwise specified.

At times it is convenient to talk about games without mentioning preference relations. These structures are called *strategic game forms*, and they are denoted $\mathbb{F}_s, \mathbb{F}'_s, \mathbb{F}''_s, \ldots$. The game form $\mathbb{F}_s$ can be associated to the preference relation $\succeq_i$ and the resulting game is denoted $(\mathbb{F}_s, \succeq_i)$. At other times it is convenient to evaluate outcomes without making reference to players' strategic possibilities. We can do this by making use of the classical notion of Pareto optimality [51], that only considers outcomes and players' preferences.

Definition 2 (Pareto Optimality) *Let* N *be a finite set of players,* W *a set of alternatives and* $\succeq_i$ *a preference relation over* W*.* $x \in W$ *is called* weakly Pareto optimal *if there is no* $y \in W$ *for which* $y \succ_i x$ *for all* $i \in N$*; it is called* strongly Pareto optimal *if there is no* $y \in W$ *for which* $y \succeq_i x$ *for all* $i \in N$ *and* $y \succ_i x$ *for some* $i \in N$*.*

That an outcome is Pareto Optimal, either in its weak or strong form, suggests that no change to another outcome is possible that makes at least one individual better off without making any other individual worse off.

Example 2 (Pareto Optimality) *To give a flavour of Pareto optimality in games, let us consider the partial convergence game of Figure 2.2. The strategy profile* (U, R) [1] *is neither weakly Pareto optimal nor strongly Pareto optimal, as the strategy profile* (U, L) *is better for both players. However if strategy profile* (U, L) *were not an available outcome,* (U, R)

[1]Notice that in the game representation of Figure 2.2 the strategy profiles identify the outcomes, therefore those games have a bijective outcome function.

would still be weakly Pareto optimal but would not be strongly Pareto optimal any longer. The latter because (L, D) *would be strictly better for at least one player without making the others worse off.*

Cornerstone of game theory is the notion of rational behaviour [6], i.e. what a player should do provided what he can choose and what his preferences are. [2] The concepts of *best response* and *dominant strategy* are possibly the most used to formalize what it means for an individual to act rationally.

Definition 3 (Best Response and Dominant Strategy) *Let* $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ *be a game and let* $i, j \in N$*. Let* σ *be a strategy profile. Player j's strategy* σ_j *is called,*

- best response *if and only if* $\forall \sigma'_j$, $o(\sigma) \succeq_j o(\sigma'_j, \sigma_{-j})$;
- dominant strategy *iff* $\forall \sigma'_j, \forall \sigma'_{-j}$, $o(\sigma_j, \sigma'_{-j}) \succeq_j o(\sigma')$.

Substantially, a strategy is a best response for a player if, fixing the strategies of the other players, there is no other strategy that can guarantee a better outcome; and it is a dominant strategy if it is a best response for all possible strategies of the other players.

Example 3 (Best Response and Dominant Strategy) *In the partial convergence game of Figure 2.2 Row's strategy in the strategy profile* (U, R) *is best response, while Column's strategy in the same strategy profile is not best response, as the first player cannot profitably deviate to* (D, R)*, while the second player can do this, moving to* (U, L)*. For this reason R cannot be a dominant strategy for Column. Notice that U is instead a dominant strategy for Row.*

The Coordination game of Figure 2.2 is also of interest. Even though (U, L) *and* (D, R) *are there strongly Pareto optimal outcomes, and both players have no incentive to deviate from* (U, L) *or* (D, R)*, no player has a dominant strategy.*

Two major *solution concepts*, i.e. sets that contain those outcomes to be reached in a game provided that players act rationally, will be considered: Nash equilibrium, to be referred to alsof as best-response equilibrium (*BR*-equilibrium), and dominant strategy equilibrium (*DS*-equilibrium), that extend Definition 3 to talk about rational behaviour by all players.

Definition 4 (Equilibria) *Let* $\mathbb{G}$ *be a game. A strategy profile* σ *is a:*

- *BR*-equilibrium *(Nash equilibrium) if and only if* $\forall i \in N, \sigma_i$ *is a best response;*
- *DS*-equilibrium *if and only if* $\forall i \in N, \sigma_i$ *is a dominant strategy.*

In words, a strategy profile is a Nash equilibrium if each individual strategy is a best response for the player holding it, and it is a dominant strategy equilibrium if each individual strategy is a dominant strategy for the player holding it.

[2] Aumann's classical definition of rationality ("A person's behavior is rational if it is in his best interests, given his information", [7]) considers epistemic notions such as knowledge and belief with which our work is not concerned.

Example 4 (Equilibria) *To illustrate BR and DS equilibria, let us consider the games described in Example 3. Partial convergence game has one BR-equilibrium, namely* (U, L)*: as already observed, no player has an incentive to deviate from it. The fact that* $(U, L) \prec_{Column} (D, L)$ *is a sufficient condition for a deviation by the column player, as he is not capable of moving to* (D, L) *without the help of his opponent.* (U, L) *is also the unique DS-equilibirum. As for the coordination game, the observations made in Example 3 are sufficient to conclude that* (U, L) *and* (D, R) *are the only BR-equilibria, while no DS-equilibrium is present.*

As previously discussed, strategic game models suggest a representation of simultaneous play. This fact is also witnessed by the formulation of the outcome function, that goes from tuples of strategy profiles to outcomes. However for our purposes it is also convenient to define *subgames*, that are intended to represent game restrictions caused by moves of some players.

Definition 5 (Subgame of a strategic game) *Let* $\mathbb{G} = (N, W, \Sigma_i, \succeq_i, o)$ *be a game,* σ *be a strategy profile, and* $C \subseteq N$*. The subgame of* $\mathbb{G}$ *defined by* σ_C *is a game* $\mathbb{G} \downarrow \sigma_C = (N', W', \Sigma'_i, \succeq'_i, o')$ *such that:*

- $N' = N \setminus C$*;*
- $W' = W \setminus \{s \mid \exists \sigma'$ *such that* $s = o(\sigma')$ *and* $\sigma'_C \neq \sigma_C\}$*;*
- *for all* $i \in N \setminus C$*,* $\Sigma'_i = \Sigma_i$*;*
- *for all* $i \in N \setminus C$*,* $\succeq'_i = \succeq_i$*;*
- $o' : \bigtimes_{i \in N \setminus C} \Sigma'_i \to W'$ *is such that for all* $\sigma' \in \bigtimes_{i \in N \setminus C} \Sigma'_i$*,* $o'(\sigma') = o(\sigma', \sigma_C)$*.*

A subgame $\mathbb{G} \downarrow \sigma_C$ of $\mathbb{G}$ is nothing but what is obtained from $\mathbb{G}$ once the coalitional strategy σ_C of the set of players in C is fixed. It should be thought of as a snapshot representing what is still 'left to play' once the players in C have made their choice.

Example 5 *Matrix representations, such as the ones given in Figure 2.2, are extremely useful to have a grasp of subgames. Consider the prisoner's dilemma for instance, and the strategy* U *by the row player. The subgame* $\mathbb{G} \downarrow U_{Row}$*, where* $\mathbb{G}$ *represents the prisoner's dilemma and* U_{Row} *the fact that the row player plays up, can be represented as a matrix where the move D is cut out of the picture. In this matrix, what is left for Column to play is a choice between* (U, R) *and* (U, L)*. Figure 2.3 applies this transformation to each game in Figure 2.2.*

2.2.2 Coalitional games

In addition to the strategic games of Definition 1 a great amount of attention will be devoted to *coalitional games* (also called *cooperative games*), that abstractly represent the power of groups of players by so-called *effectivity functions* [50]. Effectivity functions are defined as follows:

	L	R
U	2,2	0,3

Prisoner's dilemma

	L	R
U	3,3	2,2

Full convergence

	L	R
U	1,1	0,0

Coordination

	L	R
U	3,3	2,2

Partial convergence

Figure 2.3: Examples of $\mathbb{G} \downarrow U_{Row}$ for each game in Figure 2.2

Definition 6 (Effectivity function) *Let N be a finite sets of players and W a set of outcomes. An* effectivity function *is a function $E : 2^N \to 2^{2^W}$, such that for each $C \subseteq N$, $E(C)$ is closed under supersets.*

An effectivity function assigns to every coalition a set of sets of states. Intuitively, if $X \in E(C)$ the coalition is said to be able to *force* or *determine* that the outcome of the interaction will be some member of the set X. Along these lines, each set of states $X \in E(C)$ will be referred to as a *choice* of coalition C, while the set $E(C)$ will be called the *choice set* of coalition C. Under this interpretation, closure under superset is quite a natural property: if a coalition is able to force the game to end up inside set X then it is also able to force it to end up in each Y with $X \subseteq Y$.

Generally speaking, a number of properties can be assigned to effectivity functions, depending on the features to be modelled. For present purposes the following definitions will come to hand.

Definition 7 *Let $C, C' \subseteq N$ be a coalition. An effectivity function is*

- Closed under finite unions, *if for all C, $E(C)$ is closed under finite unions;*
- Outcome monotonic, *if for all C, $E(C)$ is closed under supersets;*
- Regular, *if for all C, $X \in E(C)$ and $Y \in E(\overline{C})$ implies that $X \cap Y \neq \emptyset$;*
- Superadditive, *if for all C, C' with $C \cap C' = \emptyset$, $X \in E(C)$ and $Y \in E(C')$ implies that $X \cap Y \in E(C \cup C')$;*
- N-maximal, *if $X \notin E(\emptyset)$ implies that $\overline{X} \in E(N)$;*
- Coalition monotonic, *if for all C, C' with $C \subseteq C'$, $X \in E(C)$ implies that $X \in E(C')$;*
- Closed-world *or satisfies* inability of the empty coalition *(IOEC) if $C = \emptyset$ implies that $E(C) = \{W\}$;*
- Playable, *if it is outcome monotonic, superadditive, for all C we have that $E(C)$ contains the unit and $\emptyset \notin E(C)$;*
- Determined, *if playable and closed-world.*

The above mentioned properties acquire a natural reading once we understand effectivity functions as a set of sets that a coalition can force. Closure under finite union says that a coalition being able to force the interaction to end up either in a set A or in a set B can force the interaction to end up in the set $A \cup B$. Outcome monotonicity, that implies closure under finite unions, says that if a coalition is able to force the outcome of the interaction to belong to a particular set, then that coalition is also able to force the outcome to belong to all its supersets. Regularity says that if a coalition is able to force the outcome of an interaction to belong to a particular set, then no possible combinations of moves by the other players can prevent this from happening. When outcome monotonicity holds, an alternative way to express regularity (used for instance in [54]) is the following: for all C, $X \in E(C)$ implies that $\overline{X} \notin E(\overline{C})$, which says that if a coalition can force a particular set then its opponents cannot force the complement of that set. Superadditivity expresses the fact that two disjoint coalitions can join forces, by stating that if a set X can be forced by some coalition C and a set Y by some disjoint coalition D then the intersection of the two sets can be forced by the union of the two coalitions. N-maximality says that if the empty coalition cannot force the interaction to end up in a set X then the coalition made by all players together can force the complement of that set, i.e. $\overline{X}$. Coalition monotonicity states that the bigger a coalition is, the more the properties this coalition will be able to force, i.e. the more the sets in its effectivity function. The IOEC condition requires the empty coalition to be able to bring about only trivial consequences: if we think of an outcome as resulting from the intersection of the choices of opposing coalitions, IOEC guarantees the sets of all players taken together to determine where the interaction will end up. The last conditions, playability and determinacy, will be shown in the coming sections to be crucial when relating effectivity functions and strategic games. Playable effectivity functions have been first introduced in [54] while determined ones in [19].

As is clear from their definition, which requires outcome monotonicity, effectivity functions bear a certain redundancy. The following notion is a description of a *nonredundant* effectivity function.

Definition 8 (Nonmonotonic core) *[54] Let E be an effectivity function. The* nonmonotonic core $E^{nc}(C)$ *for $C \subseteq N$ is the set of minimal sets in $E(C)$:*

$$\{X \in E(C) \mid \neg\exists Y(Y \in E(C) \text{ and } Y \subset X)\}$$

The nonmonotonic core is a description of the minimal sets in an effectivity function. Nevertheless there can be effectivity functions defined on a domain W consisting of an infinite descending chain of sets $W \supset W_1 \supset \ldots \supset W_n \ldots$ for which none of the sets W_i is represented in the nonmonotonic core. In this sense nonmonotonic cores still eliminate too much information from the original effectivity function. To overcome this potential problem, we say that the nonmonotonic core of $E(C)$ is *complete* if it undercuts every set in $E(C)$, i.e., for every $X \in E(C)$ there exists $Y \in E^{nc}(C)$ such that $Y \subseteq X$.

Variations in the definition of effectivity functions are adopted in the literature. Dynamic effectivity functions, defined first in [54], specify at each state what out-

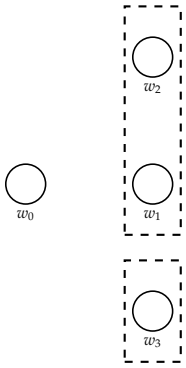

Figure 2.4: Nonmonotonic core of a choices set $E(w_0)(C)$. The dashed squares indicate the minimal sets, the smallest choices available to coalition C from w_0. In case each set $X \in E(w_0)(C)$ is superset of some dashed square, i.e. in case $E(w_0)(C) = \{\{w_1, w_2\}\{w_3\}\}^{\sup}$, the nonmonotonic core of $E(w_0)(C)$ is also complete.

comes a coalition is able to force. They will be used in Chapter 4 and Chapter 6 in connection with logical models of coalitional rationality and are defined as follows.

Definition 9 (Dynamic effectivity function)
Given a finite set of players N and a set of states W, a dynamic effectivity function *is a function* $E : W \to (2^N \to 2^{2^W})$ *such for each* $w \in W$*, and* $C \subseteq N$ *we have that* $E(w)(C)$ *is closed under supersets .*

A dynamic effectivity function E enjoys the properties in Definition 7 whenever they hold for all $E(w)$, for $w \in W$, and the nonmonotonic core of its choice sets can be naturally described. An example is given in Figure 2.4.

Effectivity functions are the characteristic feature of coalitional games, that can now be formally introduced.

Definition 10 (Coalitional game) *A* coalitional game *is a tuple* $\mathbb{C} = (N, W, E, \succeq_i)$ *where:*

- *N is a set of players;*
- *W is a set of outcomes;*
- *E is an effectivity function;*
- $\succeq_i$ *is a total preorder on S.*

Coalitional games will be denoted $\mathbb{C}, \mathbb{C}', \mathbb{C}'', \ldots$. Coalitional game forms will be denoted $\mathbb{F}_c, \mathbb{F}_c', \mathbb{F}_c'', \ldots$. Comparing the definition of coalitional games and that of strategic game it will be immediately clear that in the coalitional case effectivity functions replace the outcome function and the strategy profiles. That effectivity functions are really more abstract than outcome functions and strategy profiles will

soon be observed, but to establish the properties that effectivity functions need to have in order to correspond to strategic games will require some more effort and will be dealt with in the coming chapters. To start with we can observe that a coalitional game can be obtained from a strategic game in a canonical way (cf. [50]), by relating the coalitional effectivity function to the strategies of the players in the strategic game.

Each strategy game has therefore its own effectivity function, known as the α-effectivity function. The α-effectivity function, extensively used in the social choice literature [49, 1], contains those sets in which a coalition C can force the game $\mathbb{G}$ to end up, no matter what $\overline{C}$ does.

Definition 11 (α-Effectivity function) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a strategic game. Its α-effectivity function $E^{\alpha}_{\mathbb{G}}$ is defined as follows:*

$$X \in E^{\alpha}_{\mathbb{G}}(C) \Leftrightarrow \exists \sigma_C \forall \sigma_{\overline{C}}\, o(\sigma_C, \sigma_{\overline{C}}) \in X.$$

Coalitional games can be constructed from strategic ones making use of the α-effectivity function.

Definition 12 (Coalitional games from strategic ones) *Let $\mathbb{G} = (N, W, \Sigma_i, \succeq_i, o)$ be a strategic game. The coalitional game of $\mathbb{G}$ is $\mathbb{C}^{\mathbb{G}} = (N, S, E^{\alpha}_{\mathbb{G}}, \succeq_i)$.*

Example 6 (α-Effectivity) *Let us consider again each game $\mathbb{G}$ of Figure 2.2. Their α-effectivity functions coincide, as in the standard account preferences are not relevant in defining coalitional power and players' strategies share the same labels. $E^{\alpha}_{\mathbb{G}}(\{Column\})$, the α-effectivity function of the column player, comprises the set $\{(U, L), (U, R)\}$, as Column can only decide between L and R, and, for the same reason, the set $\{(D, L), (D, R)\}$, together with all their supersets. Likewise the α-effectivity function of the row player $E^{\alpha}_{\mathbb{G}}(\{Row\})$ is given by the set $\{(U, L), (D, L)\}$, as Row can only decide between U and D, the set $\{(D, R), (U, R)\}$, and all their supersets. As for the other coalitions, we have that $E^{\alpha}_{\mathbb{G}}(\emptyset) = \{W\}$, as the empty coalition does not interfere in the choice of the outcome, and that $E^{\alpha}_{\mathbb{G}}(N) = 2^W \setminus \emptyset$, as players together can choose any possible outcome.*

$E^{\alpha,nc}_{\mathbb{G}}(\{Column\})$, the nonmonotonic core of the α-effectivity function of the column player, is given by the set $\{(U, L), (U, R)\}$ together with the set $\{(D, L), (D, R)\}$. As for the row player $E^{\alpha,nc}_{\mathbb{G}}(\{Row\})$ is given by the set $\{(U, L), (D, L)\}$ together with the set $\{(D, R), (U, R)\}$. As for the remaining coalitions $E^{\alpha,nc}_{\mathbb{G}}(\emptyset)$ remains $\{W\}$ while $E^{\alpha,nc}_{\mathbb{G}}(N) =$ $\{(U, L), (D, R), (U, R), (D, L)\}$. Notice that all the nonmonotonic cores are complete.

The class of α-effectivity functions is clearly included in the class of effectivity functions. One central question of our work is to establish the exact nature of this inclusion. At the present stage we can make it more precise, showing that the class of effectivity functions that are α-effectivity functions of strategic games is included in the class of playable effectivity functions, as witnessed by the following results.

Proposition 2 *Let $\mathbb{G} = (N, W, \Sigma_i, \succeq_i, o)$ be a game and $E^{\alpha}_{\mathbb{G}}$ its α-effectivity function.*

Then the following holds:

- $E^{\alpha}_{\mathbb{G}}$ *is playable;*
- $E^{\alpha}_{\mathbb{G}}$ *is determined whenever o is surjective.*

Proof *The proof is a straightforward check of the conditions. The argument of* $E^{\alpha}_{\mathbb{G}}$ *being playable first appears in [54].*

The nonmonotonic core of α-effectivity functions is of particular interest for the coming part of the work. Proposition 2 has shown that α-effectivity functions are playable. The nonmonotonic core of the empty coalition in playable effectivity functions will be of particular relevance in establishing subsequent results.

Proposition 3 *For every playable effectivity function E:*

1. $E(\emptyset)$ *is a filter.*
2. $E^{nc}(\emptyset)$ *is either empty or a singleton.*

Proof 1. $E(\emptyset)$ *is non empty by safety; it is closed under supersets by outcome monotonicity, and under intersections by superadditivity (with respect to the empty coalition).*

2. *Suppose* $E^{nc}(\emptyset)$ *is nonempty, and let* $X, Y \in E^{nc}(\emptyset)$*. Then,* $\emptyset$ *is effective for each of* X *and* Y*, hence, by superadditivity, it is effective for* $X \cap Y$*. By the definition of* $E^{nc}(\emptyset)$*, it follows that* $X = X \cap Y = Y$*. Suppose instead that* $E^{nc}(\emptyset)$ *is not a singleton and that is nonempty. Then there are* $X, Y \in E^{nc}(\emptyset)$ *with* $X \neq Y$*. By superadditivity* $X \cap Y \in E^{nc}(\emptyset)$ *and we have that* $X \cap Y \subset Y$ *or* $X \cap Y \subset X$*. Contradiction.*

Proposition 2 has established that the class of playable effectivity functions includes the class of effectivity functions that are α-effectivity function of some strategic game; consequently Proposition 3 applies to all α-effectivity functions. However α-effectivity functions enjoy other interesting properties, that can be observed looking at their nonmonotonic core.

Proposition 4 *For every* α*-effectivity function* $E^{\alpha}_{\mathbb{G}} : 2^N \to 2^{2^W}$*, the following hold:*

1. *The nonmonotonic core of* $E^{\alpha}_{\mathbb{G}}(\emptyset)$ *is the singleton set* $\{Z\}$ *where*

$$Z = \{x \in W \mid x = o(\sigma_N) \text{ for some } \sigma_N\}.$$

2. $E^{\alpha}_{\mathbb{G}}(\emptyset)$ *is the principal filter generated by* Z*.*

Proof *For both claims it suffices to observe that* $Z \in E^{\alpha}_{\mathbb{G}}(\emptyset)$ *and that for every* $U \in E^{\alpha}_{\mathbb{G}}(\emptyset)$*,* $Z \subseteq U$*. Therefore,* $E^{nc}(\emptyset) = \{Z\}$ *for* $E = E^{\alpha}_{\mathbb{G}}$ *and* $E^{\alpha}_{\mathbb{G}}(\emptyset)$ *is the principal filter generated by* Z*.*

We can observe that while Proposition 3 has formulated the nonmonotonic core of playable effectivity functions in terms of filters, Proposition 4 has formulated the nonmonotonic core of α-effectivity functions in terms of principal filters. The results in Chapter 3 will show that this difference is crucial if want to characterize the class of effectivity functions that correspond to, or *represent*, strategic games. Formally

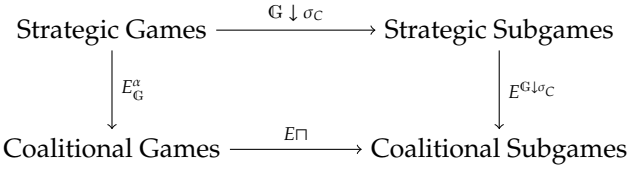

Figure 2.5: Relations between games and subgames. The operation $\mathbb{G} \downarrow \sigma_C$ (Definition 5), that restricts a strategic game with a strategy of a group of players, transforms a game into a subgame, while $E^{\mathbb{G}\downarrow\sigma_C}$ (Definition 13) describes what coalitions can do in a subgame. $E^{\alpha}_{\mathbb{G}}$ (Definition 11) describes the α-effectivity function of a strategic game, extracting its coalitional structure. The picture will be completed in Chapter 3, defining the operation $E\sqcap$ of choice restriction of an effectivity function, that transforms a coalitional game into a coalitional subgame.

an effectivity function E such that $E = E^{\alpha}_{\mathbb{G}}$, for $\mathbb{G}$ being a strategic game, is said to *represent* $\mathbb{G}$. For a dynamic effectivity function the same terminology applies. In that case, if for some $w \in W$, $E(w) = E^{\alpha}_{\mathbb{G}}$, E is said to *represent* $\mathbb{G}$ at world w.

In the same way we have done for strategic games, representability can be lifted to subgames. This definition will turn out to be useful when studying coalitional rationality in strategic games. To lift representability to subgames we introduce the notion of effectivity function for a subgame, that mimics the features of Definition 5.

Definition 13 (Coalitional subgames) *Let $\mathbb{G} = (N, W, \Sigma_i, \geq_i, o)$ be a game. The coalitional subgame $\mathbb{C}^{\mathbb{G}\downarrow\sigma_C} = (N', S', E^{\mathbb{G}\downarrow\sigma_C}, \geq'_i)$ of $\mathbb{G}$ is a coalitional game where the entries different from $E^{\mathbb{G}\downarrow\sigma_C}$ follow Definition 5 and the sub-α-effectivity function $E^{\mathbb{G}\downarrow\sigma_C}$ is defined for each $C' \subseteq \overline{C}$ as follows:*

$$X \in E^{\mathbb{G}\downarrow\sigma_C}(C') \Leftrightarrow \exists \sigma_{C'} \forall \sigma_{\overline{C}\setminus C'},\ o(\sigma_{\overline{C}}, \sigma_C) \in X.$$

Intuitively, a set X belongs to $E^{\mathbb{G}\downarrow\sigma_C}(C')$ when coalition C' is able to force the outcome of the game to end up in X *provided* coalition C has chosen strategy σ_C. The coalitional subgame (Definition 13) reflects the notion of subgame (Definition 5) in the same way the notion of coalitional game (Definition 11) reflects the notion of game (Definition 1). Figure 2.5 illustrates the relation between the structures.

As to solution concepts for coalitional games we consider the *core*, "the cooperative solution concept that is perhaps best known to economists" [6]. The core is based on a dominance relation among outcomes: an outcome x is dominated by an outcome y if there is a coalition that can achieve y and whose members prefer it to x. The core collects the outcomes that are not dominated. As Abdou and Keiding put it in [1] (p.53) "this notion captures the idea that group choice should be robust against coalitional improvements, i.e. no coalition should be so badly off in society's choice that it could by itself establish something better for everyone in the coalition".

Definition 14 (The core) *Let $\mathbb{C} = (N, W, E, \succeq_i)$ be a coalitional game. We say that a state s is dominated in $\mathbb{C}$ if for some C and $X \in E(C)$ it holds that $x \succ_i s$ for all $x \in X, i \in C$. The core of $\mathbb{C}$, in symbols $CORE(\mathbb{C})$ is the set of undominated states.*

Intuitively, the core is the set of those states in the game that are stable, i.e., for which there is no coalition that is at the same time able and interested to deviate from them. An other appealing rewording, again by Abdou and Keiding, is that of considering the outcomes not in the core as those encountering an "effective opposition" [1] (p.52).

2.3 Logic

The logical systems that we will work with are modal languages, i.e. extensions of the language of propositional logic with a set of modal operators $\Box_1, ..., \Box_n, \ldots$, defined on a countable set of atomic propositions $Prop = \{p_1, p_2, \ldots\}$, on which the set of *well-formed formulas* is inductively built [67]. Each well-formed formula φ of a modal language $\mathcal{L}$, henceforth simply called *formula*, is defined as follows:

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

where $\Box_i \in \{\Box_1, \ldots, \Box_n, \ldots\}$ and $p \in Prop$.

A model for this language is a tuple $M = (W, R_1, \ldots, R_n, \ldots, V)$, consisting of a set of *worlds* or *states* W commonly referred to as *domain*; an *accessibility relation* R_i for each modal operator $\Box_i$, defined via so-called neighbourhood functions [23], i.e. functions $R_i : W \to 2^{2^W}$; and a *valuation function* $V : Prop \to 2^W$ that assigns to each atomic proposition a subset of W, with the intuitive understanding that each atomic proposition is assigned the set of worlds where this proposition is true. A model without a valuation function is called a *frame*. As a general convention a multimodal language with modalities $\Box_1, \ldots, \Box_n, \ldots$ will be denoted by $\mathcal{L}^{f(\Box_1),\ldots,f(\Box_n),\ldots}$, specifying its modal operators where the function f associates to each modality a symbol representing it. The symbols will be systematically introduced as intuitive shorthands for the modalities. Let Δ be a modal language consisting of modalities $\Box_1, ..., \Box_n, \ldots$ and let $M = (W, R_1, \ldots, R_n, \ldots, V)$ be a model for this language. The satisfaction relation of a formula $\varphi \in \Delta$ with respect to a pair (M, w), where $w \in W$, is defined according to the following truth conditions:

$$\begin{aligned}
M, w \models p \quad &\text{if and only if} \quad w \in V(p) \\
M, w \models \neg\varphi \quad &\text{if and only if} \quad M, w \not\models \varphi \\
M, w \models \varphi \wedge \psi \quad &\text{if and only if} \quad M, w \models \varphi \text{ and } M, w \models \psi \\
M, w \models \Box_i\varphi \quad &\text{if and only if} \quad \varphi^M \in R_i(w)
\end{aligned}$$

where $\varphi^M = \{w \in W \mid M, w \models \varphi\}$ is called the *truth set* or the *extension* of φ.

A formula φ of a modal language Δ:

- *holds at a state w* of model M whenever $M, w \models \varphi$;
- is *valid in a model M*, denoted $\models_M \varphi$, if and only if $M, w \models \varphi$ for every $w \in W$, where W is the domain of M;
- is *valid in a class of models $\mathcal{M}$*, denoted $\models_{\mathcal{M}} \varphi$, if and only if it is valid in every $M \in \mathcal{M}$;
- is *valid in a frame F*, denoted $\models_F \varphi$, if and only if for every valuation V we have that $\models_{(F,V)} \varphi$;
- is *valid in a class of frames $\mathcal{F}$*, denoted $\models_{\mathcal{F}} \varphi$, if and only if it is valid in every $F \in \mathcal{F}$.

The set of formulas of Δ that are valid in a class of models $\mathcal{M}$ is denoted $\Delta_{\mathcal{M}}$ (for frames the denotation is $\Delta_{\mathcal{F}}$). For a set of formulas Σ, we write $M, w \models \Sigma$ to say that $M, w \models \sigma$, for all $\sigma \in \Sigma$. We say that a set of formulas Σ semantically entails a formula φ in a class of models $\mathcal{M}$, denoted $\Sigma \models_{\mathcal{M}} \varphi$, if for every $M \in \mathcal{M}$ we have that $\models_M \Sigma$ implies $\models_M \varphi$.

A modal rule

$$\frac{\varphi_1, ..., \varphi_n}{\psi} \tag{2.1}$$

is sound in a class of models $\mathcal{M}$ if $\varphi_1, ..., \varphi_n \models_{\mathcal{M}} \psi$.

Let us recall, following [23], that a modal logic Δ is called *classical* if it is closed under the rule of equivalence, i.e. for each $\Box$ in the language Δ we have:

$$\frac{\varphi \leftrightarrow \psi}{\Box\varphi \leftrightarrow \Box\psi} \tag{2.2}$$

It is called *monotonic* if it is classical and it is moreover closed under the rule of monotonicity, i.e. for each $\Box$ in the language Δ we have:

$$\frac{\varphi \rightarrow \psi}{\Box\varphi \rightarrow \Box\psi} \tag{2.3}$$

It is called *normal* if it is monotonic, it is closed under the rule of generalization and contains the K axiom, i.e. for each $\Box$ in the language Δ we have

$$\frac{\varphi}{\Box\varphi} \tag{2.4}$$

and $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$.

2.3.1 Coalition Logic

The logical language used to reason about effectivity functions is Coalition Logic [54]. Coalition Logic is multimodal language, where modalities are of the form $[C]\varphi$ and represent the fact that a certain coalition C can force a certain formula φ to be true. The language of Coalition Logic is denoted $\mathcal{L}^{[C]}$ and it is made by formulas that are defined as follows:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [C]\varphi$$

where p ranges over *Prop* and C ranges over the subsets of N. The other boolean connectives are defined as usual.

The modalities are interpreted in neighbourhood structures [23] induced by the effectivity functions and called *Coalition Models*.

Definition 15 (Coalition Models) *A* Coalition Model *is a triple*

$$(W, E, V)$$

where:

- W *is a nonempty set of states;*
- $E : W \longrightarrow (2^N \longrightarrow 2^{2^W})$ *is a* dynamic *effectivity function;*
- $V : W \longrightarrow 2^{Prop}$ *is a valuation function.*

The satisfaction relation of the formulas of the form $[C]\varphi$ with respect to a pair M, w is defined as follows:

$$M, w \models [C]\varphi \quad \text{if and only if} \quad \varphi^M \in E(w)(C)$$

where, $\varphi^M = \{w \in W \mid M, w \models \varphi\}$. As outcome monotonicity is taken to be a property of all effectivity functions, the rule of monotonicity is valid in Coalition Logic, which is therefore a monotonic modal logic [39]. Figure 2.6 gives an example of a Coalition Model.

The rule of monotonicity takes this form for each $C \subseteq N$:

$$\frac{\varphi \rightarrow \psi}{[C]\varphi \rightarrow [C]\psi} \tag{2.5}$$

As usual with neighbourhood structures, relations between set theoretical and logical properties are fairly immediate to spot. Standard correspondence results between class of frames and neighbourhood functions [23] can be automatically used for Coalition Logic.

Proposition 5 *Let $F = (W, E)$ be a Coalition Frame, and C, C' arbitrary coalitions. The following results hold:*

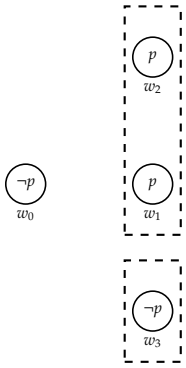

Figure 2.6: Coalition Models. The modalities are interpreted in dynamic effectivity functions that specify the neighbourhood function. In the picture the effectivity function $E(w_0)(N) = \{\{w_2, w_1\}, \{w_3\}\}^{\text{sup}}$ — as usual the minimal sets are represented by the dashed lines — and the valuation function $V(p) = \{w_1, w_2\}$ — represented by the atomic proposition assigned to the worlds where it is satisfied —make sure that the following statements hold: $M, w_0 \models [N]p$, i.e. at w_0 coalition N can achieve p and $M, w_0 \models [N]\neg p$, i.e. at w_0 coalition N can achieve $\neg p$.

- $\models_F [C]\top$ *if and only if for all* $w \in W$, $E(w)(C)$ *contains the unit;*
- $\models_F \neg[\emptyset]\neg\varphi \rightarrow [N]\varphi$ *if and only if E is regular and outcome monotonic;*
- $\models_F \neg[C]\bot$ *if and only if* $\emptyset \notin E(w)(C)$ *for each* $w \in W$*;*
- $\models_F [C']\varphi \wedge [C'']\psi \rightarrow [C' \cup C''](\varphi \wedge \psi)$ *if and only if E is superadditive;*
- $\varphi \rightarrow \psi \models_F [C]\varphi \rightarrow [C]\psi$ *if and only if E is outcome monotonic.*

Proof *The proof is standard and given in [54].*

Correspondence results allow us to distinguish by modal means a number of class of frames. However expressivity of the modal operators strongly limit the capacity of the language to discern classes of structures. To this extent the reader should notice that the logics of both determined and playable effectivity frames share the fact that $\models_F [\emptyset]\top$. However this proposition, whose interpretation is that for each $w \in W, \{W\} \in E(w)(\emptyset)$, is not sufficient to make a formal distinction between $E(w)(\emptyset)$ in the two different classes of effectivity functions, which will be a topic of discussion in Chapter 3.

Cooperative Game Models

Coalition models, which are nothing but cooperative game forms with a valuation function, can be naturally extended with preference relations.

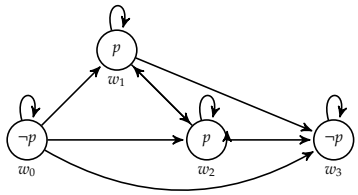

Figure 2.7: Preference Models. The arrows represent the relation $\preceq_i$, that links worlds according to the desirability of player i. The valuation given ensures that the following holds: $M, w_0 \models \Box_i^{\preceq} \Diamond_i^{\preceq} \neg\varphi$.

Definition 16 *We call* Cooperative Game Model *a Coalition Model extended by a preference relation $\preceq_i$ for each $i \in N$.*

A Cooperative Game Model without a valuation function, i.e. a tuple $(N, W, E, \succeq_i)$ will be also referred to as a Cooperative Game Frame.

2.3.2 Preference Logic

In order to talk about the preference relations of a Cooperative Game Model two standard modalities to reason about preferences [17, 63, 66] will be used.

The first, $\Diamond_i^{\preceq}$, indicates what holds at some world that is better than the present one, the second, $\Diamond_i^{\succ}\varphi$, indicates what holds at some world that is strictly worse than the current one. Their interpretation is given as follows:

$$M, w \models \Diamond_i^{\preceq}\varphi \quad \text{if and only if} \quad M, w' \models \varphi, \text{ for some } w' \text{ with } w \preceq_i w'$$

$$M, w \models \Diamond_i^{\succ}\varphi \quad \text{if and only if} \quad M, w' \models \varphi, \text{ for some } w' \text{ with } w' \prec_i w$$

Formulas with an occurrence of a weak preference operator such as $\Box_i^{\preceq}\varphi$ are meant to express the fact that φ is a property that holds in all worlds that are worse than the current one.

As is clear from the interpretation of the modal operators fundamental properties of the language, such as validity of certain formulas, are strictly dependent on the underlying relation. Also for preference total preorders, standard correspondence result are of great use.

Proposition 6 *Let $F = (W, \succeq_i)$ be a frame with a preference relation $\succeq_i$. The following results hold:*

- $\models_F \varphi \to \Diamond_i^{\preceq}\varphi$ *if and only if* $\succeq_i$ *is reflexive;*
- $\models_F \Diamond_i^{\preceq}\Diamond_i^{\preceq}\varphi \to \Diamond_i^{\preceq}\varphi$ *if and only if* $\succeq_i$ *is transitive;*

Proof *The results are standard and given in [10].*

The reader will notice that the property of connectedness has not been characterized. In fact this is not possible in a modal language like $\mathcal{L}^{\preceq}$, that can only talk about local properties of relations [10]. Some exceptional cases are however noteworthy. $\mathcal{L}^{\preceq}$ is extremely close to Boutilier's *CO*-logic, [17] a modal logic defined on modalities $\Diamond^{\preceq}, \Diamond^{\prec}$, interpreted on a total preference preorder and expressive enough to be able to define conditional preferences, for example sentences such as "from all worlds satisfying formula φ, the best worlds satisfy formula ψ", which will be object of study in Chapter 4.

For the purpose of expressing global properties of relations is convenient to use a universal (or global) modality A. This modality expresses properties of all the states in a domain W of a model M and it is interpreted as follows.

$$M, w \models A\varphi \quad \text{if and only if} \quad M, w' \models \varphi, \text{ for all } w' \in W$$

The formula $\neg A \neg \varphi$ will be abbreviated $E\varphi$. The symbol E is the existential dual of A and it indicates that a certain formula holds at some state in the model. Notice that with the global modality we have a genuine addition of expressivity (together with further gains, as shown in [31]), therefore we can express validity in a model by truth in a world, witness the fact that $M, w \models A\varphi \Leftrightarrow \models_M \varphi$.

The following frame correspondence results should give a flavour of the power of the global modality together with the preference and coalition logic modalities.

Proposition 7 *Let $F = (W, E, \succeq_i)$ be a frame with a preference relation $\succeq_i$ and an effectivity function W. The following results hold:*

- $F \models (\varphi \wedge \Box_i^{\preceq} \psi) \rightarrow A(\psi \vee \varphi \vee \Diamond_i^{\preceq} \varphi)$ *if and only if $\succeq_i$ is trichotomous;*
- $F \models A\varphi \leftrightarrow [\emptyset]\varphi$ *if and only if E has IOEC.*

Proof *The first case is standard [10]. For the second case, assume that $\models_F A\varphi \leftrightarrow [\emptyset]\varphi$ while not $E(w)(\emptyset) = \{W\}$ for every w in any frame $F = (W, E)$ in the class of Coalitional Frames $\mathcal{F}$. As both W and $E(w)(\emptyset)$ are nonempty, there is a $W' \neq W$ s.t $W' \in E(w)(\emptyset)$. Let us construct a model M, based on F that consists of the following valuation function: each atom p is false in all $w' \in W'$ and true in $W \setminus W'$. Now $M \not\models Ap \leftrightarrow [\emptyset]p$. Contradiction. For the other direction assume $E(w)(\emptyset) = \{W\}$ for a given w in an arbitrary model M based on F, and that $M, w \models A\varphi$. Then $\varphi^M = W$ and $M, w \models [\emptyset]\varphi$ follows. Assume now that $M, w \models [\emptyset]\varphi$. It has to be the case that $\varphi^M = W$ by assumption. So also that $M, w \models A\varphi$, which concludes the proof.*

Proposition 7 shows that empowering the language $\mathcal{L}^{[C],\preceq}$, the language of Coalition Logic extended with the $\Diamond_i^{\preceq}$ modality, with the global modality allows both the characterization of connected frames and the distinction between playable and determined effectivity functions. In other works the global modality, combined with

preference logics, has been shown to be expressive enough to define binary combination of preferences over formulas [66], bringing further the achievements of [17].

For all these reasons the language $\mathcal{L}^{[C],\preceq,g}$, i.e. the language $\mathcal{L}^{[C],\preceq}$ enriched with the global modality, will be our starting point for reasoning about the strategic aspect of interaction introduced in Chapter 1.

Part I

Strategic Reasoning and Coalitional Games

Chapter 3

Coalitional Games

Normative aspects of game theory may be subclassified using various dimensions. One is whether we are advising a single player (or group of players) on how to act best in order to maximize payoff to himself, if necessary at the expense of other players; and the other is advising society as a whole (or a group of players) of reasonable ways of dividing payoff among themselves. The axis I'm talking about has the strategist (or the lawyer) at one extreme, the arbitrator (or judge) at the other.

Robert J. Aumann, *What is game theory trying to accomplish?* [6]

3.1 Introduction

In cooperative game theory [56, 50], but also in related disciplines such as social choice theory [49, 1], effectivity functions have been used as an abstract representation of coalitional power. In its general form given in Definition 6, this abstraction does not consider players' preferences nor opponent possibilities but it is limited to providing a set of possible choices for a coalition without committing to a model of rational decision making, i.e. what coalitions should do with those possible choices. The purpose of this chapter is to provide a model of coalitional rationality where coalitions are treated as decision makers that, by coordinating their members, have an opinion on the possible outcomes of the interaction, reflect on their opponents' possibilities and take a joint decision.

The theory of games has come up with rigorous models of rational decision making for individuals, which can be immediately illustrated by a classical example.

Example 7 (Motivating Example) *Let us consider a version of the prisoner's dilemma depicted in Figure 2.2, already extensively discussed in the introductory chapter. We first focus on Row and reflect on what is best for him to do. After the choice L by Column, the choice D becomes preferable to the choice U — yielding* $(3,0)$ *instead of* $(2,2)$*. The same holds in case Column moved R — yielding* $(1,1)$ *instead of* $(0,3)$*. These two facts, exhausting all possibilities, are enough to conclude that whatever Column does Row is better off by playing D. If we now turn to Column we get a similar conclusion: playing R is better than playing L, no matter what Row does. In the common sense reading the choices U and L*

are cooperative moves respectively of Row and Column, while D and R are their defective counterparts. What gives the situation the flavour of a dilemma is that, as already observed in Example 3, the outcome (D, R) *is not Pareto optimal and yet it is the one that has been argued for on rational grounds.*

Looked at from an individual perspective, a prisoner's dilemma is an interactive situation in which the advantages of cooperation are overruled by the incentive for individual players to defect. But looked at from a coalitional perspective, the situation changes: using the tools presented in Definition 11, we can extract from the prisoner's dilemma the information regarding the strategic ability of the groups of players involved. Considering the reasoning patterns of coalitions instead of individuals makes new possibilities arise.

Example 8 (Motivating Example cont.) *Even though cooperative moves* U *and* L *may be irrational from an individual perspective, the coalition* $\{Row, Column\}$ *can choose the strategy profile* (U, L) *without fearing to be ruling out a strictly better alternative. In fact, the only outcome that coalition* $\{Row, Column\}$ *would rationally not choose is the outcome* (D, R)*, i.e. the result of individually rational reasoning:* (D, R) *is the only outcome among the available ones for which there exists an alternative that would make both players better off.*

Together with the prisoner's dilemma there are plenty of strategic games in which larger coalitions would rationally choose outcomes that their members would not choose on individual grounds. We believe that an appropriate model of coalitional rationality should be able to make this distinction. However, as previously observed, cooperative games furnish a rather abstract representation of coalitional power and cannot express concepts such as rationality of a coalitional choice. The chapter will bridge the gap by formulating a model of coalitional decision making based on coalitional power and individual preferences.

We will consider:

- A number of preference orders over choices within a coalitional effectivity function, in order to compare what choices are better for a certain coalition;
- The notion of choice restriction, seen as restriction on a coalitional effectivity function induced by the opponents' moves, in order to reason on situations brought about by the possible moves of the other players;
- A definition of optimal choice based on a combination of the previous two notions, in order to establish what is best to do for a certain coalition.

The added value of working out a theory of optimal choices in effectivity functions is that of being able to talk about coalitional rationality in strategic games, but also in extremely abstract coalitional games that do not correspond to any strategic game. In interactions when a coalition C can bring about the property X and and its opponent $\overline{C}$ can bring about $\overline{X}$ — i.e. whose effectivity function lacks the property of playability typical of strategic games (Proposition 2) — understanding what a

coalition should do is even more necessary. Our model of coalitional rationality will be general enough to be applicable to those types of interactions.

Even though the theory of optimal choices will be formulated for the general case, the peculiar features of strategic games will be dealt with:

- On the one hand, we will characterize the class of effectivity functions that correspond to strategic games, correcting a well-known result in the literature from [54];
- On the other hand, we will discuss the various meanings of choice in strategic interaction and propose an alternative to effectivity functions, in order to abstractly represent coalitional power in games.

The models of coalitional rationality, elaborated for the general case, will naturally incorporate the features of coalitional ability in strategic games.

Chapter structure: In Section 3.2, based on joint work with Jan Broersen, Rosja Mastop and John-Jules Meyer [18], a notion of coalitional rationality is studied, introducing an order on coalitional effectivity functions that takes into account individual preference relations and opponents' possibilities. Section 3.3 deals with the relation between coalitional games and strategic games. In Subsection 3.3.1, based on joint work with Wojtek Jamroga and Valentin Goranko [30], a relation is established between strategic games and a class of effectivity functions, correcting a believed correspondence in the literature. Subsection 3.3.2, based on joint work with Jan Broersen, Rosja Mastop and John-Jules Meyer [19], elaborates on alternative representations of coalitional ability in games to model desirable features of strategic interaction. A section discussing the achievements and the related literature will conclude the chapter.

3.2 Coalitional Rationality

Cooperative games have been introduced in Chapter 2 as structures endowed with a preference relation and an effectivity function. Due to the abstraction level of these two constructs, coalitional games can only provide a general representation of what a group of players can force and of what its members individually prefer. However, in order to model coalitional rationality, we need to express that within an effectivity function a certain choice is the best among the available ones. Hereby we are confronted with four problems:

1. The preference relations are formulated for individual players, but we are interested in coalitional preferences.
2. The preference relations are formulated as relations over outcomes, but we are interested in preferences within effectivity functions, which are sets of sets of outcomes.

3. The classical notions of optimality, such as Pareto optimality (Definition 2), do not concern players' strategic possibilities, but we are interested in formulating a notion of optimality within an effectivity function.

4. The classical notions of optimality do not concern opponents' strategic possibilities, but we are interested in formulating the notion of optimality that takes into account what the opponents can do.

Our objective is to address these points one by one, formulating a notion of coalitionally rational choice as an analogue to the notion of dominant strategy available in strategic games (Definition 3). Section 3.2.1 will study a lifting of preference relations to sets, addressing point 2; Section 3.2.2 will make use of this lifting to define a notion of optimality within a coalitional effectivity function without taking into account the opponents' possibilities — what we call a *Pareto optimal choice* — addressing points 1 and 3; Section 3.2.3 will finally endow Pareto optimal choices with the capacity of reasoning on the opponents — what we call an *undominated choice* — addressing the last point. A number of results will show that these notions are natural generalizations of the notion of dominant strategy in strategic games for the coalitional case.

3.2.1 Lifting preferences to sets

As effectivity functions are sets of sets of states and preference relations are orders on states, the first step towards a model of coalitional rationality is to define a lifting of preference relations from states to sets of states. A number of contributions already exist in the literature tackling this and various results are available [27, 56, 42, 46, 65, 41]. However none of them has handled the problem of preference lifting within a coalitional effectivity function, which is the concern of the present chapter.

Being interested in comparing choices according to individual preferences, we will consider a relatively small number of feasible preference orders on choices, namely those relating pairs of sets according to a quantified comparison of their elements. More specifically for two sets X, Y we will consider preference relations of the form

$$X \succsim_i^{(Q_1,Q_2)} Y$$

where Q_1 and Q_2 can be an existential or a universal quantifier and $\succsim_i$ either a weak or a strict preference order. For example $X \succsim_i^{(\exists,\forall)} Y$ means that for player i there exists an element in X that is better than all elements of Y. Figures 3.1, 3.2, 3.4, 3.3 illustrate the types of lifting.

Their formal definition goes as follows.

Definition 17 (Individual preferences for sets of states) *Let $\succsim_i$ be a preference order, W a set of alternatives, and $Q_1, Q_2 \in \{\forall, \exists\}$. Then $\succsim_i^{(Q_1,Q_2)} \subseteq 2^W \times 2^W$ is defined as follows:*

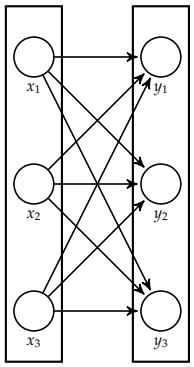

Figure 3.1: The $(\forall, \forall)$ lifting. The arrows indicate the relation $\succeq_i$ for player i. The left set X is better than the right set Y, as each element $x \in X$ is better than each element $y \in Y$.

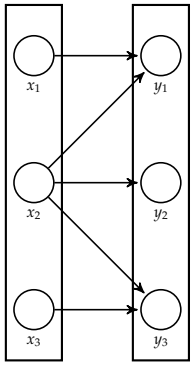

Figure 3.2: The $(\forall, \exists)$ lifting. The left set X is better than the right set Y, as for each element $x \in X$ there is an element $y \in Y$ with $x \succeq_i y$.

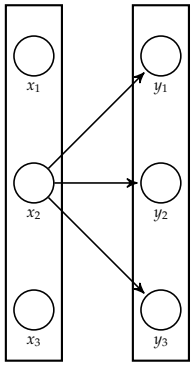

Figure 3.3: The $(\exists, \forall)$ lifting. The left set X is better than the right set Y, as there is an element $x \in X$ that is better than all elements $y \in Y$.

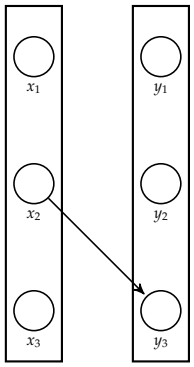

Figure 3.4: The $(\exists, \exists)$ lifting. The left set X is better than the right set Y, as there is an element $x \in Y$ and an element $y \in Y$ with $x \succeq_i Y$.

$$X \succeq_i^{(\forall,\forall)} Y \quad \text{if and only if for all } x \in X, y \in Y, \quad x \succeq_i y$$

$$X \succeq_i^{(\forall,\exists)} Y \quad \text{if and only if for all } x \in X \text{ there exists } y \in Y \text{ such that} \quad x \succeq_i y$$

$$X \succeq_i^{(\exists,\forall)} Y \quad \text{if and only if there exists } x \in X \text{ such that for all } y \in Y \quad x \succeq_i y$$

$$X \succeq_i^{(\exists,\exists)} Y \quad \text{if and only if there exists } x \in X, y \in Y \text{ such that} \quad x \succeq_i y$$

For the strict order $\succ_i^{Q_1,Q_2} \subseteq 2^W \times 2^W$ the definition is obtained by substituting every occurrence of $\succeq_i$ with $\succ_i$.

The preference liftings are applicable to a variety of settings but look particularly suited to modelling strategic decisions. To this extent in [66], which studies similar liftings, it is noticed how they "make excellent sense while choosing best moves in a game"[66].

Let us look again at games with the new definitions at hand.

Example 9 (Lifting the prisoners) *Definition 12 and Example 6 have shown how the effectivity function of a two person strategic game like the prisoner's dilemma should be represented. Its preference order $\succeq_i$ for $i \in N = \{Row, Column\}$ is instead induced by the numerical entries in the matrix as expected. The following statements are representative of the way of comparing arbitrary sets of outcomes in our game:*

- $\{(R,U),(R,D),(L,U)\} \succ_{Column}^{(\forall,\forall)} \{(L,D)\}$, *i.e. the worst that can happen to Column is that he cooperates while Row defects;*
- $\{(R,D)\} \succ_{Column}^{(\forall,\exists)} \{(R,U),(R,D),(L,U),(L,D)\}$, *i.e. defection by both players is not the worst that can happen to Column;*
- $\{(L,D),(R,U)\} \succeq_{Column}^{(\exists,\forall)} \{(R,D),(L,U)\}$, *i.e. in some cases diffformity of choice (one player cooperates while the other defects) can be better for Column than any uniform choice (both players cooperating or both defecting);*
- $\{(L,D),(R,U)\} \succeq_{Column}^{(\exists,\exists)} \{(R,D),(L,U)\}$, *i.e. in some cases diffformity of choice (one player cooperates while the other defects) can be better for Column than some uniform choice (both players cooperating or both defecting);*

Different preference liftings emphasize different aspects of betterness. The $(\forall,\forall)$ lifting for instance emphasizes a notion of *absolute* betterness: expressions of the form $\{(R,U),(R,D),(L,U)\} \succ_{Column}^{(\forall,\forall)} \{(L,D)\}$ mean that all states in the left set are strictly better for *Column* than all states in the right set. Clearly when two sets are in this relation, in order to exclude the choice of the second set, it is even needless to look at what the opponents might do. The $(\forall,\exists)$ lifting is instead close to a notion of *safe* betterness: $\{(R,D)\} \succ_{Column}^{(\forall,\exists)} \{(R,U),(R,D),(L,U),(L,D)\}$ means that it is safer for *Column* to choose $\{R,D\}$ than a set containing all outcomes, as *Column* might *risk*

ending up in $\{(L, D)\}$. It should be kept in mind, though, that the notion of risk is simply a synonym of presence of multiple possibilities and does not (yet) make the role of the opponents' possible moves explicit. The other two types of preference liftings, $(\exists, \forall)$ and $(\exists, \exists)$, function as a sort of dual of the $(\forall, \forall)$ and $(\forall, \exists)$ types: they emphasize that a choice A is not to be absolutely dispreferred to a choice B in case $A \geq_{Column}^{(\exists,\forall)} B$ or in case $A \geq_{Column}^{(\exists,\exists)} B$.

Of particular interest is the role of preference liftings within a coalitional effectivity function.

Example 10 (Lifting the prisoners (cont.)) *Let us now focus on the sets in $E(w)(Column)$ and see how the preference liftings can be applied there. We have that:*

- *neither $\{(U, R), (D, R)\} \geq_{Column}^{(\forall,\forall)} \{(U, L), (D, L)\}$ nor $\{(U, L), (D, L)\} \geq_{Column}^{(\forall,\forall)} \{(U, R), (D, R)\}$, as $(L, D) \geq_{Column} (R, U)$ and $(R, D) \geq_{Column} (L, U)$, i.e. cooperating is not absolutely better than defecting nor is defecting absolutely better than cooperating.*

- *$\{(U, R), (D, R)\} \geq_{Column}^{(\forall,\exists)} \{(U, L), (D, L)\}$ while not $\{(U, L), (D, L)\} \geq_{Column}^{(\forall,\exists)} \{(U, R), (D, R)\}$, i.e. defection is 'safer' than cooperation.*

Once again, the $(\forall, \forall)$ lifting properly mimics a notion of absolute betterness while the milder $(\forall, \exists)$ adds an element of risk and uncertainty, especially in relation with its negation, as with the expressions in Example 10 stating that defection is 'safer' than cooperation.

The example has shown that liftings indeed make perfect sense when comparing choices in an effectivity function and directs us towards the formulation of the notion of Pareto optimal choices (Definition 19), which are nothing but the maxima of the preference order induced by the liftings within a coalitional effectivity function. But first let us look more closely at some general properties of the preference liftings, in relation to the underlying preference order.

Properties of preference liftings

It goes without saying that many properties of the preference liftings are directly inherited from the underlying preference order. In our case this is a total preorder over the outcomes. Yet, most of the above defined liftings are not total preorders.

Proposition 8 *Let $(\geq_i)_{i \in N}$ be a preference order, W a set of alternatives, $Q_1, Q_2 \in \{\forall, \exists\}$ and $\geq_i^{(Q_1,Q_2)}, >_i^{(Q_1,Q_2)}$ the preference relations as given in Definition 17. We have that:*

1. *$\geq_i^{\forall,\exists}$ is a total preorder;*

2. *If $\geq_i^{(Q_1,Q_2)} \neq \geq_i^{\forall,\exists}$ then $\geq_i^{(Q_1,Q_2)}$ is not a total preorder;*

3. *$>_i^{(Q_1,Q_2)}$ is not a total preorder.*

Proof *To prove $\succeq_i^{\forall,\exists}$ is a total preorder it suffices to prove that $\succeq_i^{\forall,\exists}$ is transitive and complete. For transitivity assume $A \succeq_i^{\forall,\exists} B$ and $B \succeq_i^{\forall,\exists} C$. We need to prove that $A \succeq_i^{\forall,\exists} C$. If A is empty the proof is trivial, otherwise take an arbitrary $x \in A$. From the assumptions it follows that there exists $y \in B$ such that $x \succeq_i y$. Again from the assumptions it follows that there exists a $z \in C$ such that $y \succeq_i z$. But $\succeq_i$ is transitive so $x \succeq_i z$. To prove completeness, let us take two arbitrary nonempty sets $A, B \subseteq W$. If $A \subseteq B$ we have by reflexivity of $\succeq_i$ that $A \succeq_i^{\forall,\exists} B$. Suppose instead that $A \not\subseteq B$. So there exists $x \in A$ such that $x \notin B$. Let us consider the set $X = A \setminus B$. Suppose there exists $x \in X$ such that for no $y \in B$ we have that $x \succeq_i y$ (otherwise simply $A \succeq_i^{\forall,\exists} B$). But by completeness of $\succeq_i$ we have that $y \succeq_i x$ for all $y \in B$, that is $B \succeq_i^{\forall,\exists} A$.*

To prove that no other preference relation over sets of the form we have defined is a total preorder we need to find a counterexample for each case. Let us consider the sets $A = B = \{x\}$ for some $x \in W$. It does not hold that $A \succ_i^{(Q_1,Q_2)} B$ nor that $B \succ_i^{(Q_1,Q_2)} A$ for each Q_1, Q_2, which proves that none of these relations is a total preorder. To see that $\succeq_i^{(\exists,\exists)}$ is not a total preorder consider sets $A = \{1,3\}$, $B = \{2,7\}$, $C = \{6,5\}$ that have the natural numbers $\mathbb{N}$ as domain, with the naturally induced preference relation. We have that $A \succeq_i^{(\exists,\exists)} B$, $B \succeq_i^{(\exists,\exists)} C$ and that it is not the case that $A \succeq_i^{(\exists,\exists)} C$ which falsifies transitivity. To see that $\succeq_i^{(\forall,\forall)}$ is not a total preorder consider set $A = \{a, b\}$ with $a \succ_i b$. It is not the case that $A \succeq_i^{(\forall,\forall)} A$, which falsifies completeness. To see that $\succeq_i^{(\exists,\forall)}$ is not a total preorder consider sets $A = \{x \in \mathbb{N} \mid x \text{ is even}\}$ and $B = \{x \in \mathbb{N} \mid x \text{ is odd}\}$, where again the preference relation is induced by the order on the numbers. It is not the case that $A \succeq_i^{(\exists,\forall)} B$ nor that $B \succeq_i^{(\exists,\forall)} A$, which falsifies completeness.

A number of properties have been studied against which to compare a preference lifting. Proposition 8 shows how the structural properties of a preference order are reflected in the types of lifting that we are studying.

3.2.2 Pareto optimal choices

The preference liftings in Definition 17 can be used to order players' choice sets. Mimicking the classical notion of Pareto optimality, used to classify states, we can now introduce Pareto optimality for choices. The intuition is that, given a choice set $\mathcal{X}$ and a preference relation $\succsim$ over subsets of its domain ($\succ$ denoting its strict counterpart), a choice $X \in \mathcal{X}$ is a *Pareto optimal* if there is no choice $Y \in \mathcal{X}$ that dominates X according to $\succsim$. [1]

Definition 18 (Pareto optimal choice) *Let W be a set of alternatives, $C \subseteq N$ a set of players and $E(w)(C)$ a choice set of coalition C at w defined on the domain W. A choice $X \in E(w)(C)$ is* Pareto optimal choice *for coalition C at w if, and only if, for no $Y \in E(w)(C)$, $Y \succsim_i X$ for all $i \in C$.*

[1] As a convention, in order to increase readability, when introducing new concepts involving a quantified preference relation over sets, the relation will be anonymously denoted $\succsim$. Only later, once the notion is explained, further properties coming from the various liftings will be discussed.

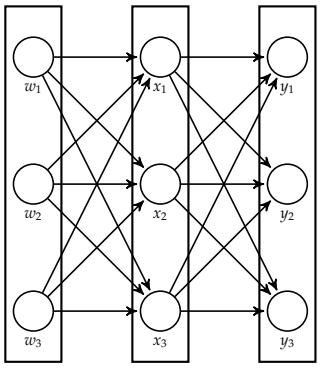

Figure 3.5: A $(\forall, \forall)$ Pareto optimal choice. The arrows indicate the relation $\succeq_i$ for each player i in coalition C at w modulo transitivity and reflexivity, and the sets succinctly represent choices in $E(w)(C)$. The leftmost set is Pareto optimal for C at w, as no better set exists in $E(w)(C)$.

In words, the definition says that a choice is Pareto optimal for a coalition at a certain state if there is no better alternative to that choice for all members of that coalition at that state. The betterness order is indicated by the preference relation over sets, which henceforth will be one of the preference liftings introduced in Definition 17. The choice set in which a Pareto optimal choice for a coalition C is evaluated will be, unless otherwise specified, the effectivity function of coalition C itself, at a particular state. To study the consequences of preference liftings in classifying choices the full-blown definition is needed.

Definition 19 (Quantified Pareto optimal choice) *Let $E(w)(C)$ be a choice set. $X \in E(w)(C)$ is (Q^1, Q^2)*-Pareto optimal choice *for coalition C in w if, and only if, for no $Y \in E(w)(C)$, $Y \succ_i^{(Q^1,Q^2)} X$ for all $i \in C$, and for $Q^1, Q^2 \in \{\exists, \forall\}$.*

As a convention, when the effectivity function is understood, we indicate with $PO_{C,w}(A, \succ_i^{(Q_1,Q_2)})$ the fact that set A is Pareto optimal choice for C in w according to preference relation $\succ_i^{(Q_1,Q_2)}$. When instead the effectivity function is not clear from the context, it will be said that $PO_{C,w}(A, \succ_i^{(Q_1,Q_2)})$ holds *in* a given effectivity function E. The new definition makes the role of preference liftings explicit in classifying choices. If for instance a set X is Pareto optimal choice for coalition C at w then there is no reason (no set Y exists in the choice set of C at w that is better than X according to the given preference lifting) not to choose X. Figure 3.5 and Example 11 illustrate this further.

Example 11 (Pareto optimal choices) *Let us carry on with the prisoner's dilemma, and let us represent it by an effectivity function E and a preference relation $\succeq_i$ for $i \in N = \{Row, Column\}$. As the prisoner's dilemma is a game with a surjective outcome function we have in particular that all Pareto optimal outcomes can be forced by the grand coalition,*

	L	R
U	$0, 2$	$0, 100$
D	$0, 3$	$0, 1$

Decision Problem 1

	L	R
U	$0, 2$	$0, 4$
D	$0, 3$	$0, -100$

Decision Problem 2

Figure 3.6: Two decision problems for the column player, while the row player is indifferent among the possible outcomes. The preference liftings have often unsatisfactory suggestions, ruling out reasonable alternatives such as choosing R in the game on the left or choosing L in the game on the right.

formally $\{(D, L)\}, \{(U, L)\}, \{(U, R)\} \in E(w)(N)$*. But are these outcomes also Pareto optimal choices? The answer is positive, for all preference liftings. Take for instance the* $(\forall, \forall)$ *lifting. Clearly, for each Pareto optimal outcome* x*, it is not possible to find an* $X \in E(w)(N)$ *such that* $X \succ_i^{(\forall,\forall)} \{x\}$*.*

At the level of single player coalitions the situation is a bit more complicated. How to treat the choices D and U by Row? The $(\forall, \forall)$ *lifting treats them both as Pareto optimal choices: the idea is that there is no absolute reason to choose defect or cooperate. The* $(\forall, \exists)$ *lifting instead recognizes the element of risk in the cooperative move and only indicates defection as Pareto optimal. Dually, the* $(\exists, \exists)$ *lifting declares no choice in* $E(w)(\{Row\})$ *Pareto optimal, while the* $(\exists, \forall)$ *would declare defection as Pareto optimal.*

From the last observation in the example one might conclude that $(\exists, \forall)$ and $(\forall, \exists)$ behave better in the prisoner's dilemma than the weak $(\forall, \forall)$ lifting. This is true in fact, though the other three types of liftings (and their corresponding strict counterparts) show other undesirable properties. It is not difficult to think of games where the $(\forall, \exists)$ and $(\exists, \forall)$ lifting do not provide a satisfactory solution.

Example 12 *Let us consider the games in Figure 3.6. In the game on the left we have that* $\{(L, U), (L, D)\} \succ_{Column}^{(\forall,\exists)} \{(R, U), (R, D)\}$*, as* (R, D) *is the worst possible outcome and, for the same reason, not* $\{(R, U), (R, D)\} \succeq_{Column}^{(\forall,\exists)} \{(L, U), (L, D)\}$*. In other words, according to the* $(\forall, \exists)$ *preference lifting there is a reason not to choose R but there is no reason not to choose L. Using Pareto optimality as a suggestion for action, the conclusion is that L is to be preferred to R, which is at least debatable. In the game on the right, instead, is the* $(\exists, \forall)$ *lifting to behave in an undesired way, by suggesting Column the move R, as* (R, D) *is for him the best possible outcome.*

The example shows that no preference lifting, at least among the ones that we have introduced, can always indicate the most reasonable move in a game and for each of them a game can be found where the given answer is not completely satisfactory. The only exception is possibly the $(\forall, \forall)$ type of lifting, which tends to provide too many suggestions, but it never excludes the desirable ones. Henceforth, this extremely weak type of lifting will be adopted by default, as a basis to define more structured concepts that *refine* its suggestions.

Before moving on to introduce undominated choices, that endow Pareto optimal ones with choice restrictions, we focus on some formal properties of the latter.

Properties of Pareto optimal choices

Here we provide formal results that relate the Pareto optimal choices with effectivity functions and strategic games together with the relation between the various forms of Pareto optimal choices associated with the preference liftings.

Pareto optimal choices and effectivity functions Pareto optimal choices are orders defined in effectivity functions, which are outcome monotonic sets of sets. How does outcome monotonicity of the effectivity function influence the optimality of choices? We can prove that the influence of outcome monotonicity depends on the lifting under consideration: while Pareto optimality enjoys monotonicity when considering relations $\succ_i^{(\forall,\forall)}, \succ_i^{(\exists,\forall)}$, it enjoys antimonotonicity when considering relations $\succ_i^{(\exists,\exists)}, \succ_i^{(\forall,\exists)}$.

Proposition 9 *Let W be a set of alternatives, $C \subseteq N$ a set of players, $\mathcal{X}$ a choice set over W, $A, B \in \mathcal{X}$, and $\succ_i^{(Q_1,Q_2)}$ for $Q^1, Q^2 \in \{\exists, \forall\}$ the usual preference relation. We have that:*

1. *$A \subseteq B$ implies that $PO_{C,w}(B, \succ_i^{(\forall,\forall)})$ whenever $PO_{C,w}(A, \succ_i^{(\forall,\forall)})$;*
2. *$A \subseteq B$ implies that $PO_{C,w}(B, \succ_i^{(\exists,\forall)})$ whenever $PO_{C,w}(A, \succ_i^{(\exists,\forall)})$;*
3. *$A \subseteq B$ implies that $PO_{C,w}(A, \succ_i^{(\exists,\exists)})$ whenever $PO_{C,w}(B, \succ_i^{(\exists,\exists)})$;*
4. *$A \subseteq B$ implies that $PO_{C,w}(A, \succ_i^{(\forall,\exists)})$ whenever $PO_{C,w}(B, \succ_i^{(\forall,\exists)})$;*

Proof *The proof is a direct consequence of Definition 19.*

The proposition shows that the $(\forall, \forall)$ and the $(\forall, \exists)$ Pareto optimal choices are monotonic (items 1 and 2), while the $(\exists, \forall)$ and the $(\exists, \exists)$ are antimonotonic (items 3 and 4). These properties are formally desirable but they allow for paradoxical interpretations. If we for instance consider the monotonic forms of Pareto optimal choices we come across a sort of Ross Paradox [47]: if it is optimal to cooperate then it is optimal to cooperate or not to cooperate.[2]. For the other two a reversed Ross paradox is available: if it is optimal to cooperate or to defect then it is optimal to defect. This suggests that the intuitive interpretation of choices should be further disentangled: *choosing X* in a certain effectivity function should be here understood as *choosing a strategy* leading to X. Different definitions of choices are possible, some of which do not enjoy properties such as outcome monotonicity: this will be a topic of discussion of Section 3.3.

[2]The original paradox states that if it is obligatory to send a mail then it is obligatory to send a mail or to burn it [47].

Strong Pareto optimal choices To complete the picture of the various forms of Pareto optimal choice we introduce its strong version, that corresponds to strong Pareto optimality for outcomes (Definition 2).

Definition 20 (Quantified strongly Pareto optimal choice) *Let $Q^1, Q^2 \in \{\exists, \forall\}$ and $E(w)(C)$ be a choice set. $X \in E(w)(C)$ is* (Q^1, Q^2)-Strongly Pareto Optimal *for coalition C in w if, and only if, for no $Y \in E(w)(C)$, $Y \geq_i^{(Q^1,Q^2)} X$ for all $i \in C$; and $Y >_i^{(Q^1,Q^2)} X$ for some $i \in C$.*

A choice is strongly Pareto optimal for some coalition if there is no choice available that is at least as good for all members of that coalition and strictly better for some. Strongly Pareto optimal choices coincide with Pareto optimal choices when considering coalitions made by one player. They significantly differ in the other cases. For instance, for $C = \emptyset$, each $X \in E(w)(C)$ is strongly Pareto optimal choice while none is Pareto optimal choice. As for monotonicity it behaves like the Pareto optimal case (Proposition 9). The following proposition states the expected inclusion relations between weakly and strongly Pareto optimal choices.

Proposition 10 *Let W be a set of alternatives, $C \subseteq N$ a set of players, $\mathcal{X}$ a choice set over W; $A, B \in \mathcal{X}$, and $>_i^{(Q_1,Q_2)}$ for $Q^1, Q^2 \in \{\exists, \forall\}$ the usual preference relation. Let us indicate with $SPO_{C,w}(A, >_i^{(Q_1,Q_2)})$ the fact that set A is Pareto optimal choice for C in w according to preference relation $>_i^{(Q_1,Q_2)}$. We have that:*

- *$SPO_{C,w}(B, >_i^{(Q_1,Q_2)})$ and $C \neq \emptyset$ implies that $PO_{C,w}(B, >_i^{(Q_1,Q_2)})$;*
- *$SPO_{C,w}(A, >_i^{(\exists,\exists)})$ and $C \neq \emptyset$ implies that $PO_{C,w}(A, >_i^{(\forall,\forall)})$.*

Proof *The proof is a direct consequence of Definitions 19 and 20.*

The proposition says that for nonempty coalitions each strongly Pareto optimal choice is a Pareto optimal choice and that each strongly Pareto optimal choice according to the $(\exists, \exists)$ preference lifting is also a Pareto optimal choice according to the $(\forall, \forall)$ one.

Pareto optimality of outcomes and of choices One feature suggested by Example 9 and Example 10 is that Pareto optimal choices reflect weak Pareto optimality of outcomes. The intuition can be made formal by the following result.

Proposition 11 *Let $\mathbb{G}$ be a game with a surjective outcome function and $E^\alpha_{\mathbb{G}}$ the effectivity function representing it at world w. For all $x \in W$ and for $Q_1, Q_2 \in \{\exists, \forall\}$ we have that:*

$$x \text{ is Pareto optimal in } \mathbb{G} \Leftrightarrow PO_{N,w}(\{x\}, >_i^{(Q_1,Q_2)}) \text{ in } E^\alpha_{\mathbb{G}}$$

Proof *From left to right, assume x to be Pareto optimal in $\mathbb{G}$. So for no $y \in W$ do we have that $y >_i x$ for all $i \in N$. By the fact that $\mathbb{G}$ has surjective outcome function we have that $E^\alpha_{\mathbb{G}}(w)(N) = 2^W \setminus \emptyset$. Suppose that $\{x\} \in E^\alpha_{\mathbb{G}}(w)(N)$ is not such that $PO_{N,w}(\{x\}, >_i^{(Q_1,Q_2)})$*

for some $Q_1, Q_2 \in \{\exists, \forall\}$. *Then there exists* $X \in E^{\alpha}_{\mathbb{G}}(w)(N)$ *such that for all (and for some)* $y \in X$, $y \succ_i x$ *for all* $i \in N$. *As* $X \neq \emptyset$ *we have that there exists* $y \in X$, $y \succ_i x$ *for all* $i \in N$. *Contradiction. The other direction is straightforward.*

The proposition shows a mapping between Pareto optimality of choices and Pareto optimality of outcomes in the corresponding game. However it holds only for surjective outcome functions. In this case therefore the effectivity function of N includes all choices of the form $\{x\}$ for x being a possible outcome of the game and establishing Pareto optimality for choices *implies* Pareto optimality for outcomes.

As to the correspondence with strongly Pareto optimal outcomes a similar result is obtained.

Proposition 12 *Let* $\mathbb{G}$ *be a game with a surjective outcome function and* $E^{\alpha}_{\mathbb{G}}$ *the effectivity function representing it at world* w. *For all* $x \in W$ *and for* $Q_1, Q_2 \in \{\exists, \forall\}$ *we have that:*

$$x \text{ is Strongly Pareto Optimal in } \mathbb{G} \Leftrightarrow SPO_{N,w}(\{x\}, \succ_i^{(Q_1,Q_2)}) \text{ in } E^{\alpha}_{\mathbb{G}}$$

Proof *It follows the same procedure as for the weak case.*

Propositions 11 and 12 show that Pareto optimal choices are generalizations of Pareto optimality on outcomes, independently of the preference lifting we might want to consider.

3.2.3 Undominated choices

The contribution of game theory to the analysis of interaction has emphasized that one fundamental aspect of rationality lies in the capability of reasoning about one's opponents. As put in [6, p.14],

> when advising what to do, you must take into account what the other players can do, and the outcome may well be a reasonable compromise.

Pareto optimal choices do not take this stance into account, as they are an order on coalitional choices that only consider the preferences of its members. In this section we define undominated choices, as those choices that remain Pareto optimal for all possible reactions of the opponents. As we have pointed out in the introductory part of this work opponents' possibilities can be modelled looking at how they transform coalitional choices: it is the notion of choice restriction that we can retrieve in strategic reasoning.

Definition 21 (Choice restriction) *Let* E *be an effectivity function, and* $X \in E(w)(\overline{C})$. *The* X-choice restriction *for* C *in* w *is the set* $E(w)(C) \sqcap X$.

Given a choice set $E(w)(C)$ its choice restriction $E(w)(C) \sqcap X$ is given by the intersection of each set in $E(w)(C)$ with X. The idea, illustrated in Figure 3.7 and Example 13, is that each possible choice of C is now *restricted* by the choice X by $\overline{C}$.

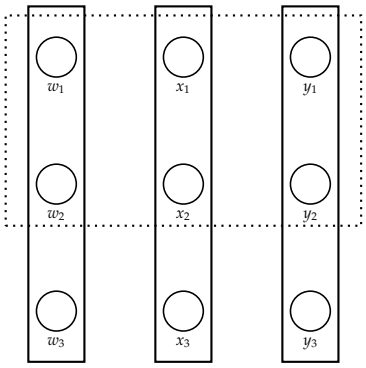

Figure 3.7: A choice restriction. The dotted rectangle, symbolizing a choice of coalition C', *restricts* the choice set of $\overline{C'}$, succinctly represented by the straight rectangles.

Example 13 *Let us take the usual example of the prisoner's dilemma. It has been previously observed that given the choice L by Column, i.e. the set* $\{(U,L),(D,L)\} \in E(w)(\{Column\})$, *the choice D, i.e. the set* $\{(D,L),(D,R)\} \in E(w)(\{Row\})$, *becomes preferable to the choice L. With the notion of choice restriction we can compute the effect of the choice L on the strategic possibilities of the row player, i.e.* $E(w)(\{Row\}) \sqcap \{(U,L),(D,L)\} = \{\{(U,L)\},\{(D,L)\}\}$. *The choices available to Row now share sets in L: this is precisely the idea of restriction of one's moves by the opponents, that cuts the possible available outcomes for a coalition.*

Combining the notions of choice restriction and Pareto optimal choice, undominated choices are immediate to define. For present purposes it is convenient to focus on only one type of Pareto optimal choice, namely the one making use of the $(\forall,\forall)$ preference lifting.

Definition 22 (Undomination) *Let* $E(w)(C)$ *be a choice set.* $X \in E(w)(C)$ *is said to be* undominated *for C at w (abbreviated* $X \rhd_{C,w}$*) if, and only if, for all* $Y \in E(w)(\overline{C})$, $(X \cap Y)$ *is a* $(\forall,\forall)$ *Pareto optimal choice in* $E(w)(C) \sqcap Y$ *for C at w.*

In words, a choice is undominated for coalition C at w if it is $(\forall,\forall)$ Pareto optimal choice in all choice restrictions induced by the opponents'possible reactions. This definition suggests that coalitional rationality consists of two dimensions: an inward Pareto-like reasoning, aiming at choosing the best among the available choices; and an outward strategic reasoning, taking the possible moves of the opponents into account. Both dimensions are represented in undominated choice, that merge Pareto optimal choices with choice restrictions, as exemplified in Figure 3.8. Hereby undomination clearly resembles the notion of dominant strategy of Definition 3. Later on, we will be able to turn this clear resemblance into a formal connection.

Example 14 *Continuing our example, we have that playing D, i.e.* $\{(D,L),(D,R)\}$, *is undominated at w for Row and so is playing R, i.e.* $\{(D,R),(U,R)\}$, *for Column, as defecting*

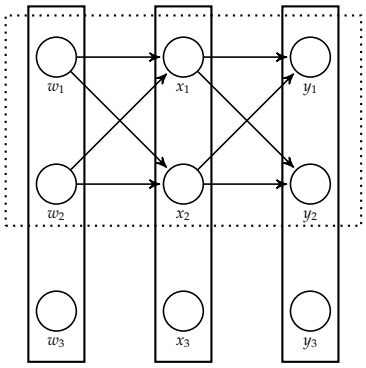

Figure 3.8: The structure of undominated choices. Pareto optimality is calculated in the choice restrictions induced by the opponents possible choices. The leftmost set is $(\forall, \forall)$ Pareto optimal in the choice restriction induced by the dotted rectangle.

in the prisoner's dilemma remains optimal whatever the opponent decides to do. However it is not the case that playing down and right, i.e. $\{(D, R)\}$ *at w for the coalition made by Row and Column, as there is a set, namely* $\{(U, L)\}$*, that dominates* $\{(D, R)\}$ *in* $E(w)(N) \sqcap W$*. In words, the notion of undomination shows formally that indeed the situation in which both players defect is individually rational but not in the interest of both players taken together.*

Undomination is the appropriate concept to refine the $(\forall, \forall)$ preference lifting, as can be seen from all games of Figure 2.2. Full convergence game, for instance, is an example of failure of $(\forall, \forall)$ Pareto optimality to indicate a reasonable choice, as both the sets corresponding to L and R are Pareto optimal. With undomination the matter clears, as the set corresponding to R is dominated, while the one corresponding to L is undominated. This can be seen as an example of how undomination, in its strongest form that uses the $(\forall, \forall)$ preference lifting, implements what Horty calls *the sure-thing principle* [41]: if an action K is to be preferred to an action K', $K \cap X$ is to be preferred to $K' \cap X$ exhausting all possibilities X. In strategic interaction, the space of possibilities is played by the opponents' possible moves.

Properties of undomination

Let us now concentrate on the structural properties of undomination, to be used later on. First we focus on choice restrictions, that as we expect from their definition, radically modify coalitional effectivity functions though preserving several important properties.

Proposition 13 *Let* $E(w)(C)$ *be a choice set and* $X \in E(w)(\overline{C})$*. The following holds*

1. $E(w)(C) \sqcap X = (E(w)(C) \sqcap X) \sqcap X$*;*
2. $E(w)(C) \sqcap X = E(w)(C)$ *whenever* $X = W$*;*

3. $\emptyset \notin E(w)(C) \sqcap X$ *whenever* E *is regular ;*

4. $X \in E(w)(C) \sqcap X$ *whenever* E *contains the unit ;*

5. $E(w)(\emptyset) \sqcap X = \{X\}$ *whenever* E *is determined ;*

Proof *The first and second item follow from the properties of the intersection and Definition 21. For the third item let us reason by contraposition and suppose* $\emptyset \in E(w)(C) \sqcap X$*. This means that either* $\emptyset \in E(w)(C)$ *or that* $X \cap Y = \emptyset$ *for some* $Y \in E(w)(C)$*. In either case* E *is not regular. For the fourth item suppose* $W \in E(w)(C)$*. But this means that there is a* $Y \in E(w)(C)$ *such that* $Y \cap X = X$*. The same reasoning applies to the fifth item.*

Proposition 13 shows that an effectivity function restricted by a set cannot be restricted further by the same set (item 1), and it is not modified when restricted by the set $\{W\}$ (item 2). Properties of effectivity functions also influence choice restrictions. Item 3 shoes that the empty set does not belong to any choice restriction of any regular effectivity function, while item 4 shows that a restricting set X is carried along to the choice restriction $E(w)(C) \sqcap X$ in every effectivity function containing the unit. Finally item 5 shows a particular property of $E(w)(\emptyset)$ in determined effectivity functions: restricting it with any set X is equivalent to obtain a choice restriction of the form $\{X\}$.

Undomination and games The relation between choice restrictions and subgames, that has been appreciated in the examples, can be formally drawn. Its formal statement will be used later on to connect undominated choices with dominant strategies in games.

Proposition 14 *Let* $\mathbb{G}$ *a strategic game and* E *the effectivity function representing it at world* w*. Let* $\sigma_{\overline{C}}$ *be a strategy of coalition* $\overline{C}$ *in* $\mathbb{G}$ *and* X *be such that* $X = \{x \mid o(\sigma_{\overline{C}}, \sigma_C) = x$ *for some* σ_C *available in* $\mathbb{G}\}$*. We have that*

$$E(w)(C) \sqcap X = E^{\mathcal{G}\downarrow\sigma_{\overline{C}}}(C)$$

Proof *Let us first show that* $E(w)(C) \sqcap X \subseteq E^{\mathcal{G}\downarrow\sigma_C}(C)$*. Assume* $Y \in E(w)(C) \sqcap X$*, for* E *an effectivity function representing* $\mathbb{G}$ *and let* X *be such that* $X = \{x \mid o(\sigma_{\overline{C}}, \sigma_C) = x$ *for some* $\sigma_C\}$ *for* $\sigma_C, \sigma_{\overline{C}}$ *being coalitional strategies in game* $\mathbb{G}$*. From the assumptions and the definition of* $E^{\alpha}_{\mathbb{G}}$ *(Definition 11) we can derive that* $X \in E(w)(\overline{C})$*. By Definition 21 there exists a set* $Z \in E(w)(C)$ *such that* $Z \cap X = Y$*. As* Z *belongs to an effectivity function representing* $\mathbb{G}$ *there exists a strategy* ρ_C *such that for all strategies* $\rho_{\overline{C}}$ $o(\rho_C, \rho_{\overline{C}}) \in Z$*. It follows that the strategy* $(\rho_C, \sigma_{\overline{C}})$ *is such that* $o(\rho_C, \sigma_{\overline{C}}) \in Z \cap X = Y$*. But by Definitions 5 and 13* $(\rho_C, \sigma_{\overline{C}})$ *is an available strategy in* $\mathcal{G} \downarrow \sigma_{\overline{C}}$ *for coalition* C*. So we have that* $Y \in E^{\mathcal{G}\downarrow\sigma_{\overline{C}}}(C)$*. For the reverse direction, assume* $Y \in E^{\mathcal{G}\downarrow\sigma_{\overline{C}}}(C)$*. This means, by Definitions 5 and 13, that there is a strategy* ρ_C *in* $\mathcal{G} \downarrow \sigma_{\overline{C}}$ *such that* $o(\rho_C, \sigma_{\overline{C}}) = \{w\} \subseteq Y$*. But as again by Definition 5* ρ_C *is also a strategy of coalition* C *in* $\mathbb{G}$*, let us consider* $Z = \{x \mid o(\rho_C, \rho_{\overline{C}})$ *for some* $\rho_{\overline{C}}\}$*. We have that* $Z \in E(w)(C)$ *and by the fact that* $Z \cap X = \{w\}$ *we also have that* $Y \in E(w)(C) \sqcap X$*.*

In a nutshell Proposition 14 states that, with effectivity functions representing strategic games, choice restrictions behave as subgames, completing the picture sketched in Figure 2.5.

Using this result we are also able to show how, for effectivity functions representing strategic games, undominated choices can be seen as dominant strategies in disguise.

Proposition 15 *Let $\mathbb{G}$ be a strategic game and E the effectivity function representing it at world w. Let σ_{-i} be a strategy of coalition $\overline{\{i\}}$ in $\mathbb{G}$ as in Definition 1. Let X be such that $X = \{x \mid o(\sigma_i, \sigma_{\bar{i}}) = x$ for some σ_{-i} available in $\mathbb{G}\}$.*

We have that

$$X \rhd_{\{i\},w} \text{ if and only if } \sigma_i \text{ is a dominant strategy for } i \text{ in } \mathbb{G}$$

Proof *($\Rightarrow$) Recall first that by Proposition 14 $Y \in E^{\mathbb{G}\downarrow\sigma_i}$ if and only if $Y \in X \sqcap E(w)(\{i\})$ for E being the effectivity function representing $\mathbb{G}$.*

Suppose σ_i is not a dominant strategy for i in $\mathbb{G}$. This is equivalent to saying that there exists σ_i' such that for some σ_{-i} we have that $o(\sigma_i', \sigma_{-i}) \succ_i o(\sigma)$. Consider now the set $X' = \{x' \mid o(\sigma_i', \sigma_{-i}) = x$ for some σ_{-i} available in $\mathbb{G}\}$ representing σ_i' and the set $Z = \{z \mid o(\rho_i, \sigma_{-i}) = z$ for some ρ_i available in $\mathbb{G}\}$ representing σ_{-i}. From the properties of E we must have that $X' \in E(w)(\{i\})$ and $Z \in E(w)(\overline{\{i\}})$. Consider now the sets $X \cap Z$ and $X' \cap Z$. We must have that $X' \cap Z \geq_i^{(\forall,\forall)} X \cap Z$, which shows that X is dominated.

($\Leftarrow$) Follows the same pattern of the previous direction.

The proposition shows that, in effectivity functions representing the choices of individual players in games, undominated strategies represent dominant strategies. The result lifts the notion of rationality available for strategic games to a more general coalitional version. The next section will bring this view further by analyzing the peculiar features of coalitional ability in games.

3.3 Coalitional Games and Strategic Games

This section is devoted to discussing the relation between coalitional games and strategic games. In Chapter 2 we have noticed how effectivity functions are rich enough to express coalitional power in strategic interaction, by defining a coalitional effectivity function of a strategic game, also called α-effectivity function (Definition 11). A natural question is whether it is possible to isolate the class of effectivity functions that precisely represent strategic games, i.e. the properties that an effectivity function needs to satisfy in order to be the α-effectivity function of some strategic game. Subsection 3.3.1 addresses this point. A second issue we will deal with, also related to the connection between undominated choices and dominant strategies, concerns the interpretation of an effectivity function in a strategic game and its flexibility in representing strategy execution. Subsection 3.3.2 is devoted to this issue.

3.3.1 Representation theorems

The representation theorem given in [54][Theorem 2.27], known as Pauly's Representation Theorem, states that an effectivity function is playable if and only if it corresponds to a strategic game. It is a generalization of already existing correspondence results in [49, 56] for strategic games with arbitrary outcome functions.

Specifically, the correspondence (called α-correspondence [54]) is formulated in two directions:

- every playable effectivity function is the α-effectivity function of some strategic game,
- each game has an α-effectivity function that is playable.

The proof of the latter claim was already recalled in Chapter 2 (Proposition 2). But the former turns out not to be correct.

Before showing this, it is instructive to recall what coalitional strategies are, following [54, p.16]:

> For notational convenience, let $\sigma_C := (\sigma_i)_{i \in C}$ denote the strategy tuple for coalition $C \subseteq N$ which consists of player i choosing strategy $\sigma_i \in \Sigma_i$.

From the interpretation given, that is also consistent with its use [54], we have that $\Sigma_\emptyset = \{\emptyset\}$, i.e. the set of strategies of coalition $\emptyset$ only contains an empty strategy. [3] The α-effectivity function of the empty coalition in a game $\mathbb{G}$ reduces then to the following:

$$X \in E^\alpha_{\mathbb{G}}(\emptyset) \Leftrightarrow \exists \sigma_\emptyset \forall \sigma_N o(\sigma_\emptyset, \sigma_N) \in X \Leftrightarrow \forall \sigma_N o(\sigma_N) \in X$$

In words sets in the α-effectivity function of coalition $\emptyset$ are supersets of the set of the possible outcomes reachable by the grand coalition. In a nutshell, the coalitional power of the empty coalition cannot hinder players from reaching a certain outcome.

For the sake of precision, the types of structures that are in [54] called *games* are usually referred to as *game forms*, because they do not consist of preference relations. As preference relations do not affect the correspondence and denoting game forms makes the notation more heavy we stick to the formulation given in [54]: we will talk of games and not of game forms and use the notation $E^\alpha_{\mathbb{G}}$ instead of the more precise but heavier $E^\alpha_{\mathbb{F}_s}$.

A Counterexample to Pauly's Representation Theorem

We can now show a counterexample to Pauly's Representation Theorem, obtained by constructing an effectivity function that is playable but cannot correspond to any strategic game.

[3] As coalitional strategies are treated as functions from sets of players to the tuples of their individual strategies, the empty strategy boils down to an empty function.

Proposition 16 *There is a playable effectivity function E for which $E \neq E^{\alpha}_{\mathbb{G}}$ for all strategic games $\mathbb{G}$.*

Proof *Consider a coalitional game frame with a single player 'a' that has the set of natural numbers $\mathbb{N}$ as the domain (i.e., $N = \{a\}, W = \mathbb{N}$), and the effectivity defined as follows:*

- $E(\{a\}) = \{X \subseteq \mathbb{N} \mid X \text{ is infinite}\}$;
- $E(\emptyset) = \{X \subseteq \mathbb{N} \mid \overline{X} \text{ is finite}\}$.

In other words, the grand coalition $\{a\}$ is effective for all infinite subsets of the natural numbers, while the empty coalition can enforce all its cofinite subsets.

E is playable and it does not correspond to any strategic game. To see this let us first verify the playability conditions. Outcome monotonicity, N-maximality, liveness and safety are straightforward to check. For superadditivity, notice that we only have two cases to verify:

1. $C = \{a\}, C' = \emptyset$;
2. $C = \emptyset, C' = \emptyset$.

For the first case, consider a set $X \in E(\{a\})$ and a set $Y \in E(\emptyset)$. To show that $X \cap Y \in E(\{a\} \cup \emptyset) = E(\{a\})$ we only need to observe that $X = (X \cap Y) \cup (X \cap \overline{Y})$. As $X \cap \overline{Y}$ is a finite set and Y cofinite, we must have that $X \cap Y$ is infinite, so $X \cap Y \in E(\{a\})$. For the second case it is sufficient to recall that the intersection of two cofinite sets is cofinite.

On the other hand, $E^{nc}(\emptyset) = \emptyset$ because there are no minimal cofinite sets. This implies, by Proposition 4, that $E \neq E^{\alpha}_{\mathbb{G}}$ for all strategic games $\mathbb{G}$.

The counterexample constructs a playable effectivity function that assigns no minimal set to the empty coalition. Using Proposition 4, which states that α-effectivity functions have a minimal set, we are able to conclude that there are playable effectivity functions that do not correspond to any strategic games.

Given this fact, it is to be expected that the rather technical argument provided in [54] fails at some point. As its proof will be readapted for an alternative characterization result, it is useful to have a look at it. However, due to its technical character, we leave its discussion to the appendix (Section A.1).

The consequences We have observed that playability conditions are not enough to characterize strategic games. This raises some relevant issues for studying game models and logics for reasoning about games:

1. What are the "truly playable" effectivity functions, i.e. the class of effectivity functions that really correspond to strategic games? How can we characterize these functions in an abstract way? This issue is discussed in Section 3.3.1.
2. Finally, what is the impact on logics for strategic ability, Coalition Logic in particular? Are the axiomatizations from [54] and [32] sound and complete for truly playable models? What logical constructs are needed to distinguish between playable and truly playable structures? These questions, which have a logical nature, are left for the next chapter.

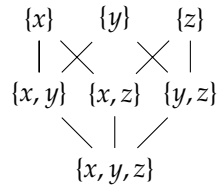

Figure 3.9: A crown

Truly Playable Effectivity Functions

The set of playable effectivity functions that α-correspond to strategic games can be characterized making use of the additional notion of *crown*.

Definition 23 *An effectivity function* $E : 2^N \to 2^{2^W}$ *is a* crown *if and only if* $X \in E(N)$ *implies that* $\{x\} \in E(N)$ *for some* $x \in X$.

Intuitively, an effectivity function is a crown if the set of all players has complete control over the outcome of the game, i.e., every choice of the players includes at least one state that the grand coalition can enforce precisely. Formally, this means that N can only force some singleton sets and all their supersets. By forming an anti-chain of singletons and drawing the cones we obtain a 'crown' as in Figure 3.9, hence the term.

Definition 24 *An effectivity function E is called* truly playable *if it is playable and is a crown.*

Several meaningful characterizations of truly playable effectivity functions are available.

Proposition 17 *The following are equivalent for every playable effectivity function* $E : 2^N \to 2^{2^W}$.

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

Proof

(1) $\Rightarrow$ *(2): immediate, by safety.*

(2) $\Rightarrow$ *(3): Let* $Z \in E^{nc}(\emptyset)$ *and let* $X \in E(\emptyset)$. *Then, by superadditivity,* $Z \cap X \in E(\emptyset)$, *and* $Z \cap X \subseteq Z$, *hence* $Z \cap X = Z$ *by definition of* $E^{nc}(\emptyset)$. *Thus,* $Z \subseteq X$. *Therefore,* $E(\emptyset)$ *is the principal filter generated by* Z, *hence* $E^{nc}(\emptyset) = \{Z\}$.

(3) $\Rightarrow$ *(1): immediate from the definitions.*

(3) ⇒ (4): Let $E^{nc}(\emptyset) = \{Z\}$ and suppose $\{x\} \notin E(N)$ for all $x \in X$ for some $X \subseteq W$. Then, by N-maximality, $S \setminus \{x\} \in E(\emptyset)$, i.e. $Z \subseteq S \setminus \{x\}$ for every $x \in X$. Then $Z \subseteq S \setminus X$, hence $S \setminus X \in E(\emptyset)$. Therefore, $X \notin E(N)$ by superadditivity and liveness. By contraposition, E is a crown.

(4) ⇒ (3): Let $Z = \{z \mid \{z\} \in E(N)\}$ and let $X \in E(\emptyset)$. Take any $z \in Z$, which is nonempty by true playability. By superadditivity we obtain that $\{z\} \cap X \in E(N)$, hence $z \in X$ by liveness. Thus, $Z \subseteq X$. Moreover, $Z \in E(\emptyset)$, for else $S \setminus Z \in E(N)$ by N-maximality, hence $\{x\} \in E(N)$ for some $x \in W \setminus Z$, which contradicts the definition of Z. Therefore, $E(\emptyset)$ is the principal filter generated by Z, hence $E^{nc}(\emptyset) = \{Z\}$.

We also observe that on finite domains playability and true playability coincide.

Proposition 18 *Every playable effectivity function $E : 2^N \to 2^{2^W}$ on a finite domain W is also truly playable.*

Proof *Straightforward, by Proposition 17.3 and the fact that every filter on a finite set is principal.*

Truly playable effectivity functions correspond to strategic game forms

The proof of Theorem 2.27 from [54] fails when we consider the effectivity function of the empty coalition (and, dually, of the grand coalition). However the proof *is* correct for the other cases. It is possible to show that the additional condition of true playability yields correctness of the original construction from [54].

Theorem 19 *A coalitional effectivity function E α-corresponds to a strategic game if and only if E is truly playable.*

The proof of this fact is to found in the appendix (Section A.2).

Theorem 19 provides a general characterization of coalitional games that represent strategic games. It shows that adding certain properties to coalitional effectivity functions, that is a model of cooperative interaction, boils down to describing a strategic game, that is a model of non-cooperative interaction. The theory of rationality, elaborated in the previous section for the abstract cooperative case, still holds for strategic games, that enjoy however extra properties, due to the type of coalitional ability that they describe.

Characterizing games with *citizen sovereignty* In coalitional games the grand coalition may be effective for all outcomes of the game. This case has often been dealt with by several works in the field of social choice theory [50], and it is referred to as *non-imposedness* or *citizen sovereignty*, as it allows players to choose freely among all possible alternatives in a decision process.

However as observed, even in coalitional games representing strategic games citizen sovereignty may not be realized. Games with citizen sovereignty are, because of the properties of playability, simply determined games and they can now be fully characterized.

Corollary 20 *A coalitional effectivity function E α-corresponds to a strategic game with a surjective outcome function if and only if E is determined.*

Proof *Follows from Proposition 2 and Theorem 19.*

Corollary 20 extends Theorem 19 to treat games with surjective outcome function, showing once again the role of the empty coalition in differentiating among various types of strategic games.

3.3.2 On coalitional choices in games

The informal reading of an effectivity function is that of a family of set of sets $\mathcal{X}$ representing the choices assigned to a coalition. However, as previously observed, this reading is ambiguous and *choosing* a set X should in fact be understood as *choosing a strategy* leading to X. In this section we analyze a different meaning of *choices*, that do not enjoy properties like outcome monotonicity.

To do this we consider a family of sets of sets for each coalition, to be understood as the *moves* that a coalition can make. This intuitive formulation can be made clear by the following formal definition, that will be referred to as the coalitional *move* function of a strategic game.

Definition 25 (Move Function) *Let $\mathbb{G} = (N, W, \Sigma_i, \succeq_i, o)$ be a game. Its coalitional move function $H^{\mathbb{G}} : 2^N \to 2^{2^W}$ is given as follows:*

$$X \in H^{\mathbb{G}}(C) \Leftrightarrow \exists \sigma_C \text{ such that } \{v \mid o(\sigma_C, \sigma_{\overline{C}}) = v \text{ for some } \sigma_{\overline{C}}\} = X.$$

Every set in the move function clearly corresponds to a strategy of the original strategic game, as it collects for each coalitional strategy all the *completions* of that strategy that can be carried out by the opponents. The following is the way a coalitional game can be obtained from a strategic game, making use of move functions.

Definition 26 (Coalitional games with move function from strategic game) *Let $\mathbb{G} = (N, W, \Sigma_i, \succeq_i, o)$ be a game. The coalitional game $\mathbb{C}^{\mathbb{G}} = (N, W, H^{\mathbb{G}}, \succeq_i)$ of $H^{\mathbb{G}}$ is a move function.*

Structures with a move function are richer than structures with an α-effectivity function. Their richness can be made precise in a formal way: all α-effectivity functions can be move functions but the converse does not hold.

Move functions have several properties that resemble α-effectivity functions.

Proposition 21 *Let $H^{\mathbb{G}}$ be the move function of $\mathbb{G}$. The following hold:*

1. *$H^{\mathbb{G}}$ is regular, superadditive, nonempty for each C;*
2. *$H^{\mathbb{G},nc}(N) = H^{\mathbb{G}}(N)$;*
3. *$X \in H^{\mathbb{G}}(\emptyset)$ if and only if $\{x\} \in H^{\mathbb{G}}(N)$ for all $x \in X$;*

4. $H^{\mathbb{G}}(C)$ *has IOEC whenever* $\mathbb{G}$ *is surjective.*

Proof *It is an easy check of the definitions.*

The proposition shows that move functions show a number of desirable properties such as regularity and superadditivity (item 1), are closely related to their nonmonotonic core (item 2) and display a duality between the empty and the grand coalition (item 3). Finally they behave as α-effectivity functions when it comes to games with surjective outcome function (item 4). Notice that, unlike the case of α-effectivity functions, $H^{\mathbb{G}}(C)$ need not be outcome monotonic.

Given the resemblance between move functions and α-effectivity functions a question suggests itself: can effectivity functions correspond to move functions and in turn characterize strategic games in the same way truly playable effectivity functions did? The answer is negative, because move functions are not monotonic, so no effectivity function can characterize them. But can we obtain a correspondence dropping monotonicity? Said otherwise, what is the class of sets of sets that correspond to move functions? We tackle this problem by introducing the notion of move set.

Definition 27 (Move sets) *Let* $\mathcal{X}_{C\subseteq N}$ *be a family of set of sets for each* $C \subseteq N$. $\mathcal{X}_{C\subseteq N}$ *is a* move set *if it superadditive, and for each disjoint coalition* $C, \overline{C}$ *we have that* $X \in \mathcal{X}_C$ *and* $Y \in \mathcal{X}_{\overline{C}}$ *we have that* $|X \cap Y| = 1$.

Move sets assign to each coalition choices in such a way that: bigger coalitions have bigger power (superadditivity condition) and choices from disjoint coalitions always amount to a single outcome (disjoint coalitions condition).

The following correspondence is conjectured, and would establish a correspondence between move sets and strategic games, in the same way we have for effectivity functions.

Conjecture 22 *A set of sets* $\mathcal{X}_{C\subseteq N}$ *for each* $C \subseteq N$ *is a move set if and only if there exists a strategic game* $\mathbb{G}$ *such that* $\mathcal{X}_{C\subseteq N} = H^{\mathbb{G}}$.

Move sets based models are very close to Kooi and Tamminga consequentialist models [44], a simplification of STIT models [8] where each coalition is associated with a partition of the domain, discussed in the section on related work.

3.4 Discussion

This section puts together the achievements obtained in this chapter, discussing the issues left open and the related work.

3.4.1 Related work

The work presented in Section 3.2 takes inspiration from Horty's seminal contribution Agency and Deontic Logic [41], where a model on coalitional rationality

is proposed, based on STIT models [8], a branching-time account of coalitional ability endowed with classical utility functions [70]. A broad discussion of the history-based models used by Horty would take this work far from the treatment of strategic interaction, which is why we resort to the discussion of the simpler *consequentialist* models, that share with Horty's models the local features that are necessary to treat one shot interactions. Consequentialist models already have been used as one-shot STIT counterpart by Tamminga and Kooi [44], who also present a model of coalitional rationality with classical utility function, that has much in common with our account.

Definition 28 (Consequentialist models) *[44]*
A consequentialist model *is a pair* (Γ, V) *where* Γ *us a choice structure and* V *is a valuation function on a countable set of propositions Prop.*

The choice structure is nothing but a description of a cooperative game frame where effectivity functions are replaced by partitions in the following way:

Definition 29 (Choice Structures) *[44]*
A choice structure *is a triple* $(W, N, Choice)$ *where* W *is a set of outcomes,* N *a finite set of players and* $Choice : 2^N \to 2^{2^W}$ *a function with the following constraints:*

- *for each* $i \in N$, $Choice(\{i\})$ *is a partition on* W*;*
- *for each* $i \in N$, *let* W *be the set of functions* s *such that* $s(i) \in Choice(\{i\})$, *for* $s(i) \in W$. *We have that for* $C \subseteq N$, $\bigcap_{i \in C} s(i) \neq \emptyset$, *i.e. the pairwise intersection of players' choices is nonempty.*
- *for* $C \subseteq N$, $Choice(C) = \{\bigcap_{i \in C} s(i) : s \in W\}$, *i.e. coalitional choices are constructed by taking the pairwise intersection of individual choices.*

Consequentialist models clearly resemble strategic games, for the way coalitional choices are constructed, though a correspondence is not yet known. A resemblance can be also observed if we look at coalitional games with move function, provided in Definition 26.

As for the notion of coalitional rationality, both in [41] and [44] a utility function is used that associates to each outcome (histories in Horty's framework) an element of an closed interval in the reals (positive reals in Horty's framework, the interval $[-5, 5]$ in Kooi and Tamminga's framework). Setting aside the fact that Kooi and Tamminga evaluate coalitional choices in the interest of other coalitions, a feature that will be dealt with later on in the chapter, for the base case both frameworks share equivalent notions of dominance:

Definition 30 (Dominance) *[44]*
Let $K, K' \in Choice(C)$ *and* $u : N \to W \to [-5, 5]$ *an utility function over the outcomes for each player. We say that* K *dominates* K' *if and only if for all* $S \in Choice(\overline{C})$ *we have that* $w \in K \cap S$ *and* $w' \in K' \cap S$ *implies that* $u_C(w) \geq u_C(w')$, *where* u_C *returns the average of individual utilities of members of* C.

This notion of dominance is very close to that of domination too, provided in Definition 22, which is however of a more general nature: we admit coalitional structures that do not enjoy properties like superadditivity or regularity, a plurality of preference liftings and we can relate it formally to dominant strategies in strategic games (Proposition 15), while the correspondence between strategic games and consequentialist models is still an open problem.

Section 3.3 deals instead with coalitional games. Here too, other works in social choice and game theory have been concerned with characterizing the class of effectivity functions corresponding to strategic games, such as [49]. However, as also observed in [54], Pauly's was the first attempt to characterize strategic games with arbitrary outcome function. In game theory textbooks, such as [51], it is often the case that outcomes are the same as strategic profiles (intending thus a bijective outcome function), and these are the structure taken as starting point in the correspondence results by [49].

3.4.2 Open issues

A number of interesting questions have been left unanswered, both in Section 3.2 and in Section 3.3. As to the former, they mostly concern the relation of undominated choices and the classical solution concepts of strategic games, such as Nash equilibrium and dominant strategy equilibrium. It has been shown by Proposition 11 and Proposition 12 that weak and strong Pareto optimality can be represented via the $(\forall, \forall)$ preference lifting and by Proposition 14 that subgames can be represented by choice restrictions. Even though the formulation of Nash equilibria and dominant strategy equilibria makes use of subgames it is not straightforward to claim that a choice is undominated if and only if it *represents* a best response or a dominant strategy, even if it is undominated in an effectivity function representing a strategic game.

To see why this is the case, let us consider an α-effectivity function of a strategic game $\mathbb{G}$, $E^{\alpha}_{\mathbb{G}}$ and let us call a set X a move of C at w if and only if it is the result of applying some strategy of C. Formally what it is meant is the following:

Definition 31 (Moves) *Let $E^{\alpha}_{\mathbb{G}}$ be an α-effectivity function of a strategic game $\mathbb{G}$. A set $X \in E^{\alpha}_{\mathbb{G}}(C)$ is called a* move *of coalition C at w if and only if there exists a coalitional strategy σ_C such that $X = \{v \mid o(\sigma_C, \sigma_{\overline{C}}) \text{ for some } \sigma_{\overline{C}}\}$*

Moves are the basic components that have been used to introduce move functions in Definition 26, and it is easy to see that every move of a coalition belongs to its coalitional effectivity function. However an effectivity function, being closed under superset, may be constituted by sets that are not moves. Hereby saying that a certain set X is undominated does not mean that X corresponds to a dominant strategy as X may not be a move.

This brings us to Section 3.3 where the most important open problem is stated in Conjecture 22 and concerns the question whether coalitional games with move functions characterize strategic games. This is conjectured, given Theorem 19 and the observations made above on dominant strategies, but a formal proof is still lacking.

It is also not known what the conditions are for move sets to correspond to partitions of the domain, in short, when their models can be turned into consequentialist models.

3.4.3 Conclusion

In this chapter a model of coalitional rationality in strategic interaction has been proposed. It stems from the abstract representation of coalitional power given by effectivity functions and it empowers it with preference relations, which allows lifting along with preferences, also the classical notions of optimality, as studied in Section 3.2. The relation of effectivity functions with strategic games, together with their interpretation of coalitional choice have been studied in Section 3.3.

Concretely in Section 3.2 the following achievements have been realized.

- Lifting of a preference order over alternatives to sets of sets of alternatives, in order to match it with the effectivity function representation, also given as a set of sets. This operation has been carried out in eight different ways, pairwise comparing the elements of the two sets (with four different alternations of the existential and universal quantifier) according to a strict and a weak preference relation. Structural properties, such as preservation of reflexivity, transitivity and completeness, have been stated in Proposition 8.

- Definition of a betterness order within an effectivity function, identifying the optimal sets. These have been called Pareto optimal choices, to mimic the classical notion defined over outcomes. A strong version of Pareto optimal choice has also been defined and both these notions have been formally related to the corresponding ones in strategic games, via Propositions 11 and 12.

- Definition of the notion of subgame, i.e. the restriction on a coalitional effectivity function induced by a possible move of the opponent, and its relating via Proposition 5 to the notion of subgame for strategic games defined in Chapter 2. Using subgames and Pareto optimality a notion of undominated choice has been defined, as a choice the remains Pareto optimal in all possible subgames.

In Section 3.3 the following achievements have been realized:

- Correction of a well-established result relating strategic games and playable effectivity functions, known as Pauly's Representation Theorem, proved in [54]. A counterexample has first been found, showing that in infinite game models there are playable effectivity functions for which no corresponding strategic game can ever be constructed (Proposition 16).

- Analysis of the original proof in [54], identifying where the wrong steps had been taken;

- Proof of correspondence between strategic games and the so-called *truly playable* class (Theorem 19);

- Discussion of the interpretation of effectivity function as coalitional choice in games, comparing it against examples taken from the literature on cooperative games. A different view of coalitional choice, closer to the consequentialist-STIT view from [8] has been proposed and discussed.

All in all, the chapter has laid the structural foundations for an exploration of the logical features of coalitional rationality, to be carried out in the coming chapter.

Chapter 4

Strategic Reasoning in Coalitional Games

It's not that assumptions don't count, but that they come after the conclusions; they are justified by the conclusions. The process goes this way: Suppose you have a set of assumptions, which logically imply certain conclusions. One way to go is to argue about the innate plausibility of the assumptions; then if you decide that the assumptions sound right, then logically you must conclude that the conclusions are right. That's the way that I reject, that's bad science.

Robert J. Aumann, *On the state of the art in game theory* [5]

4.1 Introduction

The characteristic feature of coalitional rationality, studied in the previous chapter, is the capacity to order the members' available choices considering the opponents' possibilities. The present chapter investigates the *logical structure* of this type of reasoning, describing its features within a simple mathematical language.

Vestiges of such a structure can already be found in the reasoning of individuals in the prisoner's dilemma:

1. If my opponent defects, I had better defect;
2. If my opponent cooperates, I had better defect;
3. In conclusion, I had better defect.

What is more, in coalitional reasoning decisions are obtained by merging members' choices and preferences. To say it with a slogan, the word "I", typical of strategic games, is replaced by the word "we", typical of cooperative games.

At present though, the languages to talk about coalitions, such as Coalition Logic [54] — but similar remarks hold for related logics such as Seeing To It That [8] and Alternating-time Temporal Logic [2] — do not explicitly represent preferences and only allow reasoning about what a coalition of players can achieve independently

of the moves of the other players [41, 43, 62], substantially ignoring the opponents' possibilities. Our objective is to give a unified account of coalitional rationality combining and extending already existing logics for strategic ability and preferences. In line with the study of cooperative games in the previous chapter, our logical analysis will make the following features of coalitional rationality explicit:

- **betterness**, i.e. the comparison of a coalition's possibilities, as in sentences like "we had better not do this", "we had better defect";
- **choice restriction**, i.e. the transformation in a coalitional choice space brought about by a possible move of the opponents, typical of the conditional reading of strategic decisions: "if they do this, then that will hold", "if our opponent defects, then we had better cooperate".

Once a language to reason about coalitional rationality is available, we can also start reasoning about the *regulation* of conflicts among different coalitions. Let us exemplify this point further.

In strategic interaction plenty of situations can arise in which individual preferences are not compatible and coalitions can steer the game in many possible directions. The enactment of norms can in these cases be used to regulate such conflicts. By enacting a norm we mean *the introduction of a normative constraint on individual and collective choices to achieve some systemic desiderata*. In doing so, we distinguish two perspectives: the first, called *utilitarian* or *internal*, evaluates coalitional actions only from the point of view of their rationality; the second, called *systemic* or *external*, evaluates coalitional action from the point of view of pre-stablished normative standards, the latter possibly independent of rationality constraints. The two complementary perspectives on regulation will be studied together, for the abstract case of coalitional games.

Attention will also be devoted to the specific case of coalitional games that represent strategic games. The results in Chapter 3 have questioned the capability of Coalition Logic to express their logical structure. In this respect the following issues will be addressed:

- understanding whether languages such as Coalition Logic can still be used to reason about strategic games;
- understanding the right abstraction level to express the characteristic features of coalitional ability in games.

Summing up, the present chapter aims at bringing various logical languages for preferences, coalitional ability and norms within a unified formal framework to account for coalitional rationality in strategic interaction.

Chapter structure: Section 4.2, based on joint work with Jan Broersen, Rosja Mastop and John-Jules Meyer[18], presents a logic of coalitional rationality, able to characterize many of the structural notions studied in Chapter 3, via a combination of

standard preference logics and Coalition Logic. Section 4.3, based on joint work with Jan Broersen, Rosja Mastop and John-Jules Meyer[18], deals with regulation of coalitional rationality, formalizing the internal and the external perspective on norms. Section 4.4, based on joint work with Wojtek Jamroga and Valentin Goranko [30], studies an extension of Coalition Logic to characterize truly playable effectivity functions and coalitional power in strategic games in general. Mathematical properties, such as finite axiomatization, completeness and finite model property are provided for some interesting fragments of the language. Section 4.5 finally discusses related works and issues that are left open.

4.2 Reasoning on Coalitional Rationality

This section is devoted to the construction of a logical language, based on standard modal accounts of preferences and coalitional ability [46, 54], that makes the notions of betterness and choice restriction explicit.

4.2.1 Betterness and optimality

Classically preference logics have been developed in analytic philosophy and philosophical logic to provide a precise description of notions such as having a goal, desire, intention etc.. The work of von Wright [71, 73] laid the ground for a modal logic treatment of preference, where the information of what situations are to be preferred (or dispreferred) to the present one are encoded in the modal operators. Subsequent contributions, mostly originated around the work of Johan van Benthem [65, 64], studied the rich mathematical structure behind these modalities. As typical with modal logic, properties of relational structures such as preference relations are immediately captured by the logical structure of the modal language. Examples of such modalities are the ones given in Section 2.3.1 of Chapter 2.

However preference logics have mostly been studied in isolation, without explicitly drawing their natural connection with the logics for coalitional action, with a few recent exceptions [44, 68]. Therefore, preferences, when lifted to sets, can be naturally associated with effectivity functions and share the same modal flavour of coalitional actions. As we have remarked in the previous chapter, in [66] it is rightly noticed how preference liftings can be extremely useful to describe betterness in games. It is starting from this intuition that we move on to characterize Pareto optimal choices.

Characterizing Pareto optimal choices

Pareto optimal choices are maximal sets according to a preference order on sets in a coalitional effectivity function. A first attempt to characterize them can be made using the results in [66], which logically characterize binary preference liftings similar to the ones we have discussed in Chapter 3. However those liftings are of general character:

- They only compare relative preference between two propositions and not absolute preference of a proposition with respect to all others;
- They do not restrict attention to subsets of the possible propositions, as happens in establishing maxima in a coalitional effectivity function.

Surprisingly enough not only is a language with two modalities for preference ($\Diamond_i^{\preceq}$ and $\Diamond_i^{\succ}$ defined in Section 2.3.2), one for coalitional ability ($[C]$ defined in Section 2.3.1) and the global modality (A defined in Section 2.3.2), expressive enough to characterize Pareto optimality in all its forms, but in some cases the global modality and the modality for strict preference are not even needed.

We call this language $\mathcal{L}^{\preceq,\succ,g,[C]}$ and the grammar of its formulas φ goes as follows:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond_i^{\preceq}\varphi \mid \Diamond_i^{\succ}\varphi \mid A\varphi \mid [C]\varphi$$

As usual, p is an atomic proposition and $\Box_i^{\preceq}, \Box_i^{\succ}, E, \langle C\rangle$ are used as abbreviations for $\neg\Diamond_i^{\preceq}\neg, \neg\Diamond_i^{\succ}\neg, \neg A\neg$ and $\neg[C]\neg$ respectively. Their interpretation with respect to Coalitional Game Models has already been given in Chapter 2.

The following proposition establishes a characterization of (weak) Pareto optimality with the $(\forall,\forall)$ preference lifting.

Proposition 23 *Let M be a Coalitional Game Model and φ a formula of $\mathcal{L}^{\preceq,\succ,g,[C]}$. The following holds:*

$PO_{C,w}(\varphi^M, \succ_i^{(\forall,\forall)})$ if and only if $M, w \models [C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$

Proof ($\Rightarrow$)

Let us assume that $PO_{C,w}(\varphi^M, \succ_i^{(\forall,\forall)})$, i.e. that φ^M is a Pareto optimal choice for coalition C at world w according to the $(\forall,\forall)$ preference lifting. This means, by Definition 19, that for no $X \in E(w)(C)$, $X \succ_i^{(\forall,\forall)} \varphi^M$ for all $i \in C$ and that $\varphi^M \in E(w)(C)$. But this means that for all $X \in E(w)(C)$ $\exists x \in X, \exists y \in \varphi^M, \exists j \in C$ such that $x \preceq_j y$. By the interpretation of the modal operators, no set $X \in E(w)(C)$ is such that $X \subseteq (\bigwedge_{i\in C} \neg\Diamond_i^{\preceq}\varphi)^M$. So we can conclude that $M, w \models [C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$.

($\Leftarrow$)

$M, w \models [C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$ means that $\varphi^M \in E(w)(C)$ and $(\neg \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi)^M \notin E(w)(C)$. So, by outcome monotonicity, every $X \in E(w)(C)$ contains world x such that $M, x \models \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$, which means that for some player $i \in C$ we have that $x \preceq_i y$ for some $y \in \varphi^M$. In sum $X \in E(w)(C)$, $X \succ_i^{(\forall,\forall)} \varphi^M$ for all $i \in C$.

Before going on to characterize the other forms of Pareto optimal choice a comment is needed. The formula $[C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$, which corresponds to φ being Pareto optimal choice for coalition C at w following the $(\forall,\forall)$ preference lifting, says in words that coalition C can choose φ and cannot avoid that some of its players prefer φ, which may seem far away from the concept of Pareto optimality of a choice. However applying elementary reasoning we can rewrite the formula into this equivalent one:

$$[C]\varphi \wedge \neg[C] \bigwedge_{i\in C} \Box_i^{\preceq}\neg\varphi$$

whose reading is more intuitive: coalition C can force φ and cannot force any set whose states are only worse than states satisfying $\neg\varphi$, which is much closer to the original intuition behind Pareto optimal choices.

Let us now consider the other cases.

Proposition 24 *$PO_{C,w}(\varphi^M, \succ_i^{(\forall,\exists)})$ if and only if $M, w \models [C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$*

Proof ($\Rightarrow$)

$PO_{C,w}(\varphi^M, \succ_i^{(\forall,\exists)})$ means that $\varphi^M \in E(w)(C)$ and for no $X \in E(w)(C)$, $X \succ_i^{(\forall,\exists)} \varphi^M$ for all $i \in C$. $X \succ_i^{(\forall,\exists)} \varphi^M$ for all $i \in C$ means that for all $x \in X$ there is a $y \in \varphi^M$ such that $x \succeq_i y$ and not $y \succeq_i x$ for all $i \in C$. Its negation is equivalent to saying that there exists a player j and an element $x \in X$ for any X s.t. for all $y \in \varphi^M$, not $x \succeq_j y$ or $y \succeq_j x$. In either case it is is not true that $x \succ_j y$. For this reason we have that $M, x \models \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$. In turn, using the assumptions together with the interpretation of the modal operators, this means that $M, w \models [C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$.

($\Leftarrow$)

Suppose $M, w \models [C]\varphi \wedge \langle C\rangle \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$. This means $\varphi^M \in E(w)(C)$ and, by outcome monotonicity, that for all $X \in E(w)(C)$ there is an $x \in X$ such that $M, x \models \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$. By the semantics of the modal operators there is no world $y \in \varphi^M$ such that $x \succ_j y$ for some $j \in C$. As a result $PO_{C,w}(\varphi^M, \succ_i^{(\forall,\exists)})$.

Proposition 25 *$PO_{C,w}(\varphi^M, \succ_i^{(\exists,\exists)})$ if and only if $M, w \models [C]\varphi \wedge A \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$*

Proof ($\Rightarrow$)

$PO_{C,w}(\varphi^M, \succ_i^{(\exists,\exists)})$ means that for no $X \in E(w)(C), X \succ_i^{(\exists,\exists)} \varphi^M$, for all $i \in C$ and $\varphi^M \in E(w)(C)$. By the properties of the preference relation we have that for all $x \in X$ and $y \in \varphi^M$, not $x \succ_i y$ for all $i \in C$. But $W \in E(w)(C)$, by the fact that E is outcome monotonic and nonempty. So we can conclude $M, w \models [C]\varphi \wedge A \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$.

($\Leftarrow$)

$M, w \models [C]\varphi \wedge A \bigvee_{i\in C} \neg\Diamond_i^{\succ}\varphi$ clearly implies that no $X \in E(w)(C)$ is such that $X \succ_i^{(\exists,\exists)} \varphi^M$ for $i \in C$. We already know that $M, w \models [C]\varphi$, so $PO_{C,w}(\varphi^M, \succ_i^{(\exists,\exists)})$.

Proposition 26 *$PO_{C,w}(\varphi^M, \succ_i^{(\exists,\forall)})$ if and only if $M, w \models [C]\varphi \wedge A \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$*

Proof ($\Rightarrow$)

$PO_{C,w}(\varphi^M, \succ_i^{(\exists,\forall)})$ means that for no $X \in E(w)(C), X \succ_i^{(\exists,\forall)} \varphi^M$, for all $i \in C$, and $\varphi^M \in E(w)(C)$. So for all $x \in X$ there is $y \in \varphi^M$, s.t. not $x \succ_i y$ for all $i \in C$. By connectedness $y \succeq_i x$. But $W \in E(w)(C)$, by the fact that E is outcome monotonic and nonempty. By the semantics of the modal operators we can conclude that $M, w \models [C]\varphi \wedge A \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$.

($\Leftarrow$)

$M, w \models [C]\varphi \wedge A \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi$ implies that no $X \in E(w)(C)$ is such that $X \succ_i^{(\exists,\forall)} \varphi^M$ for all $i \in C$. We already know that $M, w \models [C]\varphi$ and we can conclude that $PO_{C,w}(\varphi^M, \succ_i^{(\exists,\forall)})$.

As for the strong version of Pareto optimal choices analogous results hold, whose proofs follow the same pattern of the corresponding quantified weak versions.

Proposition 27 *The following hold:*

- $SPO_{C,w}(\varphi^M, \succ_i^{(\forall,\forall)})$ *if and only if* $M, w \models [C]\varphi \wedge \langle C\rangle(\bigvee_{i\in C} \Diamond_i^{\prec}\varphi \vee \bigwedge_{i\in C} \Diamond_i^{\preceq}\varphi)$.
- $SPO_{C,w}(\varphi^M, \succ_i^{(\forall,\exists)})$ *if and only if* $M, w \models [C]\varphi \wedge \langle C\rangle(\bigwedge_{i\in C} \neg\Diamond_i^{\succ}\varphi \vee \bigvee_{i\in C} \neg\Diamond_i^{\succeq}\varphi)$.
- $SPO_{C,w}(\varphi^M, \succ_i^{(\exists,\exists)})$ *if and only if* $M, w \models [C]\varphi \wedge A(\bigwedge_{i\in C} \neg\Diamond_i^{\succ}\varphi \vee \bigvee_{i\in C} \neg\Diamond_i^{\succeq}\varphi)$
- $SPO_{C,w}(\varphi^M, \succ_i^{(\exists,\forall)})$ *if and only if* $M, w \models [C]\varphi \wedge A(\bigwedge_{i\in C} \Diamond_i^{\preceq}\varphi \vee \bigvee_{i\in C} \Diamond_i^{\preceq}\varphi)$

Summing up, a full spectrum of maximality relations within a coalitional effectivity function can be characterized by using operators to describe properties of relational structures. In half of the cases, namely the weak and strong version of the Pareto optimal choices with the $(\forall, \forall)$ and $(\forall, \exists)$ liftings, these maximality relations among sets of states can be characterized only resorting to modal operators that are designed to express binary relations among states in a models.

4.2.2 Choice restrictions

An effectivity function encodes a specific view of coalitional ability, that of a coalition being able to force the game to end up in a certain set, whatever way the opponents decide to act. This rather strong representation does not consider the effects on a coalitional strategic ability should the opponents decide to make a particular move, which was what we tried to capture by the notion of subgame in Chapter 2. For this reason a fine-grained representation of strategic reasoning cannot be expected from a language that can only reason about effectivity functions.

As pointed out in [62], p.1:

> Much of game theory is about the question whether strategic equilibria exist. But there are hardly any explicit languages for defining, comparing, or combining strategies as such — the way we have them for actions and plans, maybe the closest intuitive analogue to strategies. True, there are many current logics for describing game structure — but these tend to have existential quantifiers saying that "players have a strategy" for achieving some purpose, while descriptions of these strategies themselves are not part of the logical language.

The expressive power of our logic for strategic ability, i.e. Coalition Logic, seems to inherit this limitation: the fact that at some model M and world w and for some coalition C we have $M, w \models [C]\varphi$ only makes reference to the power of coalition C, independently of the possible decisions of $\overline{C}$. Let us look at this in more details.

Example 15 (Saying it in Coalition Logic) *A strategic game like the prisoner's dilemma –let us again resort to its representation in Figure 2.2– can be naturally rewritten as a Cooperative Game Model. In any world w representing the prisoner's dilemma in a model PD, we therefore have that* $PD, w \models [\{Row\}](Row\ defects) \wedge \neg[\{Row\}](Column\ defects)$, *where the*

propositions are interpreted as expected. On the other hand it seems that we cannot express what {Row} can do given that *{Column} defects. If this were the case we would also be able to express that {Row} has a strategy forcing that {Row} defects and {Column} defects and a strategy forcing that {Row} cooperates and {Column} defects. Bringing this observation at the model level, we should have that PD, w* $\models$ [{*Row*}](*Row defects and Column defects*) $\wedge$ [{*Row*}](*Row cooperates and Column defects*). *By the validity of outcome monotonicity in every coalition model (see Chapter 2), we would then get PD, w* $\models$ [{*Row*}](*Column defects*), *which is at odds with our initial claim.*

The example given shows how coalition logic modalities do not easily accommodate a notion of conditional action. For this reason we introduce a modal operator to express explicitly in our language that a coalition can force some outcome *given* what its opponents do. This should not be confused with the reasoning patterns in extensive games, in which players reason on the best action to take *after* their opponents have moved, nor with the notion of ability to guarantee an outcome *independently* of what the other players do, which is the typical reading of the operators in the various game logics.

The *subgame operator*

To model choice restrictions we introduce a modal expression of the form

$$[C \downarrow \psi]\varphi$$

whose informal reading is: "in case coalition C chooses ψ, φ holds", where φ and ψ are formulas of the language $\mathcal{L}^{\leq,>,g,[C]}$ extended with modalities of the form $[C \downarrow \psi]$. We define the dual $\langle C \downarrow \psi\rangle\varphi$ as an abbreviation of $\neg[C \downarrow \psi]\neg\varphi$. Intuitively what we do is to talk about what holds in case the choice ψ of coalition C is performed. Thanks to this operator formulas of the form

$$[C \downarrow \psi][\overline{C}]\varphi$$

allow us to talk of the *restriction* in the coalitional ability of $\overline{C}$ that is caused by coalition C choosing ψ. This restriction clearly resembles the one in the definition of subgame given in Definition 5. For this reason it will be called *the subgame operator*.

Its formal interpretation goes as follows:

$$M, w \models [C \downarrow \psi]\varphi \Leftrightarrow \psi^M \in E(w)(C) \text{ implies } M, w \downarrow_{(C,\psi^M)} \models \varphi$$

The interpretation of the operator has a conditional reading: if a coalition C has a certain choice ψ^M at w, then the world where this choice is actually executed ($w \downarrow_{(C,\psi^M)}$, to be formally defined next) makes a certain proposition φ true. Notice that the capacity of C to choose ψ^M is the precondition for C to actually execute ψ^M.

The *updated world* $w \downarrow_{(C,\psi^M)}$ is so defined:

- It inherits the same valuation function as w

- It updates the effectivity function $E(w \downarrow_{(C,\psi^M)})$.

Definition 32 *Let E be an effectivity function defined on a set of outcomes W and a set of players N and let $C, C' \subseteq N$, $X \subseteq W$ and $w \in W$. $E(w \downarrow_{(C,X)})$ is defined in the following way:*

$$
\begin{array}{ll}
E(w \downarrow_{(C,X)})(C') \doteq (\{X\})^{sup} & \text{for } C' \cap C \neq \emptyset \\
E(w \downarrow_{(C,X)})(C') \doteq (E(w)(C') \sqcap X)^{sup} & \text{for } C' \cap C = \emptyset \text{ and } C' \neq \emptyset \\
E(w \downarrow_{(C,X)})(C') \doteq E(w)(C') & \text{for } C' = \emptyset
\end{array}
$$

The way the relation is updated, illustrated in Figure 4.1, deserves some comment. A distinction is made between the strategic ability update of the players who made a certain choice ψ and all the other players. After coalition C has made a choice ψ, all the coalitions involving agents belonging to C are given $(\{\psi^M\})^{\mathrm{sup}}$ as a choice set. This view maintains that a coalition comprising players in the coalition that has already chosen cannot further influence the outcome of the game. This fact implies that the subgame operator is not superadditive, in the sense given in [54], that is, bigger coalitions need not have bigger power. Said in other words, we do not allow players to make a choice within a certain coalition and then, at the same time, to make a choice within different coalitions. The models of reference are strategic games, in which strategies are decided in the beginning once and for all [51]. The other (nonempty) coalitions instead *truly update* their choice set having it restricted by the choice of C. Restriction is implemented in this case by intersecting the effectivity function with the move that has been carried out. In case for instance C chooses to force ψ and $\overline{C}$ was able to choose ξ, then given the choice by C, $\overline{C}$ is able to force $\xi \wedge \psi$. The coalitional relation at worlds different from the one where the choice is made remains instead unchanged. This means that the update is local. Again, the references are strategic games, where the sequential structure of strategies is substantially ignored. Notice also that by the last condition the empty coalition never gains power. In sum the strategic ability update is governed by three principles:

- the **irrelevance of hybrid coalitions**, that does not allow the members of the coalition that moved to further influence the interaction,
- the **restriction of opponents' choices**, that truly updates the effectivity function of the coalitions opposing the one that moved,
- the **locality of the update**, that only updates the power of nonempty coalitions at one world.

The following fact follows directly from Proposition 13:

Proposition 28 *For every C,w, $\psi^M \in E(w)(C)$, we have that*

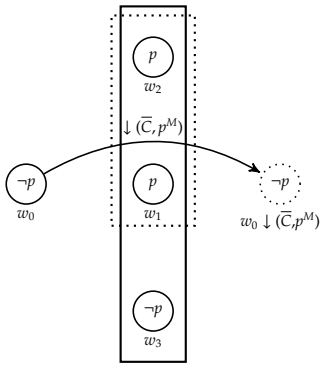

Figure 4.1: Updating worlds. $w_0 \downarrow (\overline{C}, p^M)$ inherits from w_0 the same valuation function but it updates its effectivity function. At w_0 coalition C cannot achieve the property p, as the set $\{w_1, w_2\}$ is not in its effectivity function. However, should its opponents do p, then C would also be able to achieve p, as the set $\{w_1, w_2\}$ is in its effectivity function at the updated world $w_0 \downarrow (\overline{C}, p^M)$.

1. $E(w \downarrow_{(C,\psi^M)})$ *is outcome monotonic.*
2. $E(w \downarrow_{(C,\psi^M)})$ *has IOEC whenever* $E(w)$ *has IOEC.*
3. $E(w \downarrow_{(C,\psi^M)})$ *is regular whenever* $E(w)$ *is superadditive.*

Proof *The proof is a direct consequence of the definitions.*

The proposition shows how updating worlds preserves many characteristic features of the original effectivity function, for instance outcome monotonicity (item 1). IOEC is also preserved under updates (item 2) but this is not the case with superadditivity (item 3): updating superadditive effectivity functions only preserves the property of regularity. The latter proposition shows that updates might cause the loss of important properties of the original effectivity function. Let us have a closer look at what happens.

On the nature of the updates

The definition of update allows to jump from each world in a model to its updated counterpart, within one coalition model. Figure 4.1 is once again a neat illustration of this fact. The update operation is treated as a function that takes a triple world-coalition-set as a value and returns a world. A consequence is that the coalition frames are *special frames* that contain all instances of their updates. In other words, they are *closed under subgames* .

Definition 33 (Closure under subgames) *Let* $F = (W, E)$ *be a coalition frame.* F *is said to be* closed under subgames *if and only if* $X \in E(w)(C)$ *implies that* $w \downarrow_{(C,X)} \in W$.

This is a frame condition and, as many others that we have seen so far, can be modally characterized.

Proposition 29 *Let $F = (W, E)$ be a coalition frame. The following holds:*
$F \models [C]\xi \leftrightarrow \langle C \downarrow \xi\rangle\top$ if and only if F is closed under subgames.

Proof *From right to left, it is straightforward. From left to right assume $F \models [C]\xi \leftrightarrow \langle C \downarrow \xi\rangle\top$. Consider now a set $X \in E(w)(C)$ and take a valuation function V such that $\xi^M = X$ for some M based on F. By the assumptions we have that $M, w \models \langle C \downarrow \xi\rangle\top$, which means that there is a world $w \downarrow_{(C,\xi^M)} \in W$ such that $M, w \downarrow_{(C,\xi^M)} \models \top$, i.e. F is closed under subgames.*

A different path could be taken in defining the interpretation of the subgame operator, closer to the notion of model update in Dynamic Epistemic Logic [69]. The subgame operator could be interpreted in *updated models* created from the original one, i.e. 'controlled' transformations of the original models of which they preserve some relevant properties. In this case, the updated models would be equivalent to the original model as regards the set of outcomes, the set of players and the valuation function. But they would differ in the effectivity function, changed according to the rules we have discussed, only for the world where the choice takes place. This would make the model a sort of metamodel, a model that relates other models, and how these change.

Summing up, there is a substantial difference between the two approaches:

- The first, updating the world, is not constructive. Each instance of a world updated by a coalitional choice needs to be present in the original model. We have called these kind of models *closed under subgames*.
- The second, updating the model, is constructive. A new model is created, following the rules of the update. The subgame operator is interpreted in what we have called *metamodels*.

The difference between the two approaches has consequences at the logical level. Think of the universal modality: when interpreted in models that are closed under subgames, saying that at some world w the formula $A\varphi$ holds means that φ is true at every world but also at each of its updates. Instead, interpreting $A\varphi$ in a model that is not closed under subgames says nothing about φ holding once the updated is performed. On these grounds, and in light of the result in the coming section, we prefer to work with the first, nonconstructive approach.

A surprising reduction

Even though the interpretation of the update operator in models that are closed under subgames may look complex and rather demanding, its structural behaviour is rather simple. The validities in Table 4.1 allow us to translate every sentence where the operator is occurring to a sentence where the operator is not occurring. Notice that by the interpretation of the update each relation that is not changed in the subgame does behave properly. As an example, consider the preference

modality $\Box_i^{\geq}$, whose reduction axiom can be obtained by replacing it to the A in the reduction axiom for the global modality. [1]The proof of their validity is given in the appendix (Section B.1).

What is striking about this reduction is that conditional reasoning can be expressed in Coalition Logic, contrary to what Example 15 seemed to suggest. As an example, the formula

$$[C]\varphi \wedge [\overline{C}](\varphi \rightarrow \psi)$$

shows in a natural way how the coalitional power of coalition $\overline{C}$ is changed should coalition C decide to choose φ.

Example 16 (Back to the game) *By means of the subgame operator it becomes possible to make the conditional aspect of strategic reasoning explicit. The only critical point to be taken into account is that subgames should also be represented in the models. We won't be overly formal in doing this, as the procedure boils down to copying the effectivity function of the original game and adding the instances of the subgames. In the tuple PD', w, representing the prisoner's dilemma closed under subgames we have that*

PD', w $\models$ [{*Row*} $\downarrow$ *Row defects*]([{*Column*}](*Column defects and Row defects*)$\wedge$ [{*Column*}](*Column cooperates and Row defects*))

i.e. given the choice by the row player to defect, Column can see to it that both players defect and can see to it that he cooperates and its opponent defects.

Undomination and Subgames

The logical structure of undominated choices can be clarified by making use of the subgame operator. To check that a choice is undominated, for any preference lifting that we might want to consider, we need to check Pareto optimality in each choice restriction. While Pareto optimality can be characterized with preferences and coalitional modalities, choice restrictions can be made explicit using the subgame operator. There are however structural limitations in the possible uses of the subgame operator:

- Formulas of the form $[C \downarrow \psi]\varphi$ talk about properties that hold given *one choice restriction*, namely the choice of ψ by coalition C;
- Expressions of the form $\varphi^M \triangleright_{C,w}$ (Definition 22) concern a property holding given *all choice restrictions*, namely each choice $X \in E(w)(\overline{C})$ by coalition $\overline{C}$.

Taking this structural limitation into account, undomination can be characterized in $\mathcal{L}^{\preceq,\succ,g,[C]}$ extended with the subgame operator.

The first characterization will be carried out at the model level, assuming that *every coalition can force only a finite amount of propositions*.

[1]This would not have been so if we had taken a metamodel approach. The axiom for the global modality does not hold with that approach and more restrictive assumptions on the models would be required.

Proposition 30 *Let $\{\psi_1^M, ..., \psi_n^M\} = E(w)(\overline{C})$ be a coalitional effectivity function in a Coalitional Game Model M closed under subgames. We have the following :*

$$\varphi^M \rhd_{C,w} \text{ if and only if } M, w \models \bigwedge_{\psi_k \in \{\psi_1,...,\psi_n\}} [\overline{C} \downarrow \psi_k]([C](\varphi \wedge \psi_k) \wedge \langle C \rangle \bigvee_{i \in C} \Diamond_i^{\preceq}(\varphi \wedge \psi_k))$$

Proof *This follows as a direct consequence of Definition 22 and the interpretation of the modal operators.*

The formula says that if finitely many choices are available to a coalition then undominated choices can be verified by finitely checking Pareto optimality. A finite conjunction of formulas starting with the subgame operator, plus the characterizing formulas of Pareto optimal choices, are enough to express this concept. Notice that $\varphi^M \cap \psi_k^M$ can be empty, but in this case $\bot^M = \emptyset$ is never Pareto optimal, thanks to the use of the $(\forall, \forall)$ preference lifting that makes $X \succ_i^{(\forall,\forall)} \bot^M$ trivially true for any X.

The following proposition holds when undomination is taken to be a frame condition.

Proposition 31 *Let $\mathbb{F}$ be the class of cooperative game frames closed under subgames and let $F \in \mathbb{F}$ be one of them. The following holds:*

$F \models [C]\varphi \rightarrow [\overline{C} \downarrow \psi]([C](\varphi \wedge \psi) \wedge \langle C \rangle \bigvee_{i \in C} \Diamond_i^{\preceq}(\varphi \wedge \psi))$ if and only if each $X \in E(w)(C)$ is such that $X \rhd_{C,w}$

Proof *From right to left the proof is straightforward. From left to right, let us assume that $X \in E(w)(C)$ and that $E(w)(\overline{C}) \neq \emptyset$, otherwise the proposition holds trivially. Let us consider a set $Y \in E(w)(\overline{C})$ and a valuation function V such that $V(p) = X$ and $V(q) = Y$. We have, by the characterization result of Pareto optimal choice, that $F, V \models [C]p \rightarrow [\overline{C} \downarrow q][C](p \wedge q) \wedge \langle C \rangle \bigvee_{i \in C} \Diamond_i^{\preceq}(p \wedge q)$. As a consequence, $X \rhd_{C,w}$.*

The proposition allows for interesting observations. First of all, since we are characterizing undomination as a property of the frames, we do not need any restriction on the choices of coalitions. Second, we can characterize a much finer notion of undomination and Pareto optimality of choice: we can talk about all sets in an effectivity function, and not only those that are the truth set of some proposition.

In the class of effectivity functions that have IOEC the characterization in the case of the grand coalition is rather elegant and does not require particularly restrictive assumptions.

Proposition 32 *Let $M = (W, E, \geq_i, V)$ be a Coalitional Game Model closed under subgames such that $E(w)$ has IOEC. The following holds,*

$\varphi^M \rhd_{N,w}$ if and only if $M, w \models [C]\varphi \wedge \langle C \rangle \bigvee_{i \in N} \Diamond_i^{\preceq}\varphi$

Proof *Straightforward once we notice that $E(w)(N) \sqcap X = E(w)(N)$ for each $X \in E(w)(\emptyset)$ in effectivity functions that have IOEC.*

	Axioms
A6	$[C]\xi \leftrightarrow \langle C \downarrow \xi\rangle\top$
A7	$[C \downarrow \xi]p \leftrightarrow ([C]\xi \rightarrow p)$
A8	$[C \downarrow \xi]\neg\varphi \leftrightarrow ([C]\xi \rightarrow \neg[C \downarrow \xi]\varphi)$
A9	$[C \downarrow \xi](\varphi \wedge \psi) \leftrightarrow ([C \downarrow \xi]\varphi \wedge [C \downarrow \xi]\psi)$
A10	$[C \downarrow \xi]A\varphi \leftrightarrow ([C]\xi \rightarrow A\varphi)$
A11	$[C \downarrow \xi]\Box_i^{\leq}\varphi \leftrightarrow ([C]\xi \rightarrow \Box_i^{\leq}\varphi)$
A12	$[C \downarrow \xi][C']\varphi \leftrightarrow ([C]\xi \rightarrow [C'](\xi \rightarrow \varphi))$ (for $C' \cap C = \emptyset$ and $C' \neq \emptyset$)
A13	$[C \downarrow \xi][C']\varphi \leftrightarrow A(\xi \rightarrow \varphi)$ (for $C' \cap C \neq \emptyset$)
A14	$[C \downarrow \xi][C']\varphi \leftrightarrow ([C]\xi \rightarrow [C']\varphi)$ (for $C' = \emptyset$)
	Rules
R1	$\varphi \Rightarrow [C \downarrow \xi]\varphi$
R2	$\varphi \leftrightarrow \psi \Rightarrow [C \downarrow \xi]\chi \leftrightarrow [C \downarrow \xi]\chi[\varphi/\psi]$

Table 4.1: Axioms and Rules for the subgame operator

Stocktaking The modal operators defined so far allow us to express that a choice in a coalitional effectivity function is rational for that coalition, and in some cases these operators are even reducible to a language that can only talk about abstract effectivity and preference between states. But this is not the end of the story: reasoning about rationality is also a way of reasoning about what a coalition *should do* with the choices in its effectivity function, which was the reason why we decided to move away from that abstract representation of power. It is clear that the notion of rationality itself carries a normative content: if a choice is rational, then the coalition—in some sense—should perform it. But rational choices can be many and there can be many coalitions behaving rationally, often aiming at bringing about properties that conflict with one another. How to deal then with the regulation of a strategic interaction?

4.3 Regulating Coalitional Choices

This section describes two complementary views on the regulation of coalitional choices within a unified deontic language.

- In the first, norms assume an *internal* or *utilitarian* character: actions that are permitted for a coalition are those that are best for the coalition itself (or, in a general sense, for some bigger coalition including it). The utilitarian view of norms is started in [41] and taken up and generalized in [44].
- In the second, norms assume an *external* or *systemic* character: choices are judged against general interests, specified from outside the system. This is the classical view of deontic logic, taken up in Meyer's Dynamic Deontic Logic approach [48].

Hereby the ingredients of our deontic language will be two: first, a set of formulas to express coalitional rationality, and second a set of formulas to express violations.

The internal view: a deontic logic for coalitional rationality

The utilitarian approach characteristic of the internal view on norms can be summarized by Horty's statement ([41] p.70),

> The general goal of any utilitarian theory is to specify standards for classifying actions as right or wrong: and in its usual formulation act utilitarianism defined an agent's action in some situation as right just in case the consequences of that action are at least as great in value as those of any of the alternatives open to the agent, and wrong otherwise.

Summing up, in Horty's view, to reason about what coalitions ought to do we then need to reason about their own coalitional rationality.

The language we have defined in the previous section is well-suited to define coalitional rationality. It is however convenient to expand it by introducing modal formulas of the form

$$[rational_C]\varphi$$

where C is a coalition and φ a formula of the language. Formulas of this form succinctly indicate what choices are rational for a coalition, where the notion of rationality acquires a semantics in terms of undomination. The satisfaction relation of the formulas with respect to a tuple M, w is defined as follows:

$$M, w \models [rational_C]\varphi \quad \text{if and only if} \quad \varphi^M \rhd_{C,w}$$

Also for the sake of simplicity we will work with finite Coalitional Game Models closed under subgames. By the results of the previous chapter (Proposition 30),

the operator $[rational_C]$ is in these models a mere abbreviation of formulas of the language of Coalition Logic with a preference modality and the subgame operator. Within this language we can define the classical deontic operators of forbiddance, obligation and permission. We do it in two ways, first defining the notion of absolute commands, i.e. classical deontic operators that tell coalitions what to do independently of the other coalitions, second defining the notion of policy, i.e. norms that coordinate the choices between coalitions and that generalize the first ones.

Absolute commands We define operators of the form $X^{C'}(C, \varphi)$, for $X \in \{P, O, F\}$ to say that in the interest of coalition $C' \supseteq C$ it is forbidden, permitted or obliged for C to choose φ.

Definition 34 (Deontic Operators) *For $C \subseteq C' \subseteq N$ the following statements define forbiddance, permission and obligation:*

$$\begin{aligned} F^{C'}(C, \varphi) &:= [C]\varphi \rightarrow \neg[rational_{C'}]\varphi \\ P^{C'}(C, \varphi) &:= \neg F^{C'}(C, \varphi) \\ O^{C'}(C, \varphi) &:= F^{C'}(C, \neg\varphi) \end{aligned}$$

The first operator says that in the interest of the bigger coalition C' it is forbidden for coalition C to choose φ. This is equivalent to saying that if coalition C can choose φ then it is not rational for coalition C' to choose φ, i.e. φ is undominated for C'. The second operator says that in the interest of the bigger coalition C' it is permitted for coalition C to choose φ, which is equivalent to saying that it is not true that in the interest of coalition C' is forbidden for coalition C to choose φ. The third operator says that it is obligatory for coalition C to choose φ, which is equivalent to saying that coalition C is forbidden, in the interest of the bigger coalition C', to choose $\neg\varphi$.

The deontic operators are relatively simple reductions to operators that talk about coalitional ability and rationality. The way this reduction is carried out resembles Meyer's classical account of norms in Dynamic Deontic Logic [48]. Likewise, our operators display a number of features typical of that approach.

First of all, in formulas of the form $F^{C'}(C, \varphi)$ the subformula φ, that corresponds to the choice that is forbidden for C, may not be present in its effectivity function. In Meyer's account the fact that an action α is forbidden is defined as follows:

$$F(\alpha) := [\alpha]viol$$

that is, α is forbidden if and only if each terminating execution of α leads to an undesirable state. Notice that there may be no terminating executions of α. [2] These choices are in some sense trivially forbidden, as their execution is not even possible.

[2] It is also worth noting how the utilitarian framework presented here suggests correspondence between violations and irrational choices.

Furthermore φ may not be present in the effectivity function of the bigger coalition C'. This can be obtained in games that do not enjoy the property of superadditivity (Definition 7), where bigger coalitions do not enjoy bigger power.

The other two operators also follow the same view of Dynamic Deontic Logic and for them the same remarks hold. In Meyer's account, that is reflected in our definitions, $P(\alpha) := \neg F(\alpha)$, i.e. permission is the opposite of forbiddance; and $O(\alpha) := F(\neg\alpha)$, obligation of performing an action means forbidding refraining from that action.

At times we will be concerned with the special case of rationality in the interest of the coalition itself and rationality in the interest of all the players. As for the first case, notice that saying that something is permitted in the interest of the coalition itself ($P^C(C, \varphi)$) is equivalent to saying that it is rational for that coalition ($[rational_C]\varphi$).

For each coalition we have now defined a spectrum of norms that reflect a notion of rationality ranging from self interest $F^C(C, \varphi)$ to social interest $F^N(C, \varphi)$. Let us now focus on the formal properties of this spectrum. The following validities and invalidities shed light on its logic.

validities	
1	$P^{C'}(C, \varphi) \leftrightarrow \neg O^{C'}(C, \neg\varphi)$
2	$[C]\varphi \wedge [rational_N]\varphi \leftrightarrow P^N(C, \varphi) \wedge [N]\varphi$
3	$P^{C'}(C, \varphi) \vee P^{C'}(C, \psi) \rightarrow P^{C'}(C, \varphi \vee \psi)$
4	$[C]\varphi \wedge O^N(C, \varphi) \rightarrow O^N(D, \varphi)$
5	$[rational_C]\varphi \wedge \neg[rational_N]\varphi \rightarrow F^N(C, \varphi)$
6	$F^{C'}(C, \varphi) \wedge F^{C'}(C, \psi) \rightarrow F^{C'}(C, \varphi \wedge \psi)$

	non-validities
1	$O^{C'}(C,\varphi) \wedge O^{C'}(C,\psi) \rightarrow O^{C'}(C,\varphi \wedge \psi\rangle$
2	$P^{C'}(C,\varphi \vee \psi) \rightarrow P^{C'}(C,\varphi) \vee P^{C'}(C,\psi)$
3	$O^{C'}(C,\varphi) \leftrightarrow \neg O^{C'}(C,\neg\varphi)$
4	$[rational_C]\varphi \leftrightarrow [rational_N]\varphi$
5	$O^{C'}(C,\varphi) \rightarrow P^{C'}(C,\varphi)$
6	$O^{C'}(C,\varphi) \rightarrow [C]\varphi$

The first validity says that the presence of permission is equivalent to the absence of conflicting obligations. In our framework this resembles what the legal philosophers call the *sealing legal principle*, i.e. "whatever is not forbidden is thereby permitted", [72]. When thinking of permission as rational action (even when the group of reference is a bigger coalition), the principle seems to make perfect sense, as a choice not being dominated (forbidden) is in fact a rational choice (and thereby permitted).

The second proposition makes the relation between rational action for the grand coalition and permission explicit. Whenever a proposition can be forced by the grand coalition and it is permitted for a smaller coalition then that choice is rational for the grand coalition and it can be executed by the smaller coalition. This is a way of saying that to establish the socially permitted choices we need to look at what is rational for the grand coalition.

The third proposition says that the permission of φ or the permission of ψ implies the permission of φ or ψ. Its validity is a consequence of the monotonicity condition for the $(\forall, \forall)$ lifting, that makes undominated choice monotonic. A related and even stronger validity is $P^{C'}(C,\varphi) \rightarrow P^{C'}(C,\varphi \vee \psi)$, which is a version of the Ross paradox in our system [47].

The fourth item says that if φ is a possible choice for coalition C and $\neg\varphi$ is forbidden for C taking into account the interests of the grand coalition, then $\neg\varphi$ is also forbidden for any other coalition taking into account the interests of the grand coalition. The reason for this seemingly strong consequence of forbiddance is the fact that $\neg\varphi$ is an irrational choice for the grand coalition, no matter what coalition can make that choice.

The fifth validity sums up the point of view on regulation carried by the deontic operators: when acting in the interest of the grand coalition, the conflict between social rationality and coalitional rationality is always resolved favouring the first. In a strictly utilitarian view, it makes perfect sense to say that actions can be rational without being socially rational, i.e. $F^C(C,\varphi) \wedge \neg[rational_N]\varphi$. Our framework goes beyond this strict view, generalizing rationality to the interest of bigger coalitions.

The last validity we analyze concerns prohibition, which has a conjunctive property: if choosing φ is forbidden and choosing ψ is forbidden then choosing φ and ψ together is also forbidden. As already observed in the previous chapter, if two choices are dominated and their intersection is also an available choice, then the latter must also be dominated.

The propositions that are not valid in the models are also illuminating about the properties of the deontic operators.

The first invalidity is deontic agglomeration, that is the fact that two propositions that are obligatory implies that their conjunction is also obligatory. Deontic agglomeration together with the principle of "ought implies can" (obligation implies ability) rules out, in standard deontic logic, the existence of moral conflicts [44]. In our framework instead deontic operators (and later also violation constants) point to the regulation of moral conflicts.

The second invalidity shows that a permission of choice is not equivalent to a choice of permission. If the one side is due to monotonicity of Pareto optimality, the other side is falsified by the following consideration: permission requires by definition that a coalition is able to perform the permitted choice, but effectivity functions are not closed under finite intersections (Definition 7). Notice that the proposition instead holds for the empty coalition in strategic games (Definition 11).

The third invalidity shows that a coalition need not be obliged to perform a choice between a formula and its negation. This happens when both φ and $\neg\varphi$ are undominated dominated choices. Notice though that it cannot be the case that φ and $\neg\varphi$ are both dominated, as in the one case φ^M needs to be dominated by an $X \subseteq (\neg\varphi)^M$, and it cannot be the case that some $Y \subseteq \varphi^M$ dominates $(\neg\varphi)^M$.

The next invalidity says that the rational action for a certain coalition does not necessarily coincide with that of the grand coalition, which has already been discussed.

The last invalidity states the invalidity of the classical principle of "ought implies can". When a choice is obligated it does not mean that the coalition can perform it. This is due to the fact that obligation of a formula means forbiddance of its negation, i.e. irrationality of its negation. However a dual principle is available, "ought implies can refrain", as $O^N(C,\varphi) \to [C]\neg\varphi$.

Summing up, the deontic operators presented here indicate what a coalition should do in order to behave rationally. They are arguably simple reductions to operators that talk about coalitional ability and rationality. Their simplicity has several drawbacks however, the most important of which is discussed in the coming paragraph.

Policies Let us have a look again at the definition of the prohibition operator we have just defined:

$$F^N(C,\varphi) := [C]\varphi \to \neg[\mathit{rational}_N]\varphi$$

If we apply this operator to simple games like the prisoner's dilemma, we will notice that very intuitive deontic statements, such as $F(\mathit{Row}, \mathit{Row}$ defects), are false.

The reason is because the prohibition $F(Row, Row$ defects) is of absolute nature: it says that the set of all states Row defects^{PD} is not rational for the grand coalition. But we know from the results on undomination that any choice in the effectivity function of the grand coalition containing a Pareto optimal state is necessarily undominated. In conclusion $F(Row, Row$ defects) is *not* satisfied in a world w representing the prisoner's dilemma.

However the common feeling concerning the play for the grand coalition in a prisoner's dilemma warrants the defective choice by the prisoners to be forbidden. Can we deal with this problem within the language?

The answer is positive, but to express that prisoners are forbidden to defect a new operator should be introduced, of the form

$$F(C : \varphi, C' : \psi) := [C]\varphi \wedge [C']\psi \rightarrow \neg[rational_N](\varphi \wedge \psi)$$

whose informal reading is: "in the interest of the grand coalition, choices φ by C and ψ by C' are to be forbidden".

What this operator does is to lay down a policy to which coalitions need to comply in order to promote higher interests. The operator takes into account how the intersection of two coalitional choices even when coalitionally rational can become irrational for the grand coalition and it is clearly an abbreviation of formulas of the language. In our example we could say things like $F(Row$:defect, $Column$:defect), i.e. it is forbidden for Row to defect and for $Column$ to defect. Let us write down the definition in its most general form.

Definition 35 (Policies) *Let $C_1, \ldots, C_n$ be a partition of coalition $\bigcup\{C_1, \ldots, C_n\}$. The following operators define deontic statements in the interest of coalition $\bigcup\{C_1, \ldots, C_n\}$.*

$$\begin{aligned}
F(C_1 : \varphi_1, \ldots, C_n : \varphi_n) &:= ([C_1]\varphi_1 \wedge \ldots \wedge [C_n]\varphi_n) \rightarrow \neg[rational_{\bigcup\{C_1,\ldots,C_n\}}](\varphi_1 \wedge, \ldots, \wedge\varphi_n) \\
P(C_1 : \varphi_1, \ldots, C_n : \varphi_n) &:= \neg F(C_1 : \varphi_1, \ldots, C_n : \varphi_n) \\
O(C_1 : \varphi_1, \ldots, C_n : \varphi_n) &:= F(C_1 : \neg\varphi_1, \ldots, C_n : \neg\varphi_n)
\end{aligned}$$

The norms state a policy to be applied to the coalitions involved. For instance the prohibition operator $F(C_1 : \varphi_1, \ldots, C_n : \varphi_n)$ states that in the interest of the coalition formed by the union of $C_1, \ldots, C_n$, coalition C_1 is forbidden to choose φ_1, coalition C_2 to choose φ_2 and so on until coalition C_n, which is forbidden to choose φ_n. The reason for this prohibition lies in the fact that the intersection of the respective choices $\varphi_1 \wedge, \ldots, \wedge\varphi_n$ is not rational for those coalitions taken together. Permission and obligation follow the same pattern of the absolute norms.

These newly defined operators are in fact straightforward generalizations of the utilitarian norms in Definition 34, provided the coalitional effectivity function contains the unit.

Proposition 33 *Let $M = (W, E, \geq_i, V)$ be a Cooperative Game Model where E is an effectivity function containing the unit, and let $\varphi_i = \top$ for $i = 2, \ldots, n$. The following holds:*

$$M \models F(C_1 : \varphi_1, \ldots, C_n : \varphi_n) \leftrightarrow F^{\bigcup\{C_1,\ldots,C_n\}}(C, \varphi_1)$$

Proof *$M, w \models F^{\bigcup\{C_1,\ldots,C_n\}}(C, \varphi_1)$ is equivalent to $M, w \models [C]\varphi_1 \rightarrow \neg[rational_{\{C_1,\ldots,C_n\}}]\varphi_1$. But by the fact that the effectivity function contains the unit ensures that $[C]\top$ for each $C \subseteq N$. This means that $M, w \models [C]\varphi_1 \rightarrow \neg[rational_{\{C_1,\ldots,C_n\}}]\varphi_1$ is equivalent to$M, w \models F(C_1 : \varphi_1, \ldots, C_n : \varphi_n)$ for $\varphi_i = \top$ for $i = 2, \ldots, n$.*

Substantially the policy operators generalize the deontic operators in case the effectivity function of coalitions is not empty, which is a rather mild assumption.

Let us now return to the examples and analyze them.

Example 17 (Norms of cooperation and norms of conformity) *The games of Figure 2.2 are a striking example of conflict between potential coalitions. In particular the prisoner's dilemma rules out, in its classical account based on individual rationality, the possibility for individual players (in our account single player coalitions) to achieve the Pareto optimal outcomes. However these outcomes are to be achieved in the interest of both players taken together. Let us see how to express this fact in the language.*

In the model PD of the prisoner's dilemma at world w we have the following:

$PD, w \models [Row]$Row cooperates , i.e. the row player can see to it to play U, the cooperative move, $PD, w \models [rational_{Row,Column}]$Row cooperates , i.e. it is rational for both players that row player chooses U. This allows, confirming the validity discussed before, the conclusion that $PD, w \models P(Row, Row\ cooperates)$.

However, due to the formulation of the deontic operators no single player coalition is obligated to cooperate, nor forbidden to defect, as the move D by the row player remains undominated, and thus rational, in the effectivity function of the grand coalition.

A command like "Row is forbidden to defect" is, in our logic, of absolute nature. It says that the set of states corresponding to the defective move by Row is dominated in the effectivity function of the grand coalition. But in our case these are $\{(D, L), (D, R)\}$ and D, R being Pareto optimal, the set $\{(D, L), (D, R)\}$ is certainly undominated for the grand coalition.

In these cases the policy operators are of immediate use. We have that $F^N(Row : Rowdefects, Column : Columndefects)$, i.e. in the interest of coalition N Row should not defect and Column should not defect either.

The external view: a deontic logic for coalition formation

The internal view on norms was concerned with establishing what a coalition should do in order to act rationally. The present section takes a different perspective. Starting out from a labelling on outcomes that should be understood as *undesirable* we ask ourselves what coalitions would find it rational to make choices that are violation-free. In other words, we use the labelling on outcomes to understand which coalitions are allowed forming.

To express undesirable properties of outcomes we extend our language with a special atomic propositions *viol* to be interpreted in the following way:

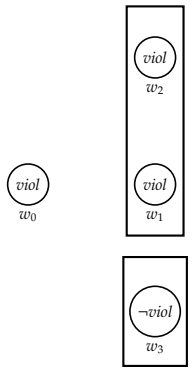

Figure 4.2: Violations and choices. The violation constant splits each model into two parts: the undesirable states, where *viol* holds, and the desirable ones, where ¬*viol* holds. As a consequence coalitional choices naturally inherit a deontic evaluation.

$$M, w \models viol \quad \text{if and only if} \quad w \in viol^M$$

Once a violation constant is introduced in the language a natural question is whether coalitions would rationally choose the desirable states. This observation allows to lift the deontic operators to coalitions, i.e. to determine what coalitions are to be forbidden, permitted, obligated if the desirable states are to be achieved.

$$\begin{aligned} F(C) &:= F^C(C, \neg viol) \\ P(C) &:= \neg F(C) \\ O(C) &:= \bigwedge_{C' \neq C} F(C') \end{aligned}$$

A coalition C is forbidden, formally $F(C)$, when the ideal states, i.e. the set of all states where *viol* does not hold, are not a rational choice for that coalition; it is permitted, formally $P(C)$, when the ideal states are a rational choice for that coalition; it is obligated, formally $O(C)$, when all the other coalitions are forbidden.

The introduction of the violation constant, labelling the undesirable states, has allowed us to lift the classical deontic operators to coalitions. In some sense this approach generalizes the utilitarian one, as violations can be introduced that are rational choices for a certain coalition. Consider therefore the following formula:

$$\neg[rational_C]\varphi \rightarrow A(\varphi \rightarrow viol)$$

When valid in a model this formula means that all irrational choices of coalition C are choices leading to violation. Consider instead this one:

$$[rational_C]\varphi \wedge A(\varphi \rightarrow viol)$$

When valid in a state this formula says that there is a rational choice for coalition C that always leads to violation. This, notice, implies that coalition C must be forbidden.

The following example gives a flavour of what is possible to do by using the violation constant in combination with the choices that are rational for certain coalitions.

Example 18 (Forbidding coalitions) *Let us make use of our violation constant in the prisoner's dilemma (whose model is denoted PD) and the coordination game (CG). In the first case we would like to forbid the states where both players defect, namely* $viol^{PD} = \{(D, R)\}$, *and in the other the states where players do not coordinate, namely* $viol^{CG} = \{(U, R), (D, L)\}$. *Let* w *be a world in both models to be used as evaluation point. We have that* $PD, w \models F(\{Row\}) \wedge F(\{Column\})$, *i.e. individual coalitions are forbidden, as they cannot force the set of ideal states, while* $PD, w \models O(\{Row, Column\}) \wedge P(\{Row, Column\})$ *as the only coalition that is not forbidden is the one made by both players. The same propositions hold true in* CG, w, *for the way we have used the violation constant.*

Stocktaking

The utilitarian approach uses the notion of coalitional rationality as semantic underpinning for the deontic operators, while what we have called the systemic approach combines the operator for rational choices with violation constants, to label the undesirable states independently of the preferences of the players. If the first approach is more concrete and directly applicable to strategic games by means of the deontic policy operators, using violation constants in combination with the coalitional rationality operator allows for an elegant lifting of norms to coalitions.

4.4 Reasoning on Coalitions in Games

The previous section has dealt with a characterization of coalitional rationality and the issues of its regulation for general coalitional games. This section will focus on strategic games and will provide a logic to characterize the coalitional power in those structures. The section will thus not be concerned with a theory of coalitional rationality, as its traits have been previously provided, but with characterizing the distinguishing properties of strategic games in a logical language.

More specifically, in this section, we investigate the impact of true playability on logics of coalitional ability. We begin by indicating that the validities of Coalition Logic do not change if we restrict models to truly playable. As a consequence, Coalition Logic (and even ATL) cannot distinguish between models based on truly playable effectivity functions and models based on playable effectivity functions.

4.4.1 True playability and Coalition Logic

The previous chapter analyzed the specific features of coalitional ability in strategic games, providing an alternative representation result to the one originally given in [54]. We can immediately observe that this new relation between effectivity functions and strategic games has no repercussions on the semantics of Coalition Logic and the soundness/completeness results for that logic. The axiomatization of playable Coalition Logic amounts in [54] to the formulas and the rules characterizing playability (Proposition 5), the axioms of propositional logic and modus ponens. In [54] it is proved how this axiomatization is sound and complete with respect to playable coalition models. The following result can be carried over.

Corollary 34 *The axiomatization of playable Coalition Logic from [54] is sound and complete wrt truly playable coalition models (and hence also strategic game models).*

Proof *To see this, let us formally define* Play *to be the class of playable coalitional models, and* TrulyPlay *as the class of models based on truly playable effectivity functions. Since* TrulyPlay $\subset$ Play*, every Coalition Logic formula valid in* Play *is valid in* TrulyPlay*, too. To see the converse, one can use the finite model property of Coalition Logic with respect to* Play *and the fact that it coincides with* TrulyPlay *on finite models.*

The results show that Coalition Logic [3] describes strategic interaction at an extremely abstract level, that is insufficient to distinguish playability from true playability. In the next section we extend the language to make this distinction possible.

4.4.2 Coalition Logic with *outcome selector* modality

Here we propose an extension of Coalition Logic, by adding a new normal modality $\langle O \rangle$ called "outcome selector". Its dual will be denoted $[O]$. The informal reading of $\langle O \rangle \varphi$ should be "there is an outcome state, enforceable by the grand coalition and satisfying φ". Instead of defining the semantics of $\langle O \rangle$ in the straightforward way (by an appropriate semantic clause), we choose a different path in order to characterize the new language axiomatically, and thus provide an axiomatic characterization of truly playable models. That is, we first expand coalition models to what we call *extended coalition models* with an additional "outcome enforceability" relation R. Later we will use axioms to impose the right behavior of R.

Definition 36 (Extended coalition frames) *An* extended (playable) coalition frame *is a neighbourhood frame $F = (W, E, R)$ where W is a set of outcomes, E a playable effectivity function, R a binary relation on W.*

[3] Furthermore, the semantics based on effectivity functions can be extended to ATL. See [29] and [54] for the fragment of ATL without "until", called *Extended Coalition Logic*). Again, it can be shown that Play and TrulyPlay determine the same sets of validities for ATL, by checking the soundness of the axiomatization for ATL given in [32] for Play, and using the completeness result for ATL with respect to strategic game models (equivalently, TrulyPlay) proved in the same paper.

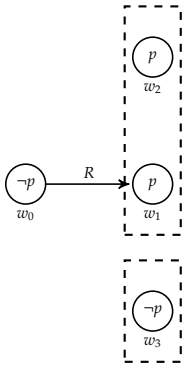

Figure 4.3: Extended Coalition Models. The relation R is not dependent on the dynamic effectivity function. There can be outcomes that are reachable from w_0 via R but that are not available choices at w_0.

An extended coalition model — an illustration is given in Figure 4.3 — is an extended coalition frame endowed with a valuation function. Given an extended coalition model M, the modality $\langle O \rangle$ is interpreted as follows.

$$M, w \models \langle O \rangle \varphi \Leftrightarrow wRs \text{ and } M, s \models \varphi \text{ for some } s \in W$$

That is, $\langle O \rangle$ has standard Kripke semantics with respect to the outcome enforceability relation R.

Note that extended coalition models do not require any interaction between the effectivity function and the relation R. However, given the intuitive reading of the relation R, the interaction suggests itself, and the following definition accounts for that.

Definition 37 (Standard coalition frames) *A* standard coalition frame *is an extended coalition frame such that, for all $w, v \in W$, we have wRv if and only if $\{v\} \in E(w)(N)$.*

A standard coalition model, depicted in Figure 4.4, is a standard coalition frame with a valuation function. Depending on the properties of the underlying effectivity functions we call extended coalition frames and models playable or truly playable.

Characterizing standard truly playable coalition frames

Proposition 35 *An extended coalition frame F is standard and truly playable if and only if $F \models [N]\varphi \leftrightarrow \langle O \rangle \varphi$.*

Proof *Left to right: Assume that F is standard and truly playable. Assume first that $(F, V), w \models [N]\varphi$ for any V and $w \in W$. By definition of E we have that $\varphi^M \in E(w)(N)$. As F is truly playable there is $v \in \varphi^M$ with $\{v\} \in E(w)(N)$. However F is also standard so wRv. But this means that $(F, V), w \models \langle O \rangle \varphi$. Conversely, if $(F, V), w \models \langle O \rangle \varphi$ then wRv for*

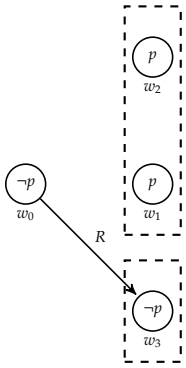

Figure 4.4: Standard Coalition Models. The relation R is now dependent on the dynamic effectivity function: each reachable outcome is an available choice and each single outcome choice is also reachable.

some $v \in \varphi^M$. F being standard we have that $\{v\} \in E(w)(N)$. By outcome monotonicity $\varphi^M \in E(w)(N)$, i.e. $(F, V), w \models [N]\varphi$.

Right to left: Assume that $F \models [N]\varphi \leftrightarrow \langle O\rangle\varphi$. Let us first prove that F is standard. Suppose wRv for some $w, v \in W$. Let V be a valuation that assigns the proposition φ only to v. We have that $M, w \models \langle O\rangle\varphi$. Then, by the assumptions we also have $M, w \models [N]\varphi$, which means that $\{v\} \in E(w)(N)$. Conversely, suppose now that $\{v\} \in E(w)(N)$. For the same valuation V we must have that $M, w \models [N]\varphi$ and by assumption that $\langle O\rangle\varphi$, which means that wRv. Thus, F is standard. To prove that F is truly playable, assume that for some $X \subseteq W$, $X \in E(w)(N)$ and let now V be a valuation function such that $\varphi^{F,V} = X$. By definition of E we have that $(F, V), w \models [N]\varphi$, hence by assumption, that $(F, V), w \models \langle O\rangle\varphi$, which means that wRv for some $v \in \varphi^{F,V}$. Then, F being standard, $\{v\} \in E(w)(N)$.

Axiomatizing standard truly playable models

We propose the following axiomatic system for the class of standard truly playable coalition models TrulyPlay, extending Pauly's axiomatization of CL. The axioms include propositional tautologies plus the following schemes:

1. $[N]\top$
2. $\neg[C]\bot$
3. $\neg[\emptyset]\varphi \rightarrow [N]\neg\varphi$
4. $[C]\varphi \wedge [C']\psi \rightarrow [C \cup C'](\varphi \wedge \psi)$ for any disjoint $C, C' \subseteq N$
5. $[N]\varphi \leftrightarrow \langle O\rangle\varphi$
6. $[O](\varphi \rightarrow \psi) \rightarrow ([O]\varphi \rightarrow [O]\psi)$.

The inference rules are: Modus Ponens, plus:

$$\frac{\varphi \to \psi}{[C]\varphi \to [C]\psi}, \text{ and } \frac{\varphi}{[O]\varphi}.$$

Notice that the modalities $[C]$ are monotone but not normal while the modality $[O]$ is normal. [4] We denote the logic axiomatized as above by TPCL. The following is routine.

Proposition 36 *TPCL is sound for the class TrulyPlay: every formula derivable in TPCL is valid in TrulyPlay.*

Now we will establish the following completeness result:

Theorem 37 (Completeness theorem) *Every formula consistent in TPCL is satisfiable in TrulyPlay. Consequently, the logic TPCL is complete for the class TrulyPlay.*

The full completeness proof is in given in the appendix (Section B.3).

Summing up, the present section has established a correspondence between formulas of Truly Playable Coalition Logic, an extension of Coalition Logic with what we have called the *outcome selector* modality, and truly playable models, that we have proved in the previous chapter to be exactly corresponding to strategic games. Moreover we have shown that the playable fragment of Coalition Logic, rather surprisingly, also corresponds to truly playable models. This apparent paradox can be easily explained: the playable fragment of Coalition Logic is simply too abstract to distinguish between truly playable and playable models, while Truly Playable Coalition Logic is enhanced with enough extra expressive power (the outcome selector modality) to make this distinction possible.

4.5 Discussion

4.5.1 Related work

A number of contributions exist in the literature touching upon the issues dealt with in this chapter, and some of them have already been taken into account as to their specific contributions. They can be divided into three main branches, namely the works aimed at:

- providing a logical account of strategic reasoning, mainly within the logic and multi-agent system community; eminent examples are the use of counterfactuals in CATL [68], the action expressions used in Coalition Action Logic [16] and the first order strategy terms in Strategy Logic [22] and the explicit strategies in [74]. Even though not directly focused on strategic interaction, the contribution in [66] makes an important step in the direction of formalizing preference liftings for strategic decisions.

[4] Despite the equivalence in axiom 5, the introduction of the new outcome modality is justified because we only want to consider frames where $[N]$ is a diamond operator of a *normal* modality, whereas the semantics of all coalition operators $[C]$ is given in terms of neighbourhood effectivity functions.

- understanding the deontic operators in terms of properties of strategic interaction; the only related contributions we consider, and up to our knowledge the only existing ones, are by Horty [41] and Kooi and Tamminga [44].
- building a logic for effectivity functions, and their generalizations. The contributions on logics for strategic interaction are plenty, starting from the seminal contribution by Rohit Parikh on the logic of games and its applications [52], however the ones using effectivity functions all originate from Marc Pauly's Coalition Logic [54, 55].

In this section we are going to focus on the approach by van der Hoek, Jamroga and Wooldridge [68] extended recently in [74], which discusses modal operators that make strategies explicit, in a similar way done by our subgame operator; and on the semantics for the deontic operators given in [41] and in [44], comparing them with our contribution.

Logics for strategic reasoning The logic Alternating-time Temporal Logic with Counterfactuals (CATL) by van der Hoek, Jamroga and Wooldridge [68], extended ATL with commitment operators of the form $C_a(\rho, \varphi)$, meaning "under the assumption that a commits to strategy ρ then φ holds". The similarity of the CATL operator and the subgame operator is evident, however a number of differences can be immediately spotted looking at the interpretation of the CATL operator.

Definition 38 (CATL models) *[68]*
The CATL operator is interpreted in models that are called Action-based Alternating Transition Systems, formally a tuple:

$$(Q, q_0, \Phi, \pi, Ag, Ac_1, \ldots, Ac_n, \rho, \tau, \mathcal{Y}_1, \ldots, \mathcal{Y}_n, \|.\|_M)$$

where Q is a nonempty set of states; $q_0 \in Q$ is the initial state; Φ is a nonempty finite set of atomic propositions; π is a valuation funciton; Ag is a finite set of players such that $|Ag| = n$ and such that each player i is assigned a unique set Ac_i of actions (the sets are pairwise disjoint); $\rho : \bigtimes_i Ac_i \to 2^Q$ is an action precondition function, specifying at which states a certain action profile can be executed; $\tau : Q \to J \to Q$, where $J \subseteq \bigtimes_i Ac_i$ is the system transition function, that specifies what the effect of the action profiles is, respecting the action precondition function. $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ are sets of so-called strategy terms for each player, also pairwise disjoint. In each model M each strategy Σ_i by player i will be denoted by the function $\|.\|_M$.

Leaving technicalities aside it is immediately evident that CATL models resemble strategic game forms, together with a mechanism to indicate for which model what strategies each player can execute. The commitment operator $C_a(\rho, \varphi)$ is interpreted by means of a model update, as follows.

Definition 39 (Commitment interpretation) *[68]*
Let M be a CATL model having Q as domain and let $q \in Q$. The formula $C_i(\sigma, \varphi)$, for σ a strategy of player i and φ a formula of the language of CATL, is interpreted as follows:

$$M, q \models C_i(\sigma, \varphi) \textit{ if and only if } M \dagger_i \|\sigma\|, q \models \varphi$$

The operation $M \dagger_i \|\sigma\|$ on M returns a new model that is made as follows:

$$(Q, q_0, \Phi, \pi, Ag, Ac_1, \ldots, Ac_n, \rho', \tau', \sim_1, \ldots, \sim_n, \|.\|_{M'})$$

where the new elements $\rho', \tau', \|.\|_{M'}$ are updated, taking into account the fact that strategy σ has been executed by player i and the remaining elements are literally copied from M.

The CATL logic, extended later in [74], was the first to deal with the issue of choice restriction in strategic reasoning and implemented it quite effectively by updating the models. However there are two issues left unanswered by CATL that we tried to deal with by means of the subgame operator:

- In CATL, strategy terms, as for example the occurrence of σ in formulas of the form $C_i(\sigma, \varphi)$, are not formulas of the language, but they act as functions from a model M to its updated version $M \dagger_i \|\sigma\|$. Contrarily, in the subgame operator, the analogue of strategies are formulas, as in expressions of the form $[C \downarrow \varphi]\psi$. The main advantage of this method is the possibility to relate strategy execution to strategic ability, in our case by means of reduction axioms.

- If in Coalition Logic expressions of the form $[\{i\}]\varphi$ are used to express the fact that a coalition $\{i\}$ *has a strategy* σ_i to achieve φ, CATL expressions make σ_i explicit, by saying for instance $C_i(\sigma, \varphi)$. In a way, from being able to express the presence of a strategy, we are now able to name that strategy in the language and study the transformations it brings about at the model level. However [5] Coalition Logic is also able to express that a coalition *does not have a strategy* to achieve φ, by simply saying $\neg[C]\varphi$; this seems particularly awkward to do only using expressions of the form $C_i(\sigma, \varphi)$ as $\neg C_i(\sigma, \varphi)$ only means that strategy σ does not achieve φ. With the subgame operator though $\neg[C \downarrow \psi]\varphi$ is directly reducible to an expression without the subgame operator occurring in it, preserving the possibility of expressing the various meanings of strategic ability.

All in all, CATL and its extensions represent a fundamental starting point for the investigation of logics for strategic reasoning and have partly inspired the construction of the subgame operator, that, even though far less structured, looks more suited to study the relation between strategic ability and strategy execution.

[5] As far we can recollect, the argument has been first used by Johan van Benthem in his invited talk "In praise of strategies, but how?", during the "First Workshop on Logics and Strategies", June 26, 2009 at the University of Groningen.

Norms and coalitions In the section of related work of the previous chapter, we have dealt with the notion of dominance in consequentialist models, taken as *simplified representation* of the branching-time STIT framework. In consequentialist models the ought operator by Horty [41] and Kooi and Tamminga [44] can be defined. While the first deals with a strictly utilitarian view of oughts, i.e. a coalition ought to do what it is optimal for itself, the second generalizes this view to take into account the preferences of other coalitions, i.e. what a coalition ought to do needs to be optimal for a coalition specified in the modal operator.

We believe that interpreting both operators on consequentialist models will not betray Horty's views, even though we are aware that more subtle meanings carried by the ought operator can only be expressed using the full-blown STIT apparatus.

Let us interpret Horty's ought operator $\odot[C \text{ cstit } : \varphi]$, with the intuitive meaning that coalition C ought to see to it that formula φ holds, using the notion of dominance in Definition 30.

Definition 40 (Horty's ought) *Let M be a consequentialist model and w a world in its domain. Let $Choice(C)_w$ be the assignment of the set $Choice(C)$ at world w according to Definition 29. We have*

$M, w \models \odot[C \text{ cstit } : \varphi]$ if and only if for each $K \in Choice(C)_w$ such that $K \not\subseteq \varphi^M$ there is an action $K' \in Choice(C)_w$ such that

- *K dominates K'*
- *$K' \subseteq \varphi^M$*
- *$K'' \subseteq \varphi^M$ implies that K' dominates K''*

Horty's notion of dominance is somewhat stronger than ours, as it is not based on a $(\forall, \forall)$ preference lifting. It can then be the case that in a choice set of a consequentialist model there are no undominated choices, even when this choice set is nonempty. For this reason Horty's ought consists of three clauses introduced by a $\forall - \exists$ quantification on the sets in a choice set. This makes it at the same time particularly expressive and particularly awkward to characterize in terms of simple modal expressions, as we did for our deontic operators.

Being concerned with expressing moral conflicts, Kooi and Tamminga generalize Horty's operator, by generalizing the underlying notion of dominance.

Definition 41 (F-dominance) *[44]*

Let $K, K' \in Choice(C)$ and $u_i : N \to W \to [-5, 5]$ a utility function over the outcomes for each player. We say that K F-dominates K' if and only if for all $S \in Choice(\overline{C})$ we have that $w \in K \cap S$ and $w' \in K' \cap S$ implies that $u_F(w) \geq u_F(w')$, where u_C returns the average of individual utilities of members of F.

The definition of their ought operator $\odot^F[C \text{ cstit } : \varphi]$ is obtained by replacing the word *dominates* by the word *F-dominates* in Definition 40.

The notion of F-dominance can be though of as a sort of *dominance for someone else*, i.e. a dominance looked at from the point of view of another coalition. We will

see that such generalizations get very close to the theory of dependence relations dealt with in the second part of the thesis. But for now we observe how Kooi and Tamminga's approach inherits the strength (but also the complexity) of Horty's ought, by adding the issue of evaluating a coalitional choice from the point of view of a different coalition. We have seen how in our case an operator of the form $F^N(C, \varphi)$, where the prohibition is made in the interest of a bigger coalition, is an absolute command, and as such does not propose a desirable solution for even simple games. Kooi and Tamminga deal with the prisoner's dilemma by showing formulas of the form $\odot^{\{Row,Column\}}[Row \text{ cstit } : \text{Row cooperates}] \wedge \odot^{\{Row\}}[Row \text{ cstit } : \text{Row defects}]$ (a notational variant of these formulas also holds with our absolute commands), however they do not focus on the operators of forbiddance and permission, making our remark on absolute commands not strictly applicable. Moreover, the scope of their work being other than analyzing the effects of simultaneous norms, it does not discuss policies to regulate games.

4.5.2 Open issues

The present chapter has left several issues unresolved. We prefer to focus on what we think are the two most fundamental ones, that concern both the characterization of strategic reasoning and the role of norms. As to the first point, we deal with the relation between the role of updates in the subgame operator and the well-known model update in Dynamic Epistemic Logics [69]. As to the second point we deal with the relation between the role of norms in our framework and how they can account for a multiplicity of other interesting related notions, such as social choice correspondences.

Choices as announcements Public Announcement Logic [69] formalizes the effect of the announcement of a true formula in each agent a's epistemic relation $R(a)$, defined as a partition on a domain W. The standard operator $[\varphi]\psi$ says that ψ holds after φ is announced. Its semantics is given as follows:

$$M, w \models [\varphi]\psi \Leftrightarrow M, w \models \varphi \text{ implies } M|\varphi, w \models \psi$$

where $M|\varphi = (W', R'(a), V')$ takes these values:

- $W' = \varphi^M$
- $R'(a) = R(a) \cap (W \times \varphi^M)$
- $V'(p) = V(p) \cap \varphi^M$

The model restriction of public announcement *eliminates worlds*. At the logical level a reduction can be shown such that every sentence from the modal language with the modal operator interpreted on the epistemic relation and the public announcement operator can be translated into a sentence from the same language without the public announcement operator occurring in it. We report the reduction axioms in Table 3.

	Axioms
	Public Announcement Axioms
A1	$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$
A2	$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$
A3	$[\varphi](\xi \wedge \psi) \leftrightarrow ([\varphi]\xi \wedge [\varphi]\psi)$
A4	$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$
	Rules
R1	$\xi \wedge (\xi \rightarrow \psi) \Rightarrow \psi$
R2	$\xi \Rightarrow [\varphi]\xi$

Table 4.2: Proof System for Public Announcement Logic

If we compare the public announcement operator to the subgame operator, we can observe the structure of the two axiom systems is very similar in the atomic and boolean case, but very different in the modal case. A subtle difference can however be observed in the atomic clause. If Public Announcement Logic reduces the atomic announcement to an implication between atoms ($[q]p \leftrightarrow (q \rightarrow p)$), the subgame operator reduces it to an implication between an atom and a choice ($[C \downarrow q]p \leftrightarrow ([C]q \rightarrow p)$). This fact witnesses that we are really reducing strategy execution to strategic ability.

The appendix will make it clear (Section B.1) that the similarity of the logics applies to the proof techniques as well, that are at least for the basic cases identical to those of Public Announcement Logic [69]. The specific differences are given, once again, by the way the coalitional relation is updated.

Majority Voting Social choice theory, preference aggregation and judgment aggregation — for their interrelation see the work in [33] — analyze various paradoxes of preference merging, like the paradox of majority voting. This says that a majority can decide for an issue p, a majority for $p \rightarrow q$, and another majority for not q, yielding *illogical* policies. Our deontic operator can be extended to satisfy a majority optimality.

Definition 42 (Majority Domination) *Given an effectivity function E, X is* majority undominated *for C in w if, and only if, (i) $X \in E(w)(C)$ and (ii) for all $Y \in E(w)(\overline{C})$, $(X \cap Y)$ is Pareto Optimal in $E(w)(C) \sqcap Y$ for a $C' \subseteq C$ such that $|C'| \geq \frac{|C|}{2} + 1$.*

We can define a majority deontic operator $M, w \models P^{\frac{1}{2}}(C, \varphi) \Leftrightarrow \varphi^M \in E(w)(C)$ and *is majority undominated* for C in w. Majority Voting Paradox arises in this framework:

$$M \not\models P^{\frac{1}{2}}(C, \varphi) \wedge P^{\frac{1}{2}}(C, \varphi \rightarrow \psi) \rightarrow P^{\frac{1}{2}}(C, \psi)$$

Here the obligation for a coalition to do φ and $\varphi \rightarrow \psi$ need not mean the obligation to do ψ. In general it would be extremely profitable for the applicability of such deontic languages to systematically relate to classical results in Social Choice Theory, as for instance Arrow's results on Social Welfare Functions [4].

There are typical conditions that such a function should have in order to be desirable:

Unanimity (or Pareto Optimality): If all individuals weakly prefer an alternative x over an alternative y, then x should be weakly preferred over y in the final decision.

Independence of Irrelevant Alternatives : Given two preference orderings $\preceq, \preceq'$ if all individuals prefer x over y according to $\preceq$ if and only if x is preferred to y in the final decision according to $\preceq$ then all individuals prefer x over y according to $\preceq'$ if and only if x is preferred to y in the final decision according to $\preceq'$.

Universal Domain : every alternative should be ranked.

Non-Dictatorship : There is no agent i such that, for all preference orderings $\preceq$, if i prefers x over y according to $\preceq$, then x is preferred over y in the final decision.

We can clearly observe some resemblance with properties satisfied by our operator $[rational_N]$, but again, systematic treatment is required.

4.5.3 Conclusion

The contribution of this chapter consists in the formalization of various aspects of cooperative interaction and its regulation. The whole enterprise has been conducted by making use of a standard logic for coalitional ability, Coalition Logic, empowered with modal operators to meet the specific needs. First, Coalition Logic has been empowered with a preference modality, to characterize notions of optimality inside an effectivity function; second, it has been empowered with the subgame operator, to reason about choice restriction; third with violation constants, to reason about regulation of coalitional choice and finally with modalities to characterize the specificity of strategic games as opposed to general coalitional games.

As for the logical characterization of strategic reasoning, the following results have been achieved:

- Logical characterization of betterness within an effectivity function, by means of a coalition and preference modality in some cases with an auxiliary global modality. The so-called Pareto optimal choices have been characterized in their strong and weak version, for all preference liftings studied in the previous

chapter. We could show that for the $(\forall, \forall)$ preference lifting, Pareto optimal choices could be expressed even without resorting to the global modality.

- Logical characterization of choice restriction, by means of an operator that could express choice execution by a coalition. This has been called the subgame operator, its role being to talk of formulas that would hold if some coalitions were to make a certain choice. The subgame operator was shown to behave rather elegantly by displaying reduction axioms to the standard coalition logic operator, yet the price to pay is that models of coalitional interaction need to contain all the instances of the subgames.

As for the regulation of coalitional interaction, the chapter has described deontic languages to tell coalitions how to behave to achieve certain desirable outcomes. Two views have been proposed:

- The *internal*, or utilitarian view, where, in order to decide whether a coalitional choice is to be obliged, prohibited or permitted, the interest of bigger coalitions need to be taken into account. Coalitional choices that are permitted for coalition C in the interest of some $C' \supseteq C$ are rational choice for coalition C', while forbiddance and obligation are defined following a standard duality in deontic logic. Refining on this we have defined the notion of policy, a simultaneous deontic command on some coalitions in order to protect the interest of those coalitions taken together. With the notion of policy, the standard reasoning in games like the prisoner's dilemma can be fully accounted for.

- The *external*, or systemic view, where desirable properties to be achieved by coalitional choice do not depend on coalitional interests. States in an interaction are split between good states and bad states, in the spirit of classical deontic logic. The labelling in a strategic context acquires a natural meaning, allowing the lifting the deontic operators to coalitions. The idea behind this lifting is that coalitions for which all rational choices are violations should not be allowed, and should instead be permitted or even obligated otherwise.

As for the logical characterization of strategic games, plenty of issues raised by the previous chapter have been answered.

- We have pointed out that Coalition Logic and ATL are not expressive enough to characterize true playability. On the other hand, they can be extended in a relatively simple way to obtain such a characterization.

- We have studied an extension of Coalition Logic with a normal *outcome selector* modality that we show sufficient for axiomatic characterization of truly playable structures.

All in all, the results have shown that a number of complex concepts in strategic interaction can be reconducted to the realm of Coalition Logic and reasoned upon with minimal extensions of that language.

Part II

Strategic Reasoning and Dependence Games

Chapter 5

Dependence Games

> *To* understand *an idea or a phenomenon — or even something like a piece of music — is to relate it to familiar ideas or experiences, to fit it into a framework in which one feels* at home.
>
> Robert J. Aumann, *What is game theory trying to accomplish?* [6]

5.1 Introduction

Traditionally, the cooperative possibilities of players in strategic games are described by merging their strategic ability and forming coalitions. This mathematically simple step, that takes the union of the players and the pairwise intersection of their choice sets, implicitly assumes that players are able and willing to join forces to achieve a common goal. However such assumptions are not applicable to a large number of scenarios, for a variety of reasons:

- Players may not be able to communicate with each other, as in the classical story of the prisoner's dilemma in Chapter 1, making coordination impossible;
- Players may not wish to form a coalition with other players, due to differences in desires, views, preferences etc.;
- Players may not have control of the process of coalitional decision making and may not trust the way the coalitional outcomes are chosen;
- The procedures employed to come up with a coalitional decision may impose additional costs on the participants (time, computational power etc.).

This chapter is devoted to a weakening of the classical view of strategic games as coalitional games, by keeping the individual perspective typical of strategic games and constraining the possibility of players to work together. We start from the observation that in some games players can do something for each other or, otherwise said, *depend on each other*. If we look at the prisoner's dilemma in Figure

2.2, for example, it is evident that the row player can do something for the column player (playing U instead of D) and that the column player can do something for the row player (playing L instead of R). As soon as players are aware of this possibility, they can *exchange favours* and reach an outcome that is satisfactory for both.

The perspective that we have just intuitively introduced radically differs from the classical coalitional account, where coalitions can be formed without considering players' preferences. Our account, that sees coalitions as resulting for reciprocal exchanging of favours among their members, will be called *dependence theory*.

The importance of dependence in multi-agent systems was not recognized until the publication of a series of papers by Castelfranchi and colleagues [21, 20], who elevated it to a paradigm to understand social interaction. Their work emphasized the necessity of building a formal theory of dependence modelling the role that cognitive phenomena such as beliefs and goals play in its definition. In the last decade, the notion of dependence has made its way into several research lines (e.g., [60, 12, 13, 59]), but still today dependence theory has several versions and no unified theory. However, the aim of the theory is clear:

> "One of the fundamental notions of social interaction is the *dependence* relation among players. In our opinion, the terminology for describing interaction in a multi-player world is necessarily based on an analytic description of this relation. Starting from such a terminology, it is possible to devise a calculus to obtain predictions and make choices that simulate human behavior" [21, p. 2].

In this view, dependence theory addresses two main issues:

- the representation of dependence relations among the players in a system;
- the use of such information as a means to obtain predictions about the behavior of the system.

While all contributions to dependence theory have thus far focused on the first point, the second challenge, "[to] devise a calculus to obtain predictions", has been mainly addressed by means of computer simulation methods (e.g., [59]) and no analytical approaches have yet been developed. We take up these two challenges from an analytical point of view and outline a theory of dependence based on standard game-theoretical notions and techniques.

The theory moves from the following definition of dependence, which is adapted from one of the definitions that can be found in the informal literature on dependence theory in multi-agent systems (e.g., [21, 20]):

> *Player i depends on player j for strategy σ_j, within a given game, if and only if σ_j is a dominant strategy (or a best response in some profile σ) not for j itself, but instead for i.*[1]

[1] This definition will be formalized in Definition 44.

The aim of the chapter is to provide a thorough analysis of the above informal definition. Concretely, it presents two results. First, it shows that dependence allows for the characterization of an original notion of reciprocity for strategic games (Theorem 40). Second, it shows that dependence can be fruitfully applied to ground cooperative solution concepts. These solution concepts are characterizable as the core of a specific class of coalitional games—here called *dependence games*—where coalitions can force outcomes only in the presence of reciprocity (Theorems 41 and 42).

Our study is meant to lay a bridge between game theory and dependence theory that, within the multi-agent systems community, are erroneously considered to be alternative, when not incompatible, paradigms for the analysis of social interaction.[2] It is our conviction that the theory of games and that of dependence are highly compatible endeavours. On the one hand dependence theory can be incorporated into the highly developed mathematical framework of game theory, obtaining the sort of mathematical foundations that are still missing. On the other hand, game theory can fruitfully incorporate a novel dependence-theoretic perspective on the analysis of strategic interaction.

With respect to this latter point, the chapter shows that dependence theory can play a precise role in games by modeling a specific way in which cooperation arises within strategic situations (Section 5.3):

> "As soon as there is a possibility of choosing with whom to establish parallel interests, this becomes a case of choosing an ally. When alliances are formed, it is to be expected that some kind of mutual understanding between the two players involved will be necessary. [...] One can also state it this way: A parallelism of interests makes a cooperation desirable, and therefore will probably lead to an agreement between the players involved." [70, p. 221]

Once this intuitive notion of "parallelism of interests" is taken to mean "mutual dependence" [21] or "dependence cycle" [60] the bridge is laid and the sort of cooperation that arises from it can be fruitfully analyzed in dependence-theoretic terms. This intuition, we will see, leads to the definition of a particular class of cooperative games.

Chapter structure: The chapter, based on joint work with Davide Grossi [35, 34], is structured as follows. Section 5.2 provides a formal definition of dependence relation between players in a strategic game, obtained by generalizing the classical ones of best response and dominant strategy. With a definition of dependence at hand, we can literally *draw* dependence relations among players and *spot* dependence cycles, i.e. strategies that can be played in favour of each other. Section 5.3 provides the solution for dependence cycles, by formulating the notion of agreement, seen as a strategy profile where each player favours some other player. Agreements

[2] An impression that was recently reiterated during the AAMAS'2009 panel discussion "Theoretical Foundations for Agents and MAS: Is game theory sufficient?".

allow us to view strategic games as dependence games, the class of cooperative games where coalitional choices are determined by agreements. Section 5.3 studies in addition how the solution concepts for cooperative games behave in dependence games. Finally, Section 5.3.5 shows an application to games with transferable utilities, making use of a theory of dependence in that setting. As for the other chapters, special attention is devoted to related literature and issues left open.

5.2 Dependence in Games

Dependence theory, as developed within artificial intelligence and multi-agent systems, has been mainly inspired by work in the social sciences such as [24]. It moves from presuppositions that are clearly shared by the theory of games—eminently the fact that the outcome of social interaction depends on the choices of different agents—but it emphasizes, rather than the strategic aspect of agents' choices, the interdependencies existing between them in terms of what they want and what they choose:

> "Sociality obviously presupposes two or more players in a common shared world. A 'Common World' implies that there is interference among the actions and goals of the players: the effects of the action of one player are relevant for the goals of another: i.e., they either favour the achievement or maintenance of some goals of the other's (positive interference), or threat some of them (negative interference)." [20, p. 161-162]

In this view, what underpins the analysis of social interaction is the idea that agents can favour or hinder each other's goals.

The present section shows how, by tweaking some basic game-theoretical notions, this perspective on social interaction can be accommodated within the theory of games. The section proceeds with a formal definition and analysis of the notion of dependence in games.

This section introduces and studies the formal properties of dependence in strategic games, obtained by generalizing the classical notions of dominant strategy and best response given in Definition 3.

5.2.1 Dependence relations

The literature on dependence theory features a number of different relations of dependence. Yet, in its most essential form, a dependence relation is a relation occurring between two players i and j with respect to a certain state (or goal) which i wants to achieve but which it cannot achieve without some appropriate action of j.

> "x depends on y with regard to an act useful for realizing a state p when p is a goal of x's and x is unable to realize p while y is able to do so." [21, p.4]

The definition in [21] acquires a variety of meanings applicable to different contexts. We start out by emphasizing the strategic aspects of dependence by reformulating its definition as follows:

> *A player i depends on a player j for the strategy σ_j when σ_j is a favour by j to i, that is, the choice by j of σ_j is in i's interest.*

Let us compare this reading with the one given by Castelfranchi in [21, p.4], which will make clear what the assumptions are upon which the theory will be developed. In fact, it might be argued that the focus of Castelfranchi's formulation seems to differ slightly from ours in a few points. Such differences, we claim, are not essential and do not lie at the core of the notion:

- Castelfranchi stresses the fact that one of the actors is not able to realize the goal which he is dependent for ("[...] while y is able to do so." [21, p.4]). An attentive reading of our formulation will reveal that the requirement is encoded in the fact that strategy σ_j, played by j in i's interest, is by definition not under control of player i.

- Castelfranchi talks about playing to reach someone else's goal, while we adopt the more immediate notion of favour. Once again, clear formulations of favour or of play to reach someone else's goal are not available in the literature, and different cognitive accounts provide different solutions. As will be clear from Definition 43, a generalization of the standard definitions of best response and dominant strategy can naturally formalize a game-theoretical notion of favour.

- Finally, while our formulation of dependence consists of a three-place relation, Castelfranchi'szzs incorporates further ingredients such as acts, while some other accounts even adopt the notion of plans (e.g. [60]). We reckon a treatment of actions and plans separate from strategies not to be fundamental for a formal theory of dependence in games and we consequently abstract away from them by using the sole notion of strategy.

To formalize favours, which constitute a fundamental ingredient of a theory of dependence, we generalize the notions of best response and dominant strategy, that are applied to strategies that a player plays in his own interest, to the notions of best response and dominant strategy *for someone else*.

Definition 43 (Best for someone else) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, $i, j \in N$ and σ be a strategy profile.*

1. *The strategy σ_j is a* best response for *i if and only if $\forall \sigma'_j \in \Sigma_j$, $o(\sigma) \succeq_i o(\sigma'_j, \sigma_{-j})$.*

2. *The strategy σ_j is a* dominant strategy for *i if and only if $\forall \sigma' \in \bigtimes_{k \in N} \Sigma_k$, $o(\sigma_j, \sigma'_{-j}) \succeq_i o(\sigma')$.*

In words, a strategy σ_j by player j is best response for player i if there is no other strategy σ'_j that guarantees a better outcome to player i than σ_j, provided the other players stick to the profile σ_{-j}; and it is a dominant strategy for player i if, no matter what strategy profile σ'_{-j} the other players stick to, the strategy σ_j by j guarantees to player i a better outcome than any other strategy by j.

Definition 43 generalizes Definition 3 by allowing the player holding the preference to be different from the player whose strategies are considered. When player i and j coincide we get Definition 3 back.

Once we have defined what it means to play in someone else's interest, a definition of dependence is straightforward. In the same fashion as Definition 43, we formulate a notion of *BR*-dependence, if we consider an underlying best response for someone else, and one of *DS*-dependence, if instead we focus on dominant strategies.

Definition 44 (Dependence) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and $i, j \in N$ and σ be a strategy profile.*

1. *Player i* BR-depends *on j for strategy σ_j—in symbols, $iR^{BR}_{\sigma}j$—if and only if σ_j is a best response for i in σ.*
2. *Player i* DS-depends *on j for strategy σ_j—in symbols, $iR^{DS}_{\sigma}j$—if and only if σ_j is a dominant strategy for i.*

Definition 44 deserves a few remarks. The first thing to notice is that the notion of dependence arising from the definition is based on an underlying notion of rationality. In our case we opted for the ones that are, arguably, most standard in a pure strategy setting like ours: best response and dominant strategy. But it must be clear that other choices are possible (e.g. strict best response, strict dominance) and that Definition 44 could be easily extended to accommodate them.

Secondly, a consequence of the definition is that, given any game, we can always associate to any profile σ a binary relation—R^{BR}_{σ} or R^{DS}_{σ}—on the set of players which describes who depends on whom for the realization of that profile. In other words, we can associate to each profile σ a graph (N, R^{BR}_{σ}), or a graph (N, R^{DS}_{σ}), which provide a structural description of the sort of dependencies at work in the underlying game. We call these graphs *dependence graphs*.

Example 19 (The dependence graph of the prisoner's dilemma) *Consider again the prisoner's dilemma in Figure 2.2. Its best response dependence graph is depicted in Figure 5.1. There we notice that, for instance, the relation $R^{BR}_{(U,R)}$ depicted in the up-right corner is such that Column depends on him/herself (it is a reflexive point), as it plays its own best response, but also on Row, as Row does not play its own best response but a best response for Column. Therefore (U, R) displays some kind of 'inbalance'. In contrast, the graph associated to (U, R) depicts a cycle of BR-dependence in which Row plays a best response for Column and, vice versa, Column for Row.*

The last remark worth making is that, in general, relations R^{BR}_{σ} and R^{DS}_{σ} do not enjoy any particular structural property. However, when they do, such structural

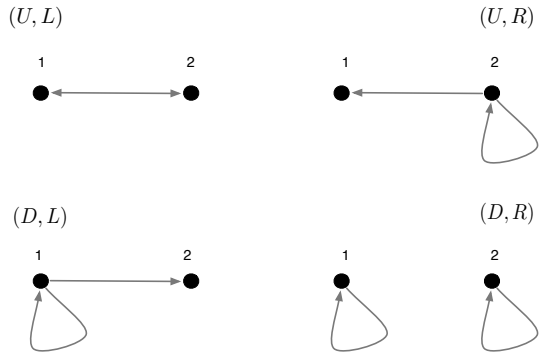

Figure 5.1: BR-dependences in the Prisoner's dilemma.

properties can have a precise game-theoretical meaning. The following simple fact gives a simple example of how structural properties of dependence graphs relate to game-theoretical properties of the underlying games.

Proposition 38 (Reflexive dependencies and equilibria) *Let $\mathbb{G}$ be a game and let $x \in \{BR, DS\}$. It holds that: for any profile σ, R^x_σ is reflexive if and only if σ is an x-equilibrium.*

Proof *Two claims must be proven. [First claim for $x = BR$] From left to right, we assume that $\forall i \in N, iR^{BR}_\sigma i$. From Definition 44, it follows that $\forall i \in N, \forall \sigma' : o(\sigma) \succeq_i o(\sigma'_i, \sigma_{-i})$, that is, σ is a Nash equilibrium. From left to right, we assume that σ is a Nash equilibrium. From this it follows that $\forall i \in N$ σ_i is a best response for i, from which the reflexivity of R^{BR}_σ follows by Definition 44. [Second claim for $x = DS$] An analogous proof applies.*

In other words, any profile in which players depend on themselves—either in a best response or in a dominant strategy sense—is an equilibrium of the corresponding type—BR or DS. Figure 5.1 offers a good pictorial example. The 'defect-defect' profile (D, R), the Nash equilibrium, indeed gives rise to a BR-dependence relation which is reflexive.

5.2.2 Dependence cycles

We have seen above how the reflexivity of dependence is related to the existence of equilibria (Proposition 38). In this section we move to a more general property of dependence relations, the existence of cycles. The literature on dependence theory in multi-agent systems puts particular emphasis on this property as cycles intuitively suggest that there exists common ground for cooperation: if an individual depends on an other individual, and the latter depends in turn on the first to achieve a specific outcome, the choice of that outcome means that the individuals are doing each other a favour. This perspective is very clearly expressed, for instance, in [12, 13], where dependence cycles are taken to signal the possibility of social interaction between players of a *do-ut-des* (give-to-get) type.

	g	$\neg g$
g	3, 3, 3	2, 4, 2
$\neg g$	4, 2, 2	1, 1, 0

g

	g	$\neg g$
g	2, 2, 4	0, 1, 1
$\neg g$	1, 0, 1	1, 1, 1

$\neg g$

Figure 5.2: A three person game. Player 1 denotes *Row*, player 2 *Column*, and player 3 chooses between the right and left matrices.

In that literature, however, dependence relations are considered as given—they do not arise from underlying structures such as games—and so are cycles, whose importance is not motivated in terms of some underlying rationale, but is taken for granted. In this and the following sections (Sections 5.2.2-5.2.4) we show how, starting from dependence relations that arise from an underlying game (Definition 44), we can give precise game-theoretical reasons for the significance of dependence cycles in strategic settings. So let us start with a definition of what a dependence cycle is.

Definition 45 (Dependence cycles) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, (N, R^x_σ) be its dependence structure for profile σ with $x \in \{BR, DS\}$, and let $i, j \in N$. An R^x_σ-dependence cycle $\mathfrak{c}$ of length $k-1$ in $\mathbb{G}$ is a tuple $(a_1, \ldots, a_k)$ such that:*

1. *$a_1, \ldots, a_k \in N$;*
2. *$a_1 = a_k$;*
3. *$\forall a_i, a_j$ with $1 \leq i \neq j < k$, $a_i \neq a_j$;*
4. *$a_1 R^x_\sigma a_2 R^x_\sigma \ldots R^x_\sigma a_{k-1} R^x_\sigma a_k$.*

Given a cycle $\mathfrak{c} = (a_1, \ldots, a_k)$, its orbit $O(\mathfrak{c}) = \{a_1, \ldots, a_{k-1}\}$ denotes the set of its elements.

In other words, cycles are sequences of pairwise different players, except for the first and the last which are equal, such that all players are linked by a dependence relation. Note that the definition allows for cycles of length 1, whose orbit is a singleton, i.e., reflexive arcs. Those are the cycles occurring at reflexive points in the graph.

We have already seen in Example 19 that the cooperative outcome of the prisoner's dilemma exhibits a cycle linking *Row* and *Column* (see Figure 5.1). Even more interesting are cycles in games with more than two players.

Example 20 (Cycles in a three-person game) *Consider the following three-person game.*[3] *A committee of three jurors has to decide whether to declare a defendant in a trial guilty or not. All the three jurors want the defendant to be found guilty, however, all three prefer that*

[3]The game can be viewed as a weak three-person variant of the prisoner's dilemma, where defection, although not being a dominant strategy, can turn out to be a best response for all players, making defection—like in the prisoner's dilemma—a Nash equilibrium.

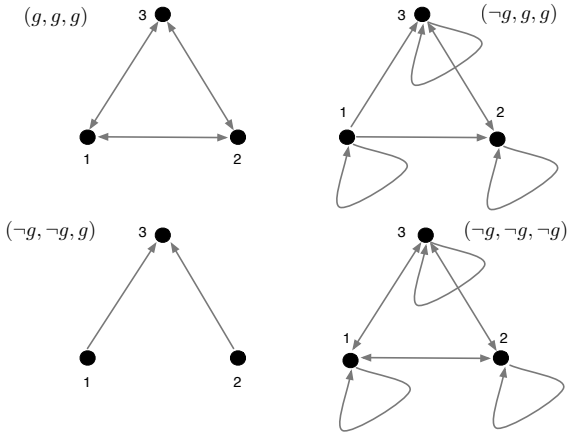

Figure 5.3: Some BR-dependencies from the game matrix in Figure 5.2 (Example 20).

the others declare the defendant guily while she declares her innocent. Also, they do not want to be the only ones declaring the defendant guilty if the other two vote for innocence. They all know each other's preferences. Figure 5.2 gives a payoff matrix for such a game. Figure 5.3 depicts some cyclic BR-dependencies inherent in the game presented. Player 1 is row, player 2 column, and player 3 picks the right or left matrix. Among the ones depicted, (g, g, g) *displays six cycles of length 3 and so does* $(\neg g, \neg g, \neg g)$*, which also contains three reflexive arcs—and hence, by Proposition 38, is a Nash equilibrium. Also* $(\neg g, g, g)$ *is a Nash equilibrium: it does not contain any cycle of length 3, but it does contain two of length two between players 2 and 3. Finally,* $(\neg g, \neg g, g)$ *does not contain any cycle.*

5.2.3 Reciprocity

We now proceed to isolate some specific forms of cycles. These will be used to define several variants of a property of strategic games which we call *reciprocity*. The idea is that, depending on the properties of the dependence cycles of a given profile, we can isolate some significant ways in which players are interconnected via a dependence relation. These will be linked, in the next section, to the existence of equilibria in appropriately transformed games.

Definition 46 (Reciprocity) *Let* $\mathbb{G}$ *be a game,* σ *a profile, and* (N, R^x_σ) *the corresponding dependence graph with* $x \in \{BR, DS\}$*. We say that:*

i) a profile σ *is* x-reciprocal *if and only if there exists a partition* $P(N)$ *of* N *such that each element* p *of the partition is the orbit of some* R^x_σ*-cycle in* (N, R^x_σ)*;*

ii) for $C \subseteq N$, a profile σ is partially x-reciprocal in C *(or C-x-*reciprocal*) if and only if there exists a partition $P(C)$ of C such that each element p of the partition is the orbit of some R_σ^x-cycle in (N, R_σ^x);*

iii) a profile σ is trivially x-reciprocal *if and only if (N, R_σ^x) is reflexive, that is, it contains $|N|$ x-cycles whose orbits are singletons;*

iv) a profile σ is fully x-reciprocal *if and only if (N, R_σ^x) contains at least one x-cycle with orbit N (i.e., a Hamiltonian cycle).*

Let us explain the above definitions by considering the case of best response dependence (BR-dependence). A profile σ is BR-reciprocal if all players belong to some cycle of BR-dependence. This is the case in both the (U, L), i.e., 'cooperate-cooperate', and (D, R), i.e., 'defect-defect', outcomes in the prisoner's dilemma (see Figures 2.2 and 5.1). The other two outcomes are not BR-reciprocal as one of the two players does not belong to the orbit of any cycle.

Along the same lines, a profile σ is partially BR-reciprocal in coalition C (or C-BR-reciprocal) if all the members of C are partitioned by cycles of BR-dependence. This means, intuitively, that independently of whether the players outside of coalition C are linked by dependencies or not, the members of C are in a situation of reciprocity in which everybody plays a best response strategy for somebody else in the coalition. So, in the prisoner's dilemma, outcome (D, L)—maximally preferred by *Row*—is $\{Row\}$-BR-reciprocal as *Row* is playing a best response for herself, hence being in a dependence relation with herself. A perfectly symmetric consideration can be made about (U, R) and *Column*.

Finally, trivial and full BR-reciprocity are special cases of BR-reciprocity. In the first case all players belong to a reflexive arc, that is, all players play their own best response strategy. In the second case there exists one Hamiltonian cycle, that is, all players are connected to one another by a path of *BR*-dependence. For example, inspecting the BR-dependencies in the prisoner's dilemma (Figure 5.1) it can be observed that: (U, L) is fully BR-reciprocal as it contains two Hamiltonian cycles: *Row* $R^{BR}_{(U,L)}$ *Column* $R^{BR}_{(U,L)}$ *Row* and *Column* $R^{BR}_{(U,L)}$ *Row* $R^{BR}_{(U,L)}$ *Column*. On the other hand, (D, R) is trivially *BR*-reciprocal as the only cycles are *Column* $R^{BR}_{(D,R)}$ *Column* and *Row* $R^{BR}_{(D,R)}$ *Row*.

To sum up, a profile is reciprocal when the corresponding dependence relation, be it a *BR*- or *DS*-dependence, clusters the players into non-overlapping groups whose members are all part of some cycle of dependencies (including degenerate ones such as reflexive links). It is partially reciprocal if its dependence graph contains at least one cycle. Trivial and full reciprocity refers to two extreme cases of reciprocity. In the first case the cycles are reflexive arcs and in the second case all players are 'visited' by one and the same cycle.

Before moving to the next section, we first provide one further illustrative example and then study the relation between the two types of reciprocity that arise from Definition 46: best response and dominant strategy reciprocity.

Example 21 (Reciprocity in the three-person game) *Let us go back to Example 20 and to its BR-dependence graph given in Figure 5.3. The graph of profile* (g, g, g) *contains cycles which all yield the partition* $\{\{1, 2, 3\}\}$ *of the set of players. It is then a fully BR-reciprocal profile. The cycles of profile* $(\neg g, g, g)$*, instead, yield two partitions:* $\{\{1\}, \{2\}, \{3\}\}$ *and* $\{\{1\}, \{2, 3\}\}$*, so that profile is BR-reciprocal, but not fully BR-reciprocal. As its graph is reflexive, it is trivially BR-reciprocal, and also partially BR-reciprocal with respect to each nonempty coalition. Interestingly, profile* $(\neg g, \neg g, \neg g)$ *it is both fully and trivially BR-reciprocal. Intuitively, in that profile each player acts in favour of some other player by playing his/her own best response strategy. Finally, profile* $(\neg g, \neg g, g)$ *does not exhibit any form of reciprocity.*

Here below we report a few relevant facts concerning the interplay between DS- and BR-reciprocity.

Proposition 39 (DS- vs. BR-reciprocity) *Let* $\mathbb{G}$ *be a game,* σ *a profile,* $C \subseteq N$*, and* (N, R^x_σ) *be its dependence graph with* $x \in \{BR, DS\}$*. The following holds:*

i) σ *is C-BR-reciprocal if and only if* σ_C *is BR-reciprocal in* $\mathbb{G} \downarrow \sigma_{\overline{C}}$*;*

ii) σ *is C-DS-reciprocal if and only if* σ_C *is DS-reciprocal in* $\mathbb{G} \downarrow \sigma'_{\overline{C}}$ *for any profile* σ'*;*

iii) if σ *is C-DS-reciprocal, then* σ *is C-BR-reciprocal, but not vice versa;*

iv) if σ *is DS-reciprocal, then* σ *is BR-reciprocal, but not vice versa.*

Proof (First claim) *From left to right. By Definition 46, if* σ *is C-BR-reciprocal, then C is the union of orbits of* R^{BR}_σ*-cycles. That is, by Definition 44, each member i of C plays a best response to* σ_{-i} *for some member j in C. Notice that, if we consider player i, best responding to* $\sigma_{C\setminus i}$ *in* $\mathbb{G} \downarrow \sigma_{\overline{C}}$ *is equivalent to best responding to* σ_{-i} *in* $\mathbb{G}$*. From right to left. It follows directly by the notion of best response and by Definitions 46 and 43. [Second claim] From left to right. By Definition 46, if* σ *is C-DS-reciprocal, then C is the union of orbits of a* R^{DS}_σ*-cycle. That is, by Definition 44, each member i of C plays a dominant strategy for some member j in C. As a dominant strategy is such no matter what the other players do, the desired result follows directly. From right to left. It follows directly by the notion of dominant strategy and by Definitions 46 and 43. [Third claim] It follows from the fact that a dominant strategy is always a best response. [Fourth claim] It follows directly from the third claim by setting* $C := N$*.*

In words, the first claim states that a profile σ is partially BR-reciprocal in a coalition C if and only if the restriction σ_C of σ to C is BR-reciprocal with respect to the subgame (recall Definition 5) obtained from $\mathbb{G}$ by fixing the strategy of the complement $\overline{C}$ of coalition C. More concisely, a profile is partially BR-reciprocal in a given coalition if and only if it is BR-reciprocal in the subgame obtained by fixing what the players do who do not belong to the coalition. The second claim is similar and states that a profile is partially DS-reciprocal in a given coalition if and only if it is DS-reciprocal in all subgames obtainable by fixing the strategies of the players who are not in the coalition. These two claims show a first interesting difference between partial BR-

and DS- reciprocity: partial BR-reciprocity is bound by the dependence structure of the current profile while partial DS-reciprocity is not. This is not surprising as the two forms of reciprocity build one on the notion of best response, and the other on the stronger notion of dominant strategy. A second important difference is pointed out by the third and fourth claims, which show that, as expected, (partial) DS-reciprocity is a stronger notion than (partial) BR-reciprocity.

Direct and indirect reciprocity

In the study of cooperation within the social sciences [25] special attention is devoted to the difference between direct reciprocity, seen as direct exchange between two participants, and indirect reciprocity, where instead a favour is returned indirectly. Our framework allows to accommodate and refine these two notions.

Definition 47 (Direct and indirect reciprocity) *Let $\mathbb{G}$ be a game and (N, R^x_σ) be its dependence structure with $x \in \{BR, DS\}$ and σ be a profile, and $C \subseteq N$.*

i) A profile σ is individually x-reciprocal *if and only if there exists a partition $P(N)$ of N such that each element p of the partition is the orbit of some R^x_σ-cycle and $|p| = 1$*

ii) A profile σ is directly x-reciprocal *if and only if there exists a partition $P(N)$ of N such that each element p of the partition is the orbit of some R^x_σ-cycle and $|p| = 2$.*

iii) A profile σ is indirectly x-reciprocal *if and only if it x-reciprocal but not directly nor individually.*

iv) A profile σ is totally indirectly x-reciprocal *if and only if it x-reciprocal but there is no partition $P(N)$ of N with an element p of the partition such that $|p| < 3$ and is the orbit of some R^x_σ-cycle.*

Example 22 (Direct reciprocity among prisoners) *The prisoner's dilemma is a two players' game. We can then expect that reciprocity can arise only in its direct form. The profile of strategy (U, L) is directly DS-reciprocal, while this does not hold for the profile of strategy (D, R) as it is reciprocal but it induces only cycles of length 1. The three persons' prisoner's dilemma of Figure 3 allows for interesting refinements as the profile of strategies $(U, L, 1)$ that arises from the third player choosing the first game is indirectly x-reciprocal. It is not totally indirectly x-reciprocal though, because there is always the possibility of cycles where two players act in direct reciprocity.*

We now proceed to give a game-theoretical interpretation of these definitions of reciprocity based on the notion of dependence cycle.

5.2.4 Reciprocity and equilibrium

We provide a characterization of reciprocity as defined in Definition 46 in terms of standard solution concepts. However, we first have to complement the set of notions provided in Section 2.2.2 with the notion of permuted game.

	L	R
U	0, 0	1, 0
D	0, 1	1, 0

$\mathbb{G}$

L	R
0, 0	0, 1
1, 0	1, 0

$\mathbb{G}^{\mu}$

Figure 5.4: The two horsemen game matrix and its permutation modeling the horse-swap.

Definition 48 (Permuted games) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, σ a profile, and $\mu : N \mapsto N$ a bijection on N. The μ-permutation of game $\mathbb{G}$ is the game $\mathbb{G}^{\mu} = (N^{\mu}, S^{\mu}, \Sigma^{\mu}, \succeq_i^{\mu}, o^{\mu})$ such that:*

- $N^{\mu} = N$;
- $S^{\mu} = S$;
- *for all $i \in N$, $\Sigma_i^{\mu} = \Sigma_{\mu(i)}$;*
- *for all $i \in N$, $\succeq_i^{\mu} = \succeq_i$;*
- $o_{\mu} : \bigtimes_{i \in N} \Sigma_{\mu(i)} \rightarrow S$ *is such that* $o_{\mu}(\mu(\sigma)) = o(\sigma)$, *where $\mu(\sigma)$ denotes the permutation of σ according to μ.*

Intuitively, a permuted game $\mathbb{G}^{\mu}$ is therefore a game where the strategies of each player are redistributed according to μ in the sense that i's strategies become $\mu(i)$'s strategies, where players keep the same preferences over outcomes, and where the outcome function assigns the same outcomes to the same profiles.

Example 23 (Two horsemen [53]) *"Two horsemen are on a forest path chatting about something. A passerby M, the mischief maker, comes along and having plenty of time and a desire for amusement, suggests that they race against each other to a tree a short distance away and he will give a prize of $100. However, there is an interesting twist. He will give the $100 to the owner of the slower horse. Let us call the two horsemen Bill and Joe. Joe's horse can go at 35 miles per hour, whereas Bill's horse can only go 30 miles per hour. Since Bill has the slower horse, he should get the $100. The two horsemen start, but soon realize that there is a problem. Each one is trying to go slower than the other and it is obvious that the race is not going to finish. [. . .] Thus they end up [. . .] with both horses going at 0 miles per hour. [. . .] However, along comes another passerby, let us call her S , the problem solver, and the situation is explained to her. She turns out to have a clever solution. She advises the two men to switch horses. Now each man has an incentive to go fast, because by making his competitor's horse go faster, he is helping his own horse to win!" [53, p. 195-196].*

Once the game of the example is depicted as the left-hand side game matrix in Figure 5.4, it is possible to view the second passerby's solution as a bijection μ which changes the game to the right-hand side version. Now *Row* can play *Column*'s moves and *Column* can play *Row*'s moves. The result is a swap of (D, L) with (U, R),

since (D, L) in $\mathbb{G}^\mu$ corresponds to (U, R) in $\mathbb{G}$ and vice versa. On the other hand, (U, L) and (D, R) stay the same, as the exchange of strategies do not affect them. As a consequence, profile (D, R), in which both horsemen engage in the race, becomes a dominant strategy equilibrium.

On the ground of these intuitions, it is possible to obtain a simple characterization of the different notions of reciprocity given in Definition 46 as the existence of equilibria in appropriately permuted games.

Theorem 40 (Reciprocity in equilibrium) *Let $\mathbb{G}$ be a game and (N, R^x_σ) be its dependence graph with $x \in \{BR, DS\}$ and σ be a profile. It holds that:*

i) σ is x-reciprocal if and only if there exists a bijection $\mu : N \mapsto N$ s.t. σ is a x-equilibrium in the permuted game $\mathbb{G}^\mu$;

ii)
- *σ is partially BR-reciprocal in C (or C-BR-reciprocal) if and only if there exists a bijection $\mu : C \mapsto C$ s.t. σ_C is a BR-equilibrium in the permuted subgame $(\mathbb{G} \downarrow \sigma_{\overline{C}})^\mu$;*
- *σ is partially DS-reciprocal in C (or C-DS-reciprocal) if and only if there exists a bijection $\mu : C \mapsto C$ s.t. σ_C is a DS-equilibrium in all permuted subgames $(\mathbb{G} \downarrow \sigma'_{\overline{C}})^\mu$ for any profile σ';*

iii) σ is trivially x-reciprocal if and only if σ is an x-equilibrium in $\mathbb{G}^\mu$ where μ is the identity over N;

iv) σ is fully x-reciprocal if and only if there exists a bijection $\mu : N \mapsto N$ s.t. σ is a x-equilibrium in the permuted game $\mathbb{G}^\mu$ and μ is such that $\{(i, j) \mid i \in N \,\&\, j = \mu(i)\}$ is a Hamiltonian cycle in N.

Proof *The theorem states eight claims: four for $x = BR$ and four for $x = DS$. [First claim for $x = BR$.] From left to right, assume that σ is BR-reciprocal and prove the claim by constructing the desired μ. By Definition 46 it follows that there exists a partition P of N such that each element p of the partition is the orbit of some R^{BR}_σ-cycle. Given P, observe that any player i belongs to at most one member p of P. Now build μ so that $\mu(i)$ outputs the successor j (which is unique) of i in the cycle whose orbit is the p to which i belongs. Since each j has at most one predecessor in a cycle, μ is an injection and since domain and codomain coincide μ is also a surjection. Now it follows that for all i, j, $iR^{BR}_\sigma j$ implies, by Definition 44, that $\sigma_{\mu(i)}$ is a best response for i in σ. But in $\mathbb{G}^\mu$ it holds that $\sigma_{\mu(i)} \in \Sigma_i$ and since σ is reciprocal, by Definition 46, we have that for all i σ_i is a best response in $\mathbb{G}^\mu$, and hence it is a Nash equilibrium. From right to left, assume μ to be the bijection at issue. It suffices to build the desired partition P from μ by an inverse construction of the one used in the left to right part of the claim. We set $iR^{BR}_\sigma j$ if and only if $\mu(i) = j$. The definition is sound w.r.t. Definition 44 because σ being a Nash equilibrium we have that $iR^{BR}_\sigma j$ if and only if j plays a best response for i in σ. Since μ is a bijection, it follows that R^{BR}_σ contains cycles whose orbits are disjoint and cover N. Therefore, by Definition 46, we can conclude that σ is BR-reciprocal. [First claim for $x = DS$.] The proof is analogous. [Second claim (i)] From right to left assume σ is C-x-reciprocal. It follows that there exists a non-empty*

C s.t. the restriction of R_{BR} to C is a cycle. By Definitions 5 and 46 this is equivalent to stating that σ_C is BR-reciprocal in $\mathbb{G} \downarrow \sigma_{\overline{C}}$. From this, by the first claim we obtain that σ_C is a BR-equilibrium in $(\mathbb{G} \downarrow \sigma_{\overline{C}})^{\mu}$ for some bijection μ. As these steps are all equivalences (Proposition 39 first equivalence) we also directly obtain the direction from left to right. [Second claim (ii)] The proof follows the line of the proof of the previous claim but exploits the second equivalence of Proposition 39. [Third claim for x = BR] Follows directly from Definition 46 and Proposition 38. [Third claim for x = DS] Follows directly from Definition 46 and Proposition 38. [Fourth claim for x = BR] Follows directly from Definition 46 and the first claim. [Fourth claim for x = DS] It can be proven in the same way.

Intuitively, the theorem connects all cycle-based forms of reciprocity identified in Definition 46 with equilibria in (sub-)games that could be obtained by appropriate permutations of the underlying game. Furthermore, the instructions for such permutations—which strategies go to which player—are provided by the existent cycles. So, if the profile is trivially reciprocal (third claim), then it is already an equilibrium, and if it is fully reciprocal (fourth claim), it then becomes an equilibrium via a permutation that follows one of the available Hamiltonian cycles over the set of players.

We hold Theorem 40 to be of particular interest for two reasons. First, it provides a clear connection between intuitions developed in the theory of dependence—such as the significance of cycles—with notions which lie at the heart of game theory—such as that of equilibrium. Second, it provides a systematic dependence-based rationale for modifications of games that allow desirable but unstable outcomes—such as the cooperative outcome in the prisoner's dilemma—to become equilibria.

Implementation via permutation

As just discussed, in view of Theorem 40, permutations can be fruitfully viewed as ways of *implementing*—in a social software sense [53]—reciprocal profiles. This terminology can be made formal as follows.

Definition 49 (Implementation as game permutation) *Let $\mathbb{G}$ be a game, σ a profile and (N, R^x_σ) be its dependence graph. Let also $\mu : N \mapsto N$ be a bijection with $C \subseteq N$. We say that:*

i) μ BR-implements σ if and only if σ is a BR-equilibrium in $\mathbb{G}^{\mu}$;

ii) μ DS-implements σ if and only if σ is a DS-equilibrium in $\mathbb{G}^{\mu}$;

iii) μ partially BR-implements σ in C if and only if σ_C is an BR-equilibrium in $(\mathbb{G} \downarrow \sigma_{\overline{C}})^{\mu}$;

iv) μ partially DS-implements σ in C if and only if σ_C is a DS-equilibrium in $(\mathbb{G} \downarrow \sigma'_{\overline{C}})^{\mu}$ for any profile σ'.

Intuitively, implementation is here understood as a way of transforming a game in such a way that the desirable outcomes, in the transformed game, are brought about at an equilibrium point. In this sense we talk about BR- or DS-implementation.

	L	R
U	2, 2	0, 3
D	3, 0	1, 1

$\mathbb{G}$

	L	R
U	2, 2	3, 0
D	0, 3	1, 1

$\mathbb{G}^{\mu}$

Figure 5.5: The prisoner's dilemma matrix and its permutation swapping *Row* with *Column* according to the cycle of profile (U, L).

Analogously, partial BR- or DS-implementation consist in the realization of the desirable outcomes as equilibria in one subgame (partial BR-implementation) or all possible sub-games (partial DS-implementation).

Example 24 (Implementation of cooperation in the prisoner's dilemma) *Just as for the two-horsemen example (Example 23) we can think of implementing the cooperative outcome* (U, L) *of the prisoner's dilemma (Figure 2.2) by permuting the game according to the permutation* μ *dictated by one of the Hamiltonian cycles present in the dependence graph of that profile (Figure 5.1):* $\mu(Row) = Column$ *and* $\mu(Column) = Row$*. In the resulting game where, essentially, Row decides whether Column plays L or R and Column whether Row plays U or D, the cooperative outcome is a Nash equilibrium by Theorem 40 (see Figure 5.5). An example of partial BR-implementation is provided by Example 20 (see also Figure 5.2). There, profile* $(\neg g, g, g)$ *is partially BR-reciprocal in coalition* $\{1, 2\}$ *(see Example 21). A permutation between* 2 *and* 3 *would yield a game such that* (g, g) *is a Nash equilibrium in the sub-game obtained by fixing the strategy of player* 1 *to* $\neg g$ *(notice, however, that in this case the identity permutation would also guarantee such result). In other words, were it so that* 1 *had already made his/her choice, swapping the strategies of* 2 *and* 3 *would lead to a stable outcome.*

5.3 Solving Dependencies: Dependence Games

The previous sections have shown how reciprocity can be given two corresponding formal characterizations: existence of cycles in a dependence structure, and existence of equilibria in a suitably permuted game (Theorem 40). In the present section, we apply the notion of reciprocity to obtain a refinement of coalitional games. The intuition behind such refinement consists, in a nutshell, in allowing coalitions to form only in the presence of some sort of reciprocity.

5.3.1 Agreements

The central solution concept for games that take dependence relations seriously will be the one of players' agreement. The key idea behind it is that, given a reciprocal profile (of some sort according to Definition 46), the players can fruitfully *agree* to transform the game by some suitable permutation of sets of strategies.

Definition 50 (Agreements and partial agreements) *Let $\mathbb{G}$ be a game, (N, R^x_σ) be its dependence structure in σ with $x \in \{BR, DS\}$, and let $i, j \in N$. A pair (σ, μ) is:*

i) an x-agreement for $\mathbb{G}$ if σ is an x-reciprocal profile, and $\mu : N \mapsto N$ a bijection which x-implements σ;

ii) a partial x-agreement in C (or a C-x-agreement) for $\mathbb{G}$, if σ is a C-x-reciprocal profile and $\mu : C \mapsto C$ a bijection which C-x-implements σ.

The set of x-agreements of a game $\mathbb{G}$ is denoted x-AGR($\mathbb{G}$) and the set of partial x-agreements, that is the set of pairs (σ, μ) for which there exists a C such that μ C-x-implements σ, is denoted x-pAGR($\mathbb{G}$).

Intuitively, a (partial) agreement, of BR or DS type, can be seen as the result of coordination (endogenous, via the players themselves, or exogenous, via a third party like in Example 23) selecting a desirable outcome and realizing it by an appropriate exchange of strategies.

Example 25 (Agreements in the prisoner's dilemma) *Let us go back to the prisoner's dilemma. Agreement $((D, R), \mu)$ with $\mu(i) = i$ for all players, is the standard DS-equilibrium of the strategic game. But there is another possible agreement, where the players swap their strategies: it is $((U, L), \nu)$, for which $\nu(i) = N\backslash\{i\}$. Here Row plays cooperatively for Column and Column plays cooperatively for Row. Of the same kind is the agreement arising in Example 23. Notice that in such an example, the agreement is the result of coordination mediated by a third party (the second passerby). Analogous considerations can also apply to Example 20 where, for instance, $((g, g, g), \mu)$ with $\mu(1) = 2, \mu(2) = 3, \mu(3) = 1$ is a BR-agreement.*

As we might expect, BR- and DS-agreements are related in the same way as BR- and DS-reciprocity (Proposition 39). In what follows we will focus only on DS-agreements and partial DS-agreements so, whenever we talk about agreements and partial agreements, we mean DS-agreements and partial DS-agreements, unless stated otherwise.

5.3.2 Dominance between agreements

As there can be several possible agreements in a game, the natural question arises of how to order them. We will do that by defining a natural notion of dominance between agreements, but first we need some auxiliary notions.

Definition 51 (*C*-candidates and *C*-variants) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and C a non-empty subset of N. An agreement (σ, μ) for $\mathbb{G}$ is a* C-candidate *if C is the union of some members of the partition induced by μ, that is: $C = \bigcup X$ where X is an element of the partition induced by μ on N. An agreement (σ, μ) for $\mathbb{G}$ is a* C-variant *of an agreement (σ', μ') if $\sigma_C = \sigma'_C$ and $\mu_C = \mu'_C$, where μ_C and μ'_C are the restrictions of μ to C. As a convention we take the set of $\emptyset$-candidate agreements to be empty and an agreement (σ, ν) to be the only $\emptyset$-variant of itself.*

In other words, an agreement (σ, μ) is a C-candidate if the partial dependence relation for σ of C and $\overline{C}$ follows exactly μ, and it is a C-variant of (σ', μ') if it differs from this latter at most in its C-part. We can now define the following notions of dominance between agreements and between partial agreements.

Definition 52 (Dominance) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and $C \subseteq N$ be a coalition. We say that:*

i) An agreement *(σ, μ) is dominated if and only if for some coalition C there exists a C-candidate agreement (σ', μ') for $\mathbb{G}$ such that for all agreements (ρ, ν) which are $\overline{C}$-variants of (σ', μ'), $o(\rho) \succ_i o(\sigma)$ for all $i \in C$.*

ii) A partial agreement *(σ_C, μ) in C is dominated if and only if for some coalition $C' \subseteq N$ there exists $(\tau_{C'}, \nu)$ which is a C'-agreement such that for all σ', τ', $o(\tau_{C'}, \tau'_{\overline{C'}}) \succ_i o(\sigma_C, \sigma'_{\overline{C}})$ for all $i \in C'$.*

The set of undominated agreements of $\mathbb{G}$ is denoted $DEP(\mathbb{G})$ and the set of undominated partial agreements is denoted $pDEP(\mathbb{G})$.

Intuitively, an agreement is undominated when a coalition C can force all possible agreements to yield outcomes which are better for all the members of the coalition, regardless of what the rest of the players can agree to do, that is, regardless of the $\overline{C}$-variants of their agreements. A partial agreement in coalition C is undominated when C can, by means of a partial permutation, force the game to end up in a set of states which are better for the member of the coalition no matter what the players in $\overline{C}$ do.

It is worth stressing the critical difference between the two notions of dominance. This difference resides in the fact that while dominance between agreements only considers deviations which are the results of agreements, dominance between partial agreements considers any form of possible deviation.

Example 26 (Dominance between partial agreements) *In the three person game of Figure 5.2, $((g_1, g_2), (\mu(1) := 2, \mu(2) := 1))$ is a partial DS-agreement in $\{1, 2\}$. This agreement, which represents a form of dependence-based cooperation between 1 and 2 dominates the partial DS-agreement in N—on a trivially DS-reciprocal profile—$((\neg g_1, \neg g_2, \neg g_3), (\mu(1) := 1, \mu(2) := 2, \mu(3) := 3))$. In fact, it is undominated, since even the partial DS-agreement in N $((g_1, g_2, g_3), (\mu(1) := 2, \mu(2) := 3, \mu(3) := 1))$ (which is also a DS-agreement) does not dominate it.*

5.3.3 Dependence-based coalitional games

Agreements exploit the dependence relations between the players in order to achieve some form of mutually beneficial cooperation. It is then natural to use agreements to study games in strategic form as some form of coalitional games where players form coalitions only in the presence of reciprocity. Standard questions of cooperative game theory can then be meaningfully asked, such as the following one:

Can we characterize the notion of dominance for agreements and partial agreements (Definition 52) in terms of a suitable notion of stability in appropriately defined coalitional games?

In order to answer this question we proceed as follows. First, starting from a game $\mathbb{G}$, we consider its representation $\mathbb{C}^{\mathbb{G}}$ as a coalitional game as illustrated in Definition 12. As Definition 12 abstracts from dependence-theoretic considerations we refine it in two ways, corresponding to the two different sorts of dependence upon which we want to build the coalitional game:

1. The first refinement is obtained by defining a coalitional game $\mathbb{C}^{\mathbb{G}}_{DEP}$ capturing the intuition that coalitions form only by means of *agreements* (Definition 50). Such games are called *dependence games*.

2. The second is obtained by defining a coalitional game $\mathbb{C}^{\mathbb{G}}_{pDEP}$ capturing the intuition that coalitions form only by means of *partial agreements* (Definition 50). Such games are called *partial dependence games*.

Having done this, we show that the core of $\mathbb{C}^{\mathbb{G}}_{DEP}$ coincides with the set of undominated agreements of $\mathbb{G}$ (Theorem 41) and, respectively, that the core of $\mathbb{C}^{\mathbb{G}}_{pDEP}$ coincides with the set of undominated partial agreements of $\mathbb{G}$ (Theorem 42). We thereby obtain a cooperative game-theoretical characterization of the notion of dominance in Definition 52, formally linking agreements to the core of classes of coalitional games.

Dependence games

We start by refining the method to obtain a coalitional game from a game in strategic form (Definition 12), thus defining the notion of dependence game.

Definition 53 (Dependence games from strategic ones) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. The dependence game $\mathbb{C}^{\mathbb{G}}_{DEP} = (N, S, E^{\mathbb{G}}_{DEP}, \succeq_i)$ of $\mathbb{G}$ is a coalitional game where the effectivity function $E^{\mathbb{G}}_{DEP}$ is defined as follows:*

$$
\begin{aligned}
X \in E^{\mathbb{G}}_{DEP}(C) \quad \Leftrightarrow \quad & \exists \sigma_C, \mu_C \text{ s.t.} \\
& \exists \sigma_{\overline{C}}, \mu_{\overline{C}} : [((\sigma_C, \sigma_{\overline{C}}), (\mu_C, \mu_{\overline{C}})) \in AGR(\mathbb{G})] \\
& \text{AND } [\forall \sigma_{\overline{C}}, \mu_{\overline{C}} : [((\sigma_C, \sigma_{\overline{C}}), (\mu_C, \mu_{\overline{C}})) \in AGR(\mathbb{G}) \\
& \text{IMPLIES } o(\sigma_C, \sigma_{\overline{C}}) \in X]].
\end{aligned}
$$

where $\mu : N \to N$ is a bijection.

This somewhat intricate formulation states nothing but that the effectivity function $E^{\mathbb{G}}_{DEP}(C)$ associates with each coalition C the states which are outcomes of agreements (and hence of reciprocal profiles), and which C can force via partial agreements (σ_C, μ_C) regardless of the partial agreements $(\sigma_{\overline{C}}, \mu_{\overline{C}})$ of $\overline{C}$.

We obtain the following theorem.

Theorem 41 (*DEP* vs. *CORE*) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. It holds that, for all agreements (σ, μ):*

$$(\sigma, \mu) \in DEP(\mathbb{G}) \;\Leftrightarrow\; o(\sigma) \in CORE(\mathbb{C}^{\mathbb{G}}_{DEP}).$$

where $\mu : N \to N$.

Proof *[Left to right:] By contraposition, assume $o(\sigma) \notin CORE(\mathbb{C}^{\mathbb{G}}_{DEP})$. By Definition 14 this means that $\exists C \subseteq N, X \in E^{\mathbb{G}}_{DEP}(C)$ s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. Applying Definition 53 we obtain that there exists an agreement $((\sigma'_C, \sigma'_{\overline{C}}), (\mu'_C, \mu'_{\overline{C}}))$ s.t. $\forall \sigma'_{\overline{C}}, \mu'_{\overline{C}}$, $o(\sigma'_C, \sigma'_{\overline{C}}) \in X$ and s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. Now, $((\sigma'_C, \sigma'_{\overline{C}}), (\mu'_C, \mu'_{\overline{C}}))$ is obviously C-candidate, and all its C-variants yield better outcomes for C than σ. Hence, by Definition 52,$(\sigma, \mu) \notin DEP(\mathbb{G})$. [Right to left:] Notice that the set up of Definition 52 implies that, if (σ, μ) is dominated, then any other agreement for σ would also be dominated. So, by contraposition, assume $(\sigma, \mu) \notin DEP(\mathbb{G})$. By Definition 52, we obtain that there exists a C-candidate agreement (σ', μ') for $\mathbb{G}$ such that for all agreements (ρ, ν) which are $\overline{C}$-variants of (σ', μ'), $o(\rho, \nu) \succ_i o(\sigma)$ for all $i \in C$. But this means, by Definition 53, that $\exists C, X$ such that $X \in E^{\mathbb{G}}_{DEP}(C)$ and $x \succ_i o(\sigma)$ for all $x \in C$. Hence, by Definition 14, we obtain $o(\sigma) \notin CORE(\mathbb{C}^{\mathbb{G}}_{DEP})$.*

Put otherwise, here is what Theorem 41 states. Given a game $\mathbb{G}$, a profile σ which is partially DS-implemented by μ (Definition 49) forms an undominated partial agreement (σ, μ) if and only if σ is in the core of the dependence game of $\mathbb{G}$. By taking Definition 46 and Theorem 40 into the picture, we thus see that Theorem 41 connects three apparently rather different properties of a strategic game $\mathbb{G}$: the existence of reciprocal profiles, the existence of DS-equilibria in permutations of $\mathbb{G}$, and the core of the dependence game built on $\mathbb{G}$.

Partial dependence games

We now move on to define the class of partial dependence games, in a way analogous to that followed for dependence games in Definition 53.

Definition 54 (Partial dependence games from strategic ones) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. The partial dependence game $\mathbb{C}^{\mathbb{G}}_{pDEP} = (N, S, E^{\mathbb{G}}_{pDEP}, \succeq_i)$ of $\mathbb{G}$ is a coalitional game where the effectivity function $E^{\mathbb{G}}_{pDEP}$ is defined as follows:*

$$\begin{aligned} X \in E^{\mathbb{G}}_{pDEP}(C) \quad \Leftrightarrow \quad & \exists \sigma_C, \mu_C \text{ s.t.} \\ & (\sigma_C, \mu_C) \in pAGR(\mathbb{G}) \\ & \text{AND } [\forall \sigma_{\overline{C}} : o(\sigma_C, \sigma_{\overline{C}}) \in X]]. \end{aligned}$$

where $\mu_C : C \to C$ is a bijection.

Partial dependence games are defined by just looking at the set of outcomes that each coalition can force by means of a partial agreement. Unlike Definition 53, Definition 54 is much closer to the standard definition of coalitional games based on strategic ones (Definition 12).

Like for dependence games, we have a characterization of the set of undominated partial agreements.

Theorem 42 (*pDEP* vs. *CORE*) *Let $\mathbb{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. It holds that, for all agreements (σ, μ):*

$$(\sigma, \mu) \in pDEP(\mathbb{G}) \;\Leftrightarrow\; o(\sigma) \in CORE(\mathbb{C}^{\mathbb{G}}_{pDEP}).$$

where $\mu : C \to C$ is a bijection with $C \subseteq N$.

Proof *[Left to right:] By contraposition, assume $o(\sigma) \notin CORE(\mathbb{C}^{\mathbb{G}}_{pDEP})$. By Definition 14 this means that $\exists C \subseteq N, X \in E^{\mathbb{G}}_{DEP}(C)$ s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. Applying Definition54 we obtain that there exists a partial agreement (ρ'_C, μ'_C) s.t. $\forall \rho'_{\overline{C}}, o(\rho'_C, \rho'_{\overline{C}}) \in X$ and s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. By Definition 52, $(\sigma, \mu) \notin pDEP(\mathbb{G})$. [Right to left:] By contraposition, assume $(\sigma, \mu) \notin pDEP(\mathbb{G})$. By Definition 52, we obtain that there exists a partial agreement (ρ'_C, μ'_C) for $\mathbb{G}$ such that for all $\rho'_{\overline{C}}$, $o(\rho, \nu) \succ_i o(\sigma)$ for all $i \in C$. But this means, by Definition 54, that $\exists C, X$ such that $X \in E^{\mathbb{G}}_{pDEP}(C)$ and $x \succ_i o(\sigma)$ for all $x \in C$. Hence, by Definition 14, we obtain $o(\sigma) \notin CORE(\mathbb{C}^{\mathbb{G}}_{DEP})$.*

Like Theorem 41, Theorem 42 establishes a precise connection between the notions of partial reciprocity in a strategic game $\mathbb{G}$, the existence of DS-equilibria in all permuted subgames of $\mathbb{G}$, and the core of the partial dependence game built on $\mathbb{G}$.

5.3.4 Coalitional, dependence, partial dependence effectivity

The coalitional game $\mathbb{C}^{\mathbb{G}}$ built on a strategic game $\mathbb{G}$ and its dependence-based counterparts $\mathbb{C}^{\mathbb{G}}_{DEP}$ and $\mathbb{C}^{\mathbb{G}}_{pDEP}$ are logically related. The following fact shows how.

Proposition 43 (Effectivity functions related) *The following relations hold:*

i) For all $\mathbb{G}$: $E^{\mathbb{G}}_{pDEP} \subseteq E^{\mathbb{G}}$;

ii) It does not hold that for all $\mathbb{G}$: $E^{\mathbb{G}}_{DEP} \subseteq E^{\mathbb{G}}_{pDEP}$; nor does it hold that for all $\mathbb{G}$: $E^{\mathbb{G}}_{pDEP} \subseteq E^{\mathbb{G}}_{DEP}$;

iii) It does not hold that for all $\mathbb{G}$: $E^{\mathbb{G}}_{DEP} \subseteq E^{\mathbb{G}}$; nor does it hold that for all $\mathbb{G}$: $E^{\mathbb{G}} \subseteq E^{\mathbb{G}}_{DEP}$.

Proof (First Claim) *Suppose not. Then for some $\mathbb{G}$, some $X \subseteq S$ and some $C \subseteq N$ we have that $X \in E^{\mathbb{G}}_{pDEP}$ and $X \notin E^{\mathbb{G}}(C)$. The latter means that $\neg\exists \sigma_C \in \bigtimes_{i \in C} \Sigma_i$ such that $\forall \sigma_{\overline{C}} o(\sigma_C, \sigma_{\overline{C}}) \in X$. However this implies by elementary logical reasoning that $\neg\exists \sigma_C \in \bigtimes_{i\ in C} \Sigma_i \neg\exists \mu_C$ such that $(\sigma_C, \mu_C) \in pAGR(\mathbb{G})$ and $\forall \sigma_{\overline{C}} o(\sigma_C, \sigma_{\overline{C}}) \in X$, i.e. that $X \notin E^{\mathbb{G}}_{pDEP}$. Contradiction. [Second Claim] To refute the first inclusion consider a prisoner's dilemma in which $\{(D, R)\} \in E^{\mathbb{G}}_{DEP}(\{Column, Row\})$ but $\{(D, R)\} \notin E^{\mathbb{G}}_{pDEP}(\{Column, Row\})$. The second inclusion instead can be refuted by any game $\mathbb{G}$ that features a partial agreement (μ_C, σ_C) for players in C but no partial agreement for players in $\overline{C}$. [Third Claim] For the first inclusion consider a game $\mathbb{G}$ with three players $1, 2, 3$ and two actions $\{a, b\}$ for each of them.*

Suppose the only possible agreement is the identity permutation $\mu(i) = i$ and (a, a, a) is a DS-equilibrium. We have that $\{o(a, a, a)\} \in E^{\mathbb{G}}_{DEP}(\{1\})$ while $\{o(a, a, a)\} \notin E^{\mathbb{G}}(\{1\})$. For the second inclusion take $\mathbb{G}$ the prisoner's dilemma game in which $\{(U, R)\} \in E^{\mathbb{G}}(\{Column, Row\})$ but $\{(U, R)\} \notin E^{\mathbb{G}}_{DEP}(\{Column, Row\})$.

The fact shows that dependence-based effectivity functions considerably modify the powers assigned to coalitions by the standard definition of coalitional games on strategic ones (Definition 12). Partial dependence effectivity functions instead really weaken the notion of coalitional ability, reducing the coalitional strategy at players' disposal. A formal consequence of Proposition 43 is the establishment of the relation between $CORE(\mathbb{C}^{\mathbb{G}})$, $CORE(\mathbb{C}^{\mathbb{G}}_{DEP})$ and $CORE(\mathbb{C}^{\mathbb{G}}_{pDEP})$, as a direct consequence of the inclusion relation among their corresponding effectivity functions.

Summing up, the results in this section have shown that agreements and partial agreements are a form of coalitional power that can be related to standard cooperative solution concepts such as the core (Theorems 41 and 42). In particular, partial agreements can be seen as a weakened form of coalitional strategies (Proposition 43), i.e. those strategies that can be executed only in the presence of mutual reciprocity among the members of a coalition. As such partial dependence games, which generalize dependence games, should be understood as an intermediate level between the individualistic perspective studied in strategic games and the group perspective analyzed in coalitional games.

5.3.5 An application to transferable utility games

The present section shows an application of dependence theory to transferable utility games [51]. In transferable utility games (in short *TU games*) the preference relations are replaced by payoff functions, that associate to each strategy profile a positive real number, with the intuitive understanding that the number symbolizes what the player gets at the state associated to that strategy profile.

Definition 55 (TU game) *A (strategic form) transferable utility game (TU game) is a tuple $\mathbb{G} = (N, \Sigma_i, p_i)$ where:*

- *N is a set non-empty set of players;*
- *Σ_i is a set of strategies for player $i \in N$;*
- *$p_i : \bigtimes_{i \in N} \Sigma_i \to \mathbb{R}_+$ is a payoff function, that associates to each player and each strategy profile a positive real number.*

So, TU games are games in strategic form where the outcome function is substituted by a payoff function where numerical payoffs encode agents' preferences. All games in Figure 2.2 are then examples of TU Games.[4]

[4]TU games are defined without an outcome function and the payoffs are directly associated to strategy profiles. TU games can be translated into standard strategic games (Definition 1) by endowing them with a bijective outcome function and a class of preference relations $\preceq_i$ induced by the payoff functions

Definition 12 allows us to translate strategic games into coalitional games by using so-called α-effectivity functions. In TU games this translation is not available as players together can only reach vectors of reals and not set of outcomes, but a similar notion can be defined, interpreting what a coalition can achieve as the best payoff that a coalition is able to achieve on its own, i.e., what we call the *value* of a coalition. We first define the coalitional payoff associated to each strategy profile.

Definition 56 (Coalitional payoff in TU Games) *Let $\mathbb{G}$ be a TU game and $\sigma \in \Sigma$ be a strategy profile. The payoff of coalition C for the strategy profile σ, $p_C(\sigma)$ is defined as follows:*

$$p_C(\sigma) = \sum_{i \in C} p_i(\sigma).$$

Taking into account the possible replies of the opponents, we are able to define the minimal payoff, namely $min_{\sigma_{\overline{C}}} p_C(\sigma)$, for each strategy σ_C that coalition C can play. In essence, we take a negative view on the opponents assuming that they will try to minimize C's payoff. Coalitions can then choose the best strategy knowing each minimal payoff, which constitutes the value of the coalition.

Definition 57 (Value of a coalition) *Let $\mathbb{G}$ be a TU game and $C \subseteq N$ be a coalition. $v^\alpha(C)$, the value that coalition C is able to guarantee, is defined as follows:*

$$v^\alpha(C) = max_{\sigma_C} min_{\sigma_{\overline{C}}} p_C(\sigma).$$

In words, the value of a coalition C is the payoff $p_C(\sigma)$ where σ_C is the most rewarding collective strategy that C can play knowing that $\sigma_{\overline{C}}$ is the toughest collective strategy by $\overline{C}$.

Definition 57 allows coalition C to select any collective strategy at its disposal. One immediate contribution of the theory of dependence is to restrict the set of available strategies via partial agreements. As done in Section 5.3, we restrict our attention to partial *DS*-agreements.

Definition 58 (Coalitional Agreements in TU Games) *Let $\mathbb{G}$ be a TU game. σ_C is a* coalitional agreement of C in game $\mathbb{G}$ if there exists $\mu : C \to C$, such that for all $i \in C$ and for all σ'_i and ρ_{-i} we have that:

$$p_{\mu(i)}(\sigma_i, \rho_{-i}) \geq p_{\mu(i)}(\sigma'_i, \rho_{-i})$$

We call $AGR^{TU}(\mathbb{G})_C$ the set of coalitional agreements of C at game $\mathbb{G}$.

Now we can define the negotiated value of a coalition, i.e., the payoff that a coalition can guarantee by undertaking agreements.

---

as expected:

$$o(\sigma) \preceq_i o(\sigma') \text{ if and only if } p_i(\sigma) \leq p_i(\sigma')$$

for each strategy profile σ, σ' and each player $i \in N$.

Definition 59 (Negotiated value of a coalition) *Let $\mathbb{G}$ be a game and $C \subseteq N$ be a coalition. $v^{\alpha}_{DEP}(C)$, the negotiated value of coalition S is able to guarantee, is defined as follows:*

$$v^{\alpha}_{DEP}(C) = max_{\sigma_C \in AGR^{TU}(\mathbb{G})_C} min_{\sigma_{\overline{C}}} p_C(\sigma).$$

Intuitively, $v^{\alpha}_{DEP}(C)$ represents the payoff that players in C can guarantee undertaking an agreement that answers the toughest collective strategy by their opponents. As a convention, if a coalition is not able to reach any agreement, v^{α}_{DEP} is set to 0. Notice the similarity with partial agreements, where the opponents of a coalition can play strategies that *need not be themselves coalitional agreements*.

Several properties are desirable for TU games. One of the most fundamental is that of superadditivity of the coalitional value, i.e the fact that coalitions can achieve more by uniting than by playing separately. In our case this translates into the following requirement: $v^{\alpha}_{DEP}(C) + v^{\alpha}_{DEP}(C') \leq v^{\alpha}_{DEP}(C' \cup C)$ for each disjoint C, C'. Under a mild assumption, this property holds.

Proposition 44 *Let $C \subseteq N, C' \subseteq N$ and $C \cap C' = \emptyset$ and $v^{\alpha}_{DEP}(C) > 0$, $v^{\alpha}_{DEP}(C) > 0$.*

$$v^{\alpha}_{DEP}(C) + v^{\alpha}_{DEP}(C') \leq v^{\alpha}_{DEP}(C' \cup C)$$

Proof *It follows from the fact that $\sigma_C \in AGR^{TU}(\mathbb{G})_C$ and $\sigma_{C'} \in AGR^{TU}(\mathbb{G})_D$ for disjoint C, C' implies that $\sigma_{C \cup C'} \in AGR^{TU}(\mathbb{G})_{C \cup C'}$ and Definition 57.*

In words the proposition says that coalitions can favourably merge their partial agreements. In general partial agreements are not superadditive, as a consequence of the fact that bigger coalitions may hinder agreements instead of favouring them, unless their separate components could already agree on something. In the latter case a partial agreement can always been obtained, by merging the two disjoint partial agreements. Finally, it is instructive to notice that superadditivity would not be obtainable under simple agreements. This depends on the fact that agreements are defined with respect to a whole strategy profile limiting considerably the possibilities of coalition formation and, therefore, of coalition merging.

In summary, the concept of partial agreement applies naturally to games with transferable utility, where the value of coalitions can be calculated looking at the payoff that players can achieve together (Definition 57). In the context of such games, partial agreements turn out to be well-behaved as they enforce, under a natural assumption, the desirable property of superadditivity (Proposition 44) which constitutes a common assumption in cooperative game theory.

5.4 Discussion

5.4.1 Related work

We relate here our game-theoretical view of dependence to existing literature in MAS. To the best of our knowledge, almost no attention has been dedicated up

till now to the relation between game theory and dependence theory. There are, however, three noteworthy exceptions:

- A study of the added value of exchanging tasks in a restricted game-like structure can be found in [14], where task-exchange is studied as a means to ease the computation of coalitions.
- Recently, the work presented in [15, 58] elaborates on ideas close to [14] applying them to a special class of games, called Boolean games.

We first briefly describe the framework in [14] and then discuss [15, 58].

Coalitions that exchange tasks

A series of papers by Boella, Sauro and van der Torre [14, 12, 57] put forward a formalization of the notions of power and dependence inspired by the work of Castelfranchi and colleagues. In their work the attempt of using dependence relations to form coalitional structures first makes its appearance. In [14] the authors frame their definitions within *task based power structures* [5].

Definition 60 (Task based power structures) *A task based power structure is a tuple* $\langle Ag, G, T, goals, power\rangle$ *where:*

- Ag *is a non-empty set of agents;*
- G *is a non-empty set of goals;*
- T *is a non-empty set of tasks;*
- $goals : Ag \to 2^G$ *is a function that associates to each agent in* Ag *the subset of goals* G *it desires to achieve;*
- $power : 2^{Ag\times T} \to 2^G$ *is a partial function that associates to a task assignment* $\tau \subseteq Ag \times T$ *a set of goals that the task assignment achieves.*

From their very first definition, the structures analyzed in [14] are characterized by a clear resemblance to strategic games. However, a variety of new concepts are introduced, such as tasks and tasks assignments, power and goals. It is worth stressing, even more because these concepts are used also in [58], the primitive character of goals and tasks, which are introduced as sets with no further formal requirement. In a task based power structure the classical definitions of rationality (such as those in Definition 3) are not immediately available and the costs and the benefits for each agents need to be independently defined.

Definition 61 (Costs and Benefits) *The benefits and costs of the task assignment* τ *for agent* $a \in Ag$ *are defined as follows:*

$$\begin{aligned} benefits(\tau, a) &= goals(a) \cap power(\tau) \\ costs(\tau, a) &= \{t \in T \mid (a, t) \in \tau\} \end{aligned}$$

[5] Similar structures are studied in [12, 57] and subsequent papers.

The authors of [14] proceed then to defining a notion of domination among task assignments. A task assignment τ_1 is dominated if there is a task assignment $\tau_2 \subseteq \tau_1$ such that all the agents involved in τ_2 enjoy higher benefits in τ_2 (or lower costs). Task assignments that are not dominated are called *do-ut-des* task assignments.

Despite the resemblance of task based power structures to strategic games, the do-ut-des links among players in [14] are of a completely different nature from ours. Apart from the lack of structure of goals, the correspondent of preferences in tasks based structures, and the consequent difficulty of formulating classical solution concepts, it appears even more problematic to relate the notion of benefit in Definition 61 to our notion of favour. While the latter is a straightforward generalization of the classical notions of best response and dominant strategy the do-ut-des in [14] is only concerned with the burden that a player bears in a task assignment and does not aim at being a formal correspondent of reciprocity in games. As authors of [14] rightly claim their definition should be interpreted as "give something to obtain something else", without concerning the strategic aspects of decision making.

Dependence in Boolean games

Recently, [15] and [58] have studied a notion of dependence for a restricted class of strategic games called Boolean games [38]. In a nutshell, Boolean games are n-player games where players act by controlling the truth value of a propositional variable, and where players' preferences are dichotomous, that is, each player has a single goal—expressed by a propositional formula—which is either fulfilled or not. The work presented in [58] then extends some of the results presented in [15] to the class of cooperative Boolean games [26], that is, a coalitional version of Boolean games. Like in our case, the authors of [15] and [58] look at dependence relations as graph-theoretical information hidden within the game structure. However, there are several important differences.

First of all the simple structure of Boolean games, and in particular the fact that players' preferences are dichotomous, allows for a definition of dependence which is considerably simpler than ours (Definition 44):

> *Player i depends on player j if and only if j controls some propositional variables which are relevant for the satisfaction of i's goal.*

It is easy to see that such a definition cannot be straightforwardly generalized to the case in which players have non-dichotomous preferences, as in that case it becomes unclear what the 'goal' of player *i* actually is. In fact, this is precisely the sort of issue that we went around by proposing Definition 44. Notice that, as a consequence, the two definitions of dependence differ radically in that the one proposed by [15] and [58] views dependence structures as properties of games, while ours views dependence structures as properties of the outcomes of games.

Secondly, it is worth mentioning an underlying difference in motivation between our work and the one presented in [15] and [58]. The latter develops the analysis of

dependence relations essentially as a means to extract graphical information which eases the complexity of computing Nash equilibria in Boolean games and the core in cooperative Boolean games. What motivates our analysis instead is rather the attempt to provide a game-theoretical foundation to dependence theory as such. This lead us to consider strategic games in their generality—rather than Boolean games—and to look at dependence as a means to characterize interesting properties of games (e.g., reciprocity) and to define a specific class of coalitional games, which has been the aim of Section 5.3.

5.4.2 Open issues

The formalization of dependence relations and agreements provided here does not consider a variety of subtleties that might play a role in interaction. We list a few of them, sketching how to extend our framework in order to incorporate these more complex features.

Partial strategy permutation. Agreements are implemented by strategy permutations among stakeholders. If this operation fits perfectly games where players are endowed with a small number of strategies, such as those of Figures 2.2 and 5.4, it seems more problematic when players are endowed with a larger number of strategies. Therefore, players may be interested in favours without necessarily having to lend control of all their actions. In this purpose, it would make sense to restrict possible permutations—exchanges of favours—to subsets of the available strategies. This could be done via a function that, when applied to a game $\mathbb{G}$, yields a game identical to $\mathbb{G}$, but where profiles are restricted to the available strategies, and where the outcome function is restricted accordingly. The intuition behind restricting the game is that players decide in advance the type of strategies that they allow to be agreed upon.

AND and OR Dependence. Our definition of players' dependence allows for situations, such as the one illustrated in Figure 5.2, where a player can be simultaneously dependent on several other players, suggesting the possibility of many possible agreements. In the literature on dependence theory (cfr. [60]) this form of dependence is usually referred to as *OR dependence*, as opposed to *AND dependence*, where instead a player is dependent on the *combined strategy* of other players, i.e. a sort of dependence not on a player but on a coalition. While the first can be easily accommodated in our framework, for the latter a generalization is required, that allows a dependence relation between a player and a coalition. The informal account in [60] suggests that AND dependencies and OR dependencies have different consequences for the stability of coalitions. If a situation of AND dependence of player i on players j and k grants the latter two players a power position (as i needs both), a situation of OR dependence allows player i to choose among the possible stakeholders in a possible agreement: in some sense players profit from

OR dependencies. A suitable generalization of the definition of dependence should be able to account for this feature.

Extensive Interaction. Dependence and agreements have been formulated for strategic games, where decisions have a one-shot nature and no temporality is involved. However dependencies are naturally present in extensive interaction as well and agreements make perfect sense there. In order to analyze dependence in extensive games we could always adopt the standard translation of an extensive game into a strategic one [51]. Dependence relations and agreements can then be retrieved in the usual way, by resorting to the strategic game we have obtained. However extensive games have special features. Their typical solution concept, for instance, is that of *sub-game perfect equilibrium*, i.e. a Nash-equilibrium that rules out incredible threats [51]. What is interesting for a theory of dependence in extensive interaction is whether analogous solution concepts can be obtained for dependence relations. A generalization of the notion of sub-game perfect equilibrium to a notion of *sub-game perfect equilibrium for someone else* could be obtained as a refinement of the notion of equilibrium based on best response for someone else that we have studied here. Such refinement should take care of ruling out strategies determining incredible favours.

5.4.3 Conclusion

Our chapter has shown that a theory of agent dependence, first introduced by Castelfranchi and colleagues, can be fully incorporated within the theory of games, where it gives rise to forms of rationality that lie between the individual perspective of strategic games and the coalitional perspective of cooperative games. Concretely what we have shown can be articulated in two directions:

- First and foremost it has been shown that the intuitive notion of dependence relation originating from social and cognitive science literature [24, 20] can be fully incorporated within the theory of games, contributing to the construction of solution concepts that account for its underlying dynamics. The standard solution concepts of best response and dominant strategies first provided in Definition 3 have been generalized to best response and dominant strategy for someone else in Definition 43, providing a basis for formalizing reciprocity in games.

- Second, once the game-theoretical account has been formulated, it has been shown that central dependence-theoretic notions such as the notion of cycle have natural game-theoretical correspondents (Theorem 40). Furthermore, dependence theory has been demonstrated to give rise to types of cooperative games where solution concepts such as the core can be applied. The relation between the various forms of cooperative games where coalitions undertake agreements (dependence and partial dependence) have been analyzed, together with the dominance they induce on agreements (Theorem 41 and 42).

The results suggest the presence of a full spectrum of cooperative solution concepts for dependence structures, that form a partial order under the inclusion relation, whose further investigation poses an interesting research challenge.

The next chapter will apply to dependence games the classical logical tools for reasoning about coalitional ability.

Chapter 6

Strategic Reasoning in Dependence Games

The other way is not to argue about the assumptions at all, but to look at the conclusions only. Do our observations jibe with the conclusions, do the conclusions sound right? If yes, then that's a good mark for the assumptions. And then we can go and derive other conclusions from the assumptions, and see whether they're right. And so on. The more conclusions we have that jibe with our observations, the more faith we can put in the assumptions. That's the way that I embrace, that's good science. Logically, the conclusions follow from the assumptions. But empirically, scientifically, the assumptions follow from the conclusions!

Robert J. Aumann, *On the state of the art in game theory* [5]

6.1 Introduction

Chapter 4 has dealt with extensively one of the best known and most used formalisms for reasoning about cooperative structures, i.e. Coalition Logic [54]. Coalition Logic is a modal logic extending propositional logic with a family of modal operators $[C]\varphi$ that are intuitively read as

"the set of players C can cooperate to achieve the property φ"

Formulas of Coalition Logic reason about what holds whenever players can form coalitions, i.e. whenever they join forces to achieve a certain goal. Generally speaking, Coalition Logic — and the same holds for similar formalisms such as Alternating-time Temporal Logic (ATL) [2] and Seeing to It That (STIT) [8] — is not concerned with investigating the reasons why cooperation should be established, but it is limited to describing the logic of coalitional ability once cooperation is already in force.

However, as we have extensively discussed in the introductory section of Chapter 5, coalitions may not always be equally likely to form, for common interest in collective action may not arise.

The theory of dependence relations studied in Chapter 5 proposes itself has a weakening of the classical theory of coalitions, clarifying the reason for players to work together. A natural question arising from this viewpoint is whether modal languages can be defined, analogous to that of Coalition Logic and siblings, where the modal operator $[C]\varphi$ is read as

"the set of players C can agree to achieve the property φ"

where agreements acquire a semantics in terms of the machinery built up in Chapter 5 to talk about dependence games. Should such a link with modal logic be established, agreements could be reasoned upon in a fully logical manner, making it possible to transfer standard logical results to purely game-theoretical ground.

Nevertheless, the jump from a Coalition Logic modality interpreted on effectivity functions to one interpreted on agreements (Definition 50) requires some work: while effectivity functions are well-behaved neighbourhood structures, agreements result from a complex interaction between preference relations and strategies in dependence games, that are not a standard semantics for modal languages. True, effectivity functions can be constructed from agreements (Definition 53), but they are a mere description of outcomes that can be achieved by agreements and they would not help in making the role of dependence relations explicit.

The aim of the chapter is to reason on agreements in a Coalition Logic-like language, interpreted on effectivity functions and preferences, and based on operators that can account for individual rationality and how it changes as a consequence of agreements.

This chapter will also devote particular attention to the use of norms to *regulate* agreements, investigating what in Chapter 4 has been called the external perspective on norms, leaving the discussion of the internal one to the conclusion. Indeed, we can find many examples of agreements violating systemic properties that we recognize as desirable. Think of cartel formation, where more companies, instead of competing to lower prices, *agree* on establishing a common level of price; the aim of such collusion (also called *the cartel agreement*) is to increase individual members' profits by reducing competition. To regulate them, we use the standard labelling on outcomes through a violation constant. The newly introduced notion of agreement, confronted with this labelling, acquires a deontic reading, on top of which we can construct the semantics of the standard modal operators of permission, forbiddance and obligation.

Our dependence-based approach to the regulation of multi-agent systems is exemplified by the following story, that is also meant to recall the building blocks of our theory of dependence.

Example 27 (Strangers on a Train) *In Patricia Highsmith's novel*[1]*, Strangers on a Train [40], which Alfred Hitchcock turned in 1951 into a movie with the same title, the following story takes place:*

[1]We thank Paul Harrenstein for having brought this example to our attention.

	N	S	O
N	$2,2$	$2,0$	$9,1$
S	$0,2$	$0,0$	$0,1$
O	$1,9$	$1,0$	$8,8$

Figure 6.1: Strangers on a train.

> *Two protagonists wish to get out of an unhappy relationship. Architect Guy Haines wants to get rid of his unfaithful wife, Miriam, in order to marry the woman he loves, Anne Faulkner. Charles Anthony Bruno, a psychopathic playboy, deeply desires his father's death. On a train to see his wife, Guy meets Bruno, who proposes the idea of exchange murders: Bruno will kill Miriam if Guy kills Bruno's father; neither of them will have a motive, and the police will have no reason to suspect either of them.*

We can illustrate our protagonists' setting, before any agreements are taken, with the two persons' matrix in Figure 6.1.

In the example, both players have the same possibilities: either do nothing (N), commit the murder of their own significant other (S), or commit the murder of the other persons' significant other (O). Let Guy be the row player and Bruno the column player. Focusing on the choices of Guy, we notice that N is a *dominant strategy* for Guy (Definition 3), as whatever strategy Bruno plays, N is a best response to that strategy. For Bruno the reasoning pattern is symmetric, therefore his strategy N is also a dominant strategy. These two facts taken together mean that the strategy profile (N, N) is a *dominant strategy equilibrium* (Definition 3).

However the story takes an interesting twist once we consider what players could do for each other. The strategy O, by Guy, is a *dominant strategy for Bruno* (Definition 43), as it is good for Bruno whatever Bruno himself decides to do. Same for Guy: the strategy O by Bruno is a *dominant strategy for Guy*. Once we identify what players can do for each other, the dependence relations can be automatically drawn: Guy *DS-depends* on Bruno for strategy O and on himself for strategy N (Definition 44), while Bruno *DS-depends on Guy* for strategy O and on himself for strategy N. Dependence *cycles* (Definition 45) suggests the possibility of reciprocal play: the profile (N, N), which is associated with two dependence cycles of length 1, is *trivially DS-reciprocal* (Definition 46), i.e. the only possible way for players to agree is to play for themselves, while the profile (O, O), which is associated with an Hamiltonian dependence cycle, is *fully DS-reciprocal* (Definition 46), i.e. players can profit by playing for each other.

In this situation two *agreements* would be possible (Definition 50): $((N, N), \mu)$ and $((O, O), \nu)$, where μ is the identity permutation and ν is the players' transposition. However notice that the outcome resulting from (O, O) is preferred by both players to the outcome resulting from (N, N), which means that $((N, N), \mu)$ is a *dominated agreement* (Definiton 52), while $((O, O), \nu)$ is undominated. The latter is also *stable*

	N	S	O
N	$2,2$	$0,2$	$1,9$
S	$2,0$	$0,0$	$1,0$
O	$9,1$	$0,1$	$8,8$

Figure 6.2: Swapping murders.

as, by Theorem 41, it belongs to the core of the resulting *dependence game* (Definition 53). Therefore $((O,O),v)$ can be considered a *rational* outcome of the dependence game: Guy would find it reasonable to kill Bruno's father only if he knew that Bruno would kill his wife, and the same for Bruno. This would be possible if Guy could *lend* his action of killing *in exchange to* Bruno's one. The proposal of swapping murders, i.e. a simultaneous exchange of favours between the strangers, suggests itself. If this agreement could take place then the game would be transformed into the one pictured in Figure 6.2, the transposition of the matrix in Figure 6.1 under swap of strategies. The swap of players shown in this game DS-implements (O,O) (Definition 49).

Imposing a normative labelling onto the strangers' example is to say that some agreements can be harmful even when rational for its stakeholders. Starting from this consideration we proceed in constructing a deontic language interpreted on agreements that is able to *solve* the strangers' game declaring their swap as undesirable and that is general enough to be applicable to a large number of interactions.

Chapter Structure: The chapter, based on joint work with Davide Grossi, Jan Broersen and John-Jules Meyer [61], is structured as follows: in Section 6.2 the theory of dependence elaborated in Chapter 5 is equipped with the tools defined in Chapter 3 to build up a semantics of coalitional rationality in undertaking agreements. In Section 6.3 we build the syntax and the semantics of a logic of agreements, introducing a *switch operator* to reason about permutations of effectivity functions. In Section 6.4 the classical deontic operators are defined on agreements, and in line with Chapter 4 they are able to reason about desirable and undesirable coalitions. The concluding section discusses the open issues and the related work.

6.2 Agreements and Coalitional Rationality

In this section we elaborate a model of agreements in terms of preferences and effectivity functions. In doing so we will follow two paths: in the first (Section 6.2.1), we make use of the notion of undomination, studied in Chapter 3 as an analogue of dominant strategy in strategic games. Besides we complement it with an operation on effectivity functions, to model permuted games. In the second (Section 6.2.2) we make use of a new notion of undomination, namely an undomination *for someone else*, as an analogue of dominant strategy for someone else in dependence games.

Finally, along the lines of Theorem 40, we investigate the assumptions under which these two representations are equivalent.

6.2.1 Permuting effectivity functions

Effectivity functions have been introduced in Chapter 2 (Definition 6) as an abstract representation of coalitional power. They represent systemic properties that a coalition can achieve by cooperating, abstracting away from the process of coalitional decision making. Effectivity functions are thereby too coarse to be taken as a model of agreements among players.

To overcome this limitation, we start out by considering individual effectivity functions — effectivity functions for single players — on which to apply the transformations induced by dependence relations.

Definition 62 (Individual effectivity functions) *Given a set of players N and a set of worlds W, an* individual effectivity function *is a function $E : W \to (N \to 2^{2^W})$.*

Individual effectivity functions, that for simplicity will simply be called effectivity functions, display the individual perspective of strategic games, assigning to each player the choices that can be made at each state. In the spirit of dependence theory, the power of *groups* of players will be given by the possible agreements that they can undertake.

Let us describe the example of the strangers with individual effectivity functions.

Example 28 *Let w be a situation representing the game in Figure 6.1. In order to avoid possible confusions due to the operation of strategy permutation we identify the outcomes with their payoff vector instead of their corresponding strategy profile, i.e. we say $(2,2)$ instead of (N,N). Guy's effectivity function $E(w)(G)$ amounts to his choices in the game closed under supersets, that is*

$$E(w)(G) = \{\{(2,2),(2,0),(9,1)\},\{(0,2),(0,0),(0,1)\},\{(1,9),(1,0),(8,8)\}\}^{sup}$$

while Bruno's is

$$E(w)(B) = \{\{(2,2),(0,2),(1,9)\},\{(2,0),(0,0),(1,0)\},\{(9,1),(0,1),(8,8)\}\}^{sup}$$

For simplicity, when no ambiguity arises, we can name sets of outcomes, writing for instance $E(w)(G) = \{N,S,O\}^{sup}$. When instead ambiguity does arise we index choices with players, for instance we say N_G to indicate that doing nothing is a choice by Guy.

In the way we have defined them (Definition 50), agreements are a reallocation of strategic ability that follows a certain dependence graph (as proved in Theorem 40). In a Cooperative Game Model however we only dispose of effectivity functions and preferences. To define agreements in these models we need to endow them with an operation that permutes effectivity functions, reassigning strategic ability. We call this operation *choice switch*.

Definition 63 (Choice switch) *Let $E(w)(i)$ be a choice set of player i at world w and μ a permutation on N. Then $E'(w)(i)$ is the* choice switch *for player i at w following permutation μ if $E'(w)(i) = E(w)(\mu(i))$.*

Substantially the choice switch assigns to a player a new effectivity function, according to a given permutation. For our purposes it is useful to dispose of a global operation of choice switch, that reallocates effectivity functions according to a certain permutation. We abbreviate with $E^{\mu}(w)$ the choice set $E(w)$ constituted by the choice switches for each player i at world w according to permutation μ.

Example 29 *Let w be a situation representing the game in Figure 6.1 and let μ be a permutation on the players such that $\mu(G) = B$. Bruno's choice switch following μ at w amounts to Guy's choices in the picture, namely*

$$E(w)(\mu(G)) = E(w)(B) = \{\{(2,2),(2,0),(9,1)\},\{(0,2),(0,0),(0,1)\},\{(1,9),(1,0),(8,8)\}\}^{sup}$$

which is the effectivity function of Bruno in Figure 6.2, representing the game scenario after the agreement is taken. For Guy the result is symmetric:

$$E(w)(\mu(B)) = E(w)(G) = \{\{(2,2),(0,2),(1,9)\},\{(2,0),(0,0),(1,0)\},\{(9,1),(0,1),(8,8)\}\}^{sup}$$

Proposition 45 *The following property holds:*

- *$E^{\mu}(w)$ preserves outcome monotonicity, regularity, superadditivity, for coalitions made by individual players.*

Proof *It is sufficient to notice that the listed properties are formulated for choice sets of any two players in case of regularity and superadditivity and, in particular, they hold for any permutation. For outcome monotonicity the same argument can be used.*

A permuted individual effectivity function encodes a sort of candidate agreement, i.e. a possible reallocation of players' strategic ability that does not take preferences into account. To obtain a proper agreement we need to identify the undominated choices for each player at each permutation, i.e. what the players find it rational to achieve if they could choose for someone else.

Definition 64 (Agreements and permuted games) *Let E be an effectivity function on W, $C \subseteq N$ a coalition, $\mu : C \to C$ a permutation, $w \in W$ a state defined on a given Coalitional Game Model M. A tuple $(\bigcap(X_i)_{i\in C}, \mu)$ with $X_i \in E(w)(i)$ is said to be an* agreement *for coalition C at world w if*

- *$X_i \rhd_{\mu^{-1}(i),w}$ in $E^{\mu}(w)$.*

The definition says that an agreement results from an exchange of strategies of individual players that are individually rational for the players receiving them. More specifically, the agreement is made by a set X that is an intersection of sets indexed by the players, and a permutation on the players. Each part of this set is an undominated choice of player i in the effectivity function of the player j indicated by the permutation.

The definition mimics the features of *DS* partial agreements of Definition 50, as:

- it is defined for a coalition and not necessarily for all the players;
- it adopts undominated choices ($X_i \rhd_{\mu^{-1}(i),w}$), analogues of dominant strategies in coalitional games.

Let us observe how this works in our example.

Example 30 *From the results of Chapter 3 (Proposition 15) we know that the choice of doing nothing by Guy and by Bruno are undominated choices in the effectivity function obtained from the game in Figure 6.1. This is because doing nothing, i.e. the profile* (N, N) *in the game, is a dominant strategy equilibrium . Once however the effectivity functions are permuted, dominant strategy equilibria also change. In the game of Figure 6.2, the choice to kill the other's significant other (the profile* (O, O)*) is now a dominant strategy. Consequently, making use of Proposition 15, the choice of doing nothing is undominated for each player and it is thereby an agreement.*

Agreements, formulated as undominated choices, inherit several properties typical of undomination. The most representative one is that of monotonicity, and its validity is shown by the following proposition.

Proposition 46 *Let* $(\bigcap(X_i)_{i\in C}, \mu)$ *be an agreement for coalition* C *at a given state* w*. Then each* (Y, μ) *such that* $\bigcap(X_i)_{i\in C} \subseteq Y$ *is an agreement for coalition* C *at* w*.*

Proof *By outcome monotonicity of effectivity functions and by the definition of the set* Y*,* Y *is such that* $Y = \bigcap(Y_i)_{i\in C}$ *for* $Y_i \in E(w)(\mu(i))$*. By monotonicity of Pareto optimality of choices (Proposition 9) we also have that* $Y_i \rhd_{\mu(i),w}$ *in* $E(w)(\mu(i))$*. This is enough to conclude, following Definition 64, that* (Y, μ) *is an agreement.*

As anticipated, what we have just described is one of the two possible ways to understand agreements, here seen as the conjunction of undominated choices in permuted effectivity functions. An alternative way is the generalization of the standard notion of undomination to undomination for someone else, which will be pursued in the coming section.

6.2.2 Coalitional rationality for someone else

A different way of formalizing agreements is suggested by the definitions of best response and dominant strategy *for someone else* that we have used in dependence games (Definition 43). Instead of permuting effectivity functions we permute preferences, generalizing undominated choices to undominated choices *for someone else*.

As we can recall from Definition 22 undominated choices have two constituents:

- Pareto optimality, i.e. undominated choices are $(\forall, \forall)$ Pareto optimal choices;
- Choice restriction, i.e. undominated choices preserve optimality in every choice restriction.

In order to define undominated choices for someone else we need to generalize the part of the definition concerning preferences.

To this end we define Pareto optimal choices for someone else, that select maxima in one's order of choices. But unlike its standard definition (Definition 19), the maxima are considered in someone else's preference order. Once again, we limit ourselves to the $(\forall, \forall)$ preference lifting. Henceforth we simply write $\succeq_i$ ($\succ_i$) for $\succeq_i^{(\forall,\forall)}$ ($\succ_i^{(\forall,\forall)}$).

Definition 65 (Pareto optimal choice for someone else) *Let E be an effectivity function, $i, j \in N$ two players, $w \in W$ a state and $X \in E(w)(i)$ a set in i's effectivity function. X is* Pareto optimal choice *by i for j (in symbols $PO_{(i \hookrightarrow j)}$) at w if, and only if, for no $Y \in E(w)(i)$, $Y \succ_j X$.*

The definition says that Pareto optimal choices for someone else are those choices in an individual effectivity function such that no better choice exists *for another given player*. Despite their name, Pareto optimal choices for someone else become standard Pareto optimal choices, i.e. for *oneself* , in case i and j coincide. Moreover, they inherit all the properties of Pareto optimal choices described in Paragraph 3.2.2. In particular they inherit monotonicity which, together with the adoption of the $(\forall, \forall)$ preference lifting, turns them into rather weak constructs. Let us have a look at Pareto optimal choices for someone else in the example.

Example 31 *In Figure 6.1 the choice N and the choice O are Pareto optimal choices by all players for themselves. As a consequence of outcome monotonicity of Pareto optimality, we have that the only choice that is not individually optimal is S, both for Guy and for Bruno. This simply means that the only choice that the strangers do not like in an absolute sense is to kill their own significant other. Pareto optimality for the other player is even less informative: all three choices for both players are Pareto optimal for the other. Once again, Pareto optimality does not represent what players should rationally do taking the opponents into account, but what they should do in an absolute sense. If a set X in an effectivity function is Pareto optimal then there is no other set Y such that all its elements are better than all the elements in X.*

In line with our considerations for the standard definition of Pareto optimality (Definition 19), the example suggests that the mere use of Pareto optimality of choice cannot provide a good characterization of individually rational choice, and even less of rational choice for someone else. Once again the limitations of Pareto optimality can be overcome by undominated choices. Here the intuition is that a choice is *undominated for player j* if it is Pareto optimal for j no matter what the other players decide to do. This is the formal definition:

Definition 66 (Undomination for someone else) *Let E be an effectivity function, $i, j \in N$ two players, $w \in W$ a state and $X \subseteq W$ a set. X is an* undominated choice *by i for j in w (in symbols $X \rhd_{i \hookrightarrow j, w}$) if and only if*

1. $X \in E(w)(i)$

2. *for all* $Y \in \bigcap E(w)(k)$ *with* $k \neq j$, $X \cap Y$ *is Pareto Optimal for* j *in* $E(w)(i) \sqcap Y$.

The definition says that for a choice X in the effectivity function of player i to be undominated for player j two conditions need to be satisfied: the first (item 1) that X is really a choice available to player i and the second that there is no better choice for player j available to player i (item 2).

If we recall Definition 22 for standard undominated choice we immediately notice the two differences: effectivity functions are not coalitional but individual and Pareto optimal choices is defined with respect to some other player.

Let us illustrate undominated choices for someone else in our motivating example.

Example 32 *In the effectivity function representing the game in Figure 6.1 the choice of doing nothing (i.e. N) is an undominated choice by each player for himself, while it is not in the effectivity function representing the game in Figure 6.2, where instead the choice of killing the other's significant other (i.e. O) is undominated by each player for himself. However if we not only want to look at individual rationality, but also at what players could do for the others, we need to resort to undomination for someone else: the choice O in Figure 6.1 is an undominated choice by each player for its opponent and the outcome* (O, O)*, resulting from both players helping each other can already be seen as a possible agreement which both players can give rise to.*

The example has made clear how favours, so central for the treatment of agreements, can be naturally incorporated in our framework: i depends on j for a choice X if j's strategy in X is a favour for i or, said formally, is undominated choice by j for i.

Definition 67 (Agreements and reciprocity) *Let* E *be an effectivity function on* W, $C \subseteq N$ *a coalition,* $\mu : C \to C$ *a permutation,* $w \in W$ *a state defined on a given Coalitional Game Model* M*. A tuple* $(\bigcap(X_i)_{i \in C}, \mu)$ *with* $X_i \in E(w)(i)$ *is said to be an* agreement *for coalition* C *at* w *if*

- $X_i \rhd_{(i \hookrightarrow \mu^{-1}(i), w)}$.

The definition says that an agreement is a set of choices for members of a coalition that are rational for some other member for that coalition. Notice the difference with the previous definition of agreements for permuted games (Definition 64): agreements are also given by the intersection of sets in players' effectivity function plus a permutation. But the permutation is now applied to preferences while in Definition 64) it is applied to effectivity functions.

We have now two definitions of agreement, the one in Definition 64 and the other in Definition 67. The following proposition shows that these two definitions are in fact equivalent.

Proposition 47 *Let* E *be an effectivity function on* W, $C \subseteq N$ *a coalition,* $\mu : C \to C$ *a permutation,* $X \subseteq W$ *a set of outcomes,* $w \in W$ *a state defined on a given Coalitional Game Model* M*. The tuple* $(\bigcap(X_i)_{i \in C}, \mu)$ *with* $X_i \in E(w)(i)$ *is an agreement for* C *at* w *in the sense of Definition 64 if and only if it is an agreement for* C *at* w *in the sense of Definition 67.*

Proof *It follows from the fact that* $X_i \rhd_{(\mu^{-1}(i),w)}$ *in* $E^{\mu}(w)$ *is equivalent to* $X_i \rhd_{(i \hookrightarrow \mu^{-1}(i),w)}$ *in* $E(w)$.

The two ways of formalizing agreement with effectivity functions are now fully disentangled and we can move on to their logical analysis.

6.3 A Logic for Agreements

In this section we introduce the syntax and the models for a modal language to reason about agreements, providing a semantics to relate them. The language, which we call $\mathcal{L}^{\leq,[i],\downarrow,sw}$, is an extension of propositional logic, with modalities to talk about preferences, single player coalitions, single player choice restriction and permutation of effectivity functions. With a few relatively small extensions, the logical language presented in Chapter 4 to reason on undominated choices, turns out to be flexible enough to express dependence relations, and also agreements.

Definition 68 (Syntax) *Let Prop be a countable set of atomic propositions. The formulas of* $\mathcal{L}^{\leq,[i],\downarrow,sw}$ *have the following grammar:*

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [i]\varphi \mid A\varphi \mid \Diamond_i^{\leq}\varphi \mid [i \downarrow \varphi]\psi \mid [sw]\varphi$$

where $p \in Prop$ and sw is a permutation on N. The informal reading of the modalities is "player i can achieve φ", "φ is globally true", "there is a better world than the current one for player i that satisfies φ", "after player i choses φ, ψ holds", "permuting effectivity functions according to sw, leads to φ".

The language is equipped with modalities to formalize both the agreements that involve the permutation of the effectivity function — via the modality $[sw]$, that reasons on the consequences of effectivity functions permutation — and the agreements that involve undomination for someone else — via the modalities $[i]$ and $\Diamond_i^{\leq}$, that reason respectively about the strategic ability of individual players and their preferences.

We first concentrate on the first approach to agreement, seeing them as equilibria in permuted interactions (following Theorem 40). Permuted interactions can be expressed using the operator $[sw]$.

The operator $[sw]$, the *switch* operator

The operator $[sw]$ accounts for the transformation in a model induced by permuting players' effectivity functions. In the same way we have done with the subgame operator (Definition 32) its interpretation is nonconstructive. Each world w has an outcoming arrow labelled with a permutation μ on players that goes to another world w' that is equivalent to w as for valuation function but differs for the players' effectivity functions, that are reallocated according to μ.

A different path could be taken, more alligned with the model update perspective of Dynamic Epistemic Logic discussed in the concluding section of Chapter 4, that

transforms each tuple model-world M, w into a different tuple M', w, interpreting the switch operator in a sort of *metamodel*. For the sake of uniformity with the subgame operator, the nonconstructive interpretation has been chosen, while the latter approach is discussed in the concluding chapter.

Definition 69 (Switch)

$$M, w \models [sw]\varphi \text{ if and only if } M, (sw, w) \models \varphi$$

The updated world (sw, w) is identical to w in all features apart from the effectivity function, which is interpreted as follows:

Definition 70 (Updated worlds for switches)

$$E((sw, w))(i) \doteq E(w)(j) \quad \text{if} \quad sw(i) = j$$

The clause regulating the update deserves a short comment. It says that updating a world means updating its effectivity function, following the given permutation. In other words, if player j had choice set $\mathcal{Y}$ at world w, then at world (sw, w) player i will have $\mathcal{Y}$ whenever $sw(i) = j$. In turn the set $\mathcal{X}$ held by player i at w will be assigned at (sw, w) to player $sw^{-1}(i)$.

As for the case of the subgame operator, coalition frames are special frames that are *closed under players permutations*. The closure can be made precise in the following way.

Definition 71 (Closure under players permutations) *Let $w \in W$ be a world,(sw, w) its update according to permutation sw, and $F = (W, E)$ be a coalition frame. F is said to be* closed under players permutations *if and only if $w \in W$ implies that $(sw, w) \in W$.*

As for the closure under subgames, it is a frame condition that can be formally characterized.

Proposition 48 *Let $F = (W, E)$ be a coalition frame. The following holds:*
$F \models \langle sw \rangle \top$ if and only if F is closed under players permutations.

Proof *From right to left, it is straightforward. From left to right assume $F \models \langle sw \rangle \top$. Consider now a world $w \in W$ and consider any permutation $sw : N \to N$. We must have that $(sw, w) \in W$.*

It is worth noticing that the switches we consider are total, while much attention has been dedicated to partial agreements, that are instead based on partial permutations (as in Definition 50). We shall see that, exploiting the features of outcome monotonicity of effectivity function and some other mild assumptions, notions analogous to partial agreements can be defined even when using total permutations.

	Axioms
A1	$[sw]p \leftrightarrow p$
A2	$[sw]\neg\varphi \leftrightarrow \neg[sw]\varphi$
A3	$[sw](\varphi \wedge \psi) \leftrightarrow ([sw]\varphi \wedge [sw]\psi)$
A4	$[sw][k]\varphi \leftrightarrow [sw^{-1}(k)]\varphi$
A5	$[sw]\Box_i^{\preceq}\varphi \leftrightarrow \Box_i^{\preceq}\varphi$
A6	$[sw'][sw]\varphi \leftrightarrow [sw' \circ sw]\varphi$
	Rules
R1	$\varphi \Rightarrow [sw]\varphi$

Table 6.1: Axioms and rules for the switch operator

6.3.1 Validities

The switch operator shares many structural features with the subgame operator. The most fundamental one is the presence of reduction axioms: also in this case the introduction of the subgame operator does not add expressive power to the language provided the models are closed under players permutations.

Proposition 49 (Reduction Axioms) *The axioms and the rules displayed in Table 6.1 are valid in Coalition Models.*

A proof is to be found in the appendix (Section B.2).

To see more clearly how the reduction works it can be observed that any formula with the switch operator occurring in it can be eventually rewritten as a formula without the switch operator occurring in it, preserving validity. Similar arguments are used in dynamic epistemic logics [69].

6.3.2 Characterization results

The coming results essentially concern the characterization power of the language with respect to the notions defined at the structural level. With these characterization results, which generalize the ones in Chapter 4 to rational choice for someone else, we can make use of the logical language to express and reason about complex interactions between preferences and choices in interdependence.

To start with, Pareto optimal choices for someone else, introduced in Definition 65, can be characterized within the language provided in Definition 68.

Proposition 50 *φ^M is Pareto optimal choice by i for j in w if and only if $M, w \models [i]\varphi \wedge \langle i\rangle\Diamond_j^{\leq}\varphi$*

The proof is a generalization of the one given in Chapter 4 (Proposition 23). We only show one direction of the proof, to give the flavour of how the generalization works. The other direction follows the same pattern of the proof of Proposition 23.

Proof ($\Rightarrow$)

Let us assume that φ^M is Pareto optimal choice by i for j in w , i.e. that φ^M is a Pareto optimal choice for player j at world w in $E(w)(i)$ according to the $(\forall,\forall)$ preference lifting. This means, by Definition 65, that for no $X \in E(w)(i)$, $X \succ_j^{(\forall,\forall)} \varphi^M$ and that $\varphi^M \in E(w)(i)$. In turn this means that for all $X \in E(w)(i)$ $\exists x \in X, \exists y \in \varphi^M$, such that $x \preceq_j y$. By the definition of effectivity functions, no set $X \in E(w)(i)$ is such that $X \subseteq (\neg\Diamond_j^{\leq}\varphi)^M$. So we can conclude that $M, w \models [i]\varphi \wedge \langle i\rangle\Diamond_j^{\leq}\varphi$.

Proposition 50 shows that saying that a choice φ is Pareto optimal for j boils down to saying that it can be performed by a player (i.e. $[i]\varphi$) and that the player cannot avoid ending up in a world that is worse for j than some φ world (i.e. $\langle i\rangle\Diamond_j^{\leq}\varphi$).

We know from Chapter 3 that Pareto optimal choices are particularly weak constructs that can however be refined by taking the opponents into account. Chapter 4 has moreover shown that the opponents' possibilities can be made formal by using the subgame operator (Section 4.2.2). In the present case its use, together with the previous result, makes for the possibility of characterizing the notion of undominated choice for someone else.

In the same fashion as what we have done with the notion of undominated choice (Proposition 31 and following ones) we put forward a variety of characterization results for undominated choice for someone else, where the generalizations apply as sketched for the case of Pareto optimal choices.

Proposition 51 *Let $\mathbb{F}$ be the class of Cooperative Game Frames with individual effectivity functions closed under subgames and let $F \in \mathbb{F}$ be one of them. Let moreover $E(w)(\bar{i}) = \bigcap E(w)(j)$(for $i \neq j$) be a set of sets obtained by superadding the choice sets of all opponents of player i. The following holds:*

$F \models [i]\varphi \rightarrow [\bar{i} \downarrow \psi]([i](\varphi \wedge \psi) \wedge \langle i\rangle \bigvee_i \Diamond_j^{\leq}(\varphi \wedge \psi))$

if and only if each $X \in E(w)(i)$ is such that X is undominated choice by i for j at w

The proof is a straightforward generalization of the one for Proposition 31, and it allows for similar observations: i) in characterizing undomination as a property of the frames, we do not need any restriction on the choices of coalitions; ii) we can characterize a much finer notion of undomination and Pareto optimality of choice: we can talk about all sets in an effectivity function, and not only those that are the truth set of some proposition.

If instead we would like to characterize undomination for someone else at the model level, we need some more restrictive assumptions, namely finiteness of effectivity functions.

Proposition 52 *Let $PO_{i\hookrightarrow j}\varphi$ abbreviate the formula characterizing the fact that φ is a Pareto optimal choices by i for j and let $\{\psi_1, ..., \psi_n\} = E(w)(\bar{i}) = \bigcap E(w)(j)$(for $i \neq j$) be the effectivity function of i's opponents. The following holds:*

$\varphi^M \rhd_{i\hookrightarrow j,w} \Leftrightarrow M, w \models \bigwedge_{\psi_i \in \{\psi_1,...,\psi_n\}} [\bar{i} \downarrow \psi_i] PO_{i\hookrightarrow j}(\varphi \wedge \psi_i)$

The proof is, once again, the generalization of the corresponding one for the rational choice by a player for himself (Proposition 30). In the same line of that proposition it shows that with finite effectivity functions, undomination for someone else can be written as a finite conjunction of formulas that make use of the subgame operator and Pareto optimality for someone else. In other words it says that an undominated choice for someone else is a Pareto optimal choice for someone else in every choice restriction. As the latter ones are finitely many a finite conjunction is sufficient to express the formula in the language.

The coming part will characterize agreements inside the language, using all the machinery that we have introduced so far. It is moreover convenient, to shorten notation, to abbreviate the syntactical correspondents of $\varphi^M \rhd_{i\hookrightarrow j,w}$ characterized in the previous propositions as $[rational_{(i\hookrightarrow j)}]\varphi$.

Characterizing agreements

As anticipated, the introduction of the switch operator in the framework makes it possible to characterize agreements without explicitly defining modal operators capturing rationality for someone else. We carry out the characterization assuming finiteness of effectivity functions and the following definition will ease the presentation of the result.

Definition 72 $\mathcal{A}_C \bigwedge_{i\in C} \varphi_i := \bigvee_{C\in\mathcal{P}(sw)} [sw] \bigwedge_{i\in C} [rational_{(i\hookrightarrow i)}]\varphi_i$

Recall that $\bigvee_{C\in\mathcal{P}(sw)}$ means that the coalition C is a union of orbits of the cycle induced by the permutation sw on N (as in Example 1). This definition draws in a formal language what a set of players can agree upon: it says that a coalition can agree on $\bigwedge_{i\in C} \varphi_i$ whenever there is a coalition C that can generate $\bigwedge_{i\in C} \varphi_i$ as a partial agreement. Notice that the coalitional ability is defined in terms of a conjunction of individually rational actions, which in turn quantify over all possible choices of one's opponents.

The syntactical and the model theoretical definition can now be related.

Proposition 53 *Let M be a finite Coalitional Game Model closed under subgames and players permutations. We have the following:*

$M, w \models \mathcal{A}_C \bigwedge_{i\in C} \varphi_i \Leftrightarrow$
there exists a permutation μ on C such that $(\bigcap(\varphi_i^M)_{i\in C}, \mu)$ is an agreement for C in w.

Proof *The result follows from Definition 72, Definition 64, Definition 66 and Proposition 52.*

Using Theorem 47 also the following result is straightforward, providing an alternative characterization of agreements in terms of undominated choices for someone else without the switch operator.

Proposition 54 *Let M be a finite Coalitional Game Model closed under subgames and players permutations. We have the following validity:*
$\mathcal{A}_C \bigwedge_{i\in C} \varphi_i \leftrightarrow \bigvee_{C\in\mathcal{P}(sw)} \bigwedge_{i\in C} [rational_{(i\rightarrowtail sw(i))}]\varphi_i$

The series of syntactic expressions characterizing agreements has shown that the language is powerful enough to account for transformations of players' strategic abilities following reciprocity cycles. The next section will label these transformations in a deontic logic fashion, aiming at pointing to the desirable ways of forming coalitions via agreements.

6.4 Deontic Operators

In the previous chapter two views of norms have been put forward: an internal perspective, following the viewpoint of [41], that interprets norms as statements concerning coalitional rationality; and an external perspective, following the viewpoint of [48], that interprets norms as statements concerning systemic rationality. Our motivating example clearly emphasizes the external-systemic perspective, as it describes a rational agreement going against desirable properties. On these grounds the chapter will be focused on the external view. The section of related work will briefly present the internal view of norms, more in line with the treatment in [41].

Along with the external view, outcomes will be labelled in accordance to their deontic status and permutations will be judged against this labelling as follows:

- Permutations are forbidden if leading to undesired outcomes (violations);
- Permutations are permitted if not forbidden;
- Permutations are obliged in case all the other possible permutations are forbidden.

The resemblance of the present definition with the one given in Section 4.3 for norms on coalitional choices shows that agreements are treated as one possible coalitional choice, and their regulation is inserted in a more general framework. However there is a notable point of difference: coalitional choices are sets of states, while agreements are sets of states endowed with a permutation on players. What is more, the latter may be defined on a subset of the set of players, giving rise to partial agreements.

To bridge the gap we will exploit outcome monotonicity of effectivity functions. We know that if a set $X \subseteq viol^M$ in some model M belongs to the effectivity function

of some player i at some world w then the set $viol^M$ does as well. In other words if a player can make a choice that, no matter how the other players choose, will lead to a state in $X \subseteq viol^M$ then it can also make a choice that, no matter how the other players choose, will lead to a violation.

Making use of this feature, we can apply the standard deontic operators to permutations.

Definition 73 (Deontic Operators on Agreements) *Let $PERM_N$ be the set of all permutations on N and let $sw \in PERM_N$. The operators $F(sw), P(sw), O(sw)$ indicate forbiddance, permission and obligation as follows.*

$$\begin{aligned} F(sw) &:= [sw] \bigvee_{i \in N} \neg [rational_i] \neg viol \\ P(sw) &:= \neg F(sw) \\ O(sw) &:= \bigwedge_{sw' \in PERM_N \neq sw} F(sw') \end{aligned}$$

Norms are here used to label players permutations. A permutation sw is forbidden if after the corresponding switch for some player the set $\neg viol^M$ is not a rational choice, it is permitted if it is not forbidden, and it is obligated if all other permutations are forbidden.

The operator $F(sw)$ and $\mathcal{A}_C$ of Definition 72 show a form of duality. The correspondence between the two will turn out to be even stricter when forbiddance is applied to coalitions and not to permutations only. For now we can show some relation between the two. The following proposition states that if some permutation is forbidden then the players together can cooperate to achieve an undesirable state.

Proposition 55 *Let F be a finite Coalitional Game Model closed under subgames and players permutations. The following holds: $F \models (\bigvee_{sw \in PERM_N} F(sw)) \rightarrow \mathcal{A}_N viol$*

Proof *Assume $M, w \models F(sw)$ for arbitrary M, w and for some permutation sw on N, that is to say $M, w \models [sw] \bigvee_{i \in N} \neg [rational_i] \neg viol$. By the interpretation of the modal operators, there is a player $sw^{-1}(k)$ for which $(\neg viol)^M$ is not an undominated choice, i.e. for each of them there is a set $X \in E(w)(sw^{-1}(k))$ for which $X \succ_{sw^{-1}(k)} (\neg viol)^M$. A fortiori $X \subseteq viol^M$ and by outcome monotonicity $viol^M$ is undominated. As by outcome monotonicity $\top^M$ is undominated, too, for all $j \neq sw^{-1}(k)$, $viol^M$ is a possible agreement of N. In other words $M, w \models \mathcal{A}_N viol$.*

The following proposition states that if some permutation is permitted then the players together can cooperate to achieve a desirable state.

Proposition 56 *Let F be a finite Coalitional Game Frame closed under subgames and players permutations. The following holds:*
$F \models (\bigvee_{sw \in PERM_N} P(sw)) \rightarrow \mathcal{A}_N \neg viol$

Proof *It follows the same pattern of the previous result.*

In both cases the converse does not hold, as $viol^M$ can be identical with the whole domain or N may not be able to agree upon a desirable property.

The validities in this section have shown that the desirability of a potential agreement — as well as its undesirability — always have some implications in terms of rational action. In particular Proposition 55 states that if some potential agreement is undesirable the grand coalition can rationally choose an undesirable state, while Proposition 56 states that if some potential agreement is permitted the grand coalition can rationally choose a desirable state.

The next section will lift these operators from permutations to coalitions.

6.4.1 A deontic logic for coalition formation

Speculating on the results of the choices that can be agreed upon by a certain coalition, it is immediate to apply the deontic statements to coalitions themselves. The idea is that coalition C is forbidden to form if and only if all the agreements it can give rise to might not lead to a desirable outcome.

Definition 74 (Deontic Operators on Coalitions)

$$\begin{aligned} F(C) &:= \bigwedge_{C\in\mathcal{P}(sw)}[sw]\bigvee_{i\in C}\neg[rational_i]\neg viol \\ P(C) &:= \neg F(C) \\ O(C) &:= F(\overline{C}) \end{aligned}$$

The operator $F(C)$ says, as anticipated, that a coalition C should not form if all agreements it can give rise to might not lead a desirable outcome; it is permitted when it is not forbidden and it is obligated when the opposite coalition is forbidden.

Notice that the expression $\bigwedge_{C\in\mathcal{P}(sw)}[sw]\bigvee_{i\in C}\neg[rational_i]\neg viol$ due to the assumption of finiteness of choices of coalitions can be described within the language. The following reveals the intimate relation between the newly defined forbiddance operator and the agreement modality:

Proposition 57 *The following is a validity of any finite Coalitional Game Frame:*
$F(C) \leftrightarrow \neg\mathcal{A}_C\neg viol$

Proof *From left to right, take an arbitrary M, w such that $M, w \models F(C)$. By definition of $F(C)$, $M, w \models \bigwedge_{C\in\mathcal{P}(sw)}[sw]\bigvee_{i\in C}\neg[rational_i]\neg viol$. This means that for all permutations sw for which $sw(C) = C$ there is some player $sw(k) \in C$ for which $(\neg viol)^M$ is not undominated in $E(w)(k)$, which in turn means that there is a set $X \in E(w)(k)$ such that $X \succ_{sw(k)} (\neg viol)^M$. (Notice on the fly that $X \subseteq viol^M$.) But this means that $M, w \not\models \mathcal{A}_C\neg viol$, i.e. $M, w \models \neg\mathcal{A}_C\neg viol$. From right to left, the proof is similar.*

The previous proposition states that forbidding a coalition is equivalent to stating that that coalition cannot avoid agreeing on an undesirable property. The following section is devoted to applying the full-blown modal apparatus we have introduced to the example of the strangers in the train.

6.4.2 Colouring strangers

The deontic operators defined in terms of agreements can be fruitfully used to succinctly reason on the relevant properties of strategic interaction. This section makes use of the characterization results obtained so far to reason on the interaction of Figure 6.1. The same type of reasoning can be extended to all interactions that can be described using single player effectivity functions.

Proposition 58 *Let M, w be a representation of the game in Figure 6.1. Let us assign the atomic proposition viol to hold in the outcome (O, O). For G, B being Guy and Bruno, $x \in \{G, B\}$, O, N, S the respective choices, the following formulas hold in M, w:*

$\neg[rational_x]O$	*players do not find it rational to kill the other player's significant other*
$\neg[rational_x]S$	*players do not find it rational to kill their own significant other*
$[rational_x]N$	*players do find it rational not to kill anyone*
$[(G,B),(B,G)][rational_x]O$	*players can agree to kill each other's significant other*
$F((G,B),(B,G))$	*it is forbidden to swap murders*
$O((G,G),(B,B))$	*it is obligatory not to swap murders*
$P((G,G),(B,B))$	*it is permitted not to swap murders*

The deontic operators precisely identify the transformations of the game structure leading to desirable and to undesirable consequences.

6.5 Discussion

6.5.1 Related work

To our knowledge, what is studied in the present chapter is the first attempt to give a dependence-theoretic semantics of deontic operators. The effort is somewhat related to what done in chapter 4, that has formulated a semantics of norms in terms of effectivity functions. There two perspective have been taken:

- the internal one, inspired by the work in [41] and [44], treating the notion of what a coalition ought to do in terms of what is rational for that coalition to do.

- the external one, inspired by the work in [48], treating the notion of what ought to be done in terms of what actions comply with predetermined systemic desiderata; or said otherwise, actions not leading to violations.

The present chapter has given a dependence-theoretic semantics for deontic operators *only following the external view*. The focus of the analysis was not to understand the norms and values induced by agreements, for instance what becomes obligatory after a contract is signed, but how agreements could be labelled if we compare them against predetermined properties that we have set up in the beginning. In other words, the present chapter has been focused on the regulation of multi-agent systems and not on the study of the normative stances arising from agreements.

It is then natural to ask what the internal side of the coin looks. We introduce a notion of contract arising from agreements, analogous to the one in Definition 35. Contracts are here viewed as a set of obligations, forbiddances and permissions applied to agreements.

Definition 75 (Contracts) *Let $i_1, \ldots, i_n$ be a set of players. The following operators define norms resulting from rational agreements.*

$$
\begin{aligned}
F_{\mathcal{A}}(i_1 : \varphi_1, \ldots, i_n : \varphi_n) &:= ([i_1]\varphi_1 \wedge \ldots \wedge [i_n]\varphi_n) \rightarrow \neg \mathcal{A}_{\bigcup\{i_1,\ldots,i_n\}} \bigwedge_{1 \leq k \leq n} \varphi_{i_k} \\
P_{\mathcal{A}}(i_1 : \varphi_1, \ldots, i_n : \varphi_n) &:= \neg F_{\mathcal{A}}(i_1 : \varphi_1, \ldots, i_n : \varphi_n) \\
O_{\mathcal{A}}(i_1 : \varphi_1, \ldots, i_n : \varphi_n) &:= F_{\mathcal{A}}(i_1 : \neg\varphi_1, \ldots, i_n : \neg\varphi_n)
\end{aligned}
$$

What the definition says is that a set of formulas, one for each player, are forbidden for those players if they are not an agreement whenever they can be executed. They are instead permitted if they are not forbidden and they are obligated if the negation of those formulas is forbidden for the respective player.

Notice that forbiddance, permission and obligation are formulated *simultaneously* for the agreement's stakeholders. In other words contracts behave as policies in Definition 35.

The newly defined notion can be applied to our initial motivating example, obviously permitting the swap of the strangers, as killing each other's significant other is a dominant strategy equilibrium of the permuted game. Hereby we notice once more the conflict between the two related deontic perspectives, one indicating what it is rational to do and the other indicating what it is required to do.

6.5.2 Open Issues

The main issue left open by the present chapter is of a logical nature, and it concerns the semantics of the switch operator. From its definition we can observe that the operator, introduced to reason on the permutations of effectivity functions, is interpreted by making use of functions that update worlds, similarly to what done for the subgame operator. However a substantial difference between the two

operators can be observed. If the reduction axioms for the subgame operator did not behave properly with a model update semantics (because of the high expressivity of the global modality), this is not the case with the switch operator.

We can namely introduce a semantics for the switch operator that updates the coalition models and not the worlds, and that validates the axioms in Table 6.1. Here is the new interpretation:

Definition 76 (Switch with Model Update) *Let $M = (W, E, V)$ be a coalition model. The operator $[sw]$ is interpreted as follows:*

$$M, w \models [sw]\varphi \text{ if and only if } M|_{sw;w}, w \models \varphi$$

The updated models $M|_{sw;w}$ are of the form $M|_{sw;w}, w = \langle W, E|_{sw;w}, \preceq_i, V\rangle$ where the only modified element, the effectivity function $E|_{sw;w}$ is defined as follows:

Definition 77 (Updated Models for Switches)

$$E|_{sw;w}(w)(k) \doteq E(w)(j) \quad \text{for} \quad (k, j) \in sw$$

$$E|_{sw;w}(w')(k) \doteq E(w)(k) \quad \text{for} \quad w \neq w'$$

Notice that the last condition corresponds to a sort of locality of updates, changing only the effectivity function at the current world. The axioms, proved in the appendix for the original semantics, can be also proved for the present one, via an immediate adaptation of the proofs.

This phenomenon poses two interesting research questions:

- What kind of expressivity of the modal language is needed to distinguish between the two semantics?
- For what class of update operators are the two semantics equivalent?

In attempting to answer these questions, the first observation that can be made concerns the role played by the global modality. In the nonconstructive case the models are closed under taking players permutations (or under subgames), as a consequence saying that the formula φ is globally true means that is true in the original effectivity functions *and in the updated ones*. This is not the case in the constructive semantics, where saying that a formula φ is globally true means that it is true in the model, but it says nothing about its updates. It is no surprise that the reduction axioms for the subgame operators, containing the global modality, do not work in the constructive case. As a further evidence, they do work in the case of the switch operator, that need not use the global modality to characterize its updates.

6.5.3 Conclusion

The contribution of the chapter consists in developing a modal logic to express dependence relations as first formalized in [35]. To that we add the machinery of

deontic logic, in order to discriminate between agreements that do and agreements that do not reach some desirable properties set up in the beginning.

Unlike the standard logics to reason about coalitionally rational action, such as ATL, STIT or CL, the capacity of a set of players to take a rational decision have been restricted to what we have called *agreements*, and formalized as a transformation of the interaction structure that exchanges *favours*, i.e. choices that are rational for someone else, among players.

Our language is based on the one we have studied in Chapter 4, which extends Pauly's Coalition Logic with preferences, to account for undominated choices. We generalize the notion of undominated choice to that of undominated choice for someone else and we consequently generalize all related characterization results. We introduced an explicit operator to talk about effectivity function permutations and showed a reduction result to the language without this operator.

The deontic language has allowed us to identify in concise terms those agreements that act accordingly or disaccordingly with the desirable properties set up in the beginning, and has revealed, by logical reasoning, a variety of structural properties of this type of collective action.

Chapter 7

Conclusion

> *Most readers will by now have understood that, in my view, scientific theories are not to be considered "true" or "false". In constructing a theory, we are not trying to get at the truth, or even to approximate to it: rather, we are trying to organize our thoughts and observations in a useful manner.*
>
> Robert J. Aumann, *What is game theory trying to accomplish?* [6]

The overall aim of this work has been that of making the structure of coalitional rationality explicit, emphasizing the reasons for self-interested individuals to join forces and achieve a common goal. Chapter 3 and Chapter 4 have been concerned with a *classical* representation of coalitional games, studying a notion of coalitional rationality based on a standard model of coalitional power, the so-called effectivity functions, and an order on it, obtained by lifting the preferences of the individuals involved in the decision. Chapter 5 and 6 have added more structure to those models, introducing the concept of dependence among individuals, and studying it as a precondition for coalition formation.

The introductory chapter has put forward a number of research questions aimed at understanding the structural and logical features of coalitional rationality. Those questions have been answered in the following way:

Coalitions and rationality In cooperative game theory abstract models of coalitional choices and preferences are adopted, but no solution concept is studied that, similarly to what done in non-cooperative game theory, identifies the best choices for a given coalition. Chapter 3, building upon these models, has studied a preference order over coalitional choices that classifies choices according to what the members of the coalition prefer and how the individuals outside the coalition can react. This order has been formally related to the notion of dominant strategy, typical of strategic games.

Cooperation and competition It is well-known from the literature of cooperative game theory how strategic games can be described as cooperative games, by identifying the cooperative possibilities of the players involved in an interaction, e.g. what the players can do together. It is also well-known, for a

restricted class of games, how a certain class of cooperative games can be related to strategic games. Chapter 3 has generalized this result, correcting a believed correspondence from [54]. Besides allowing a full description of strategic games in terms of cooperative games, the results obtained have been used to study the specificity of coalitional rationality in strategic games, the object of study of the first research question.

Coalitions and interdependence The standard models of coalitional power assign to coalitions the capacity of fully coordinating their members. In those models no vestige is found on the reasons for players to work together. Chapter 5 has studied a model of coalitions that result from reciprocal favours among their members. A new class of cooperative games have been defined, the so-called *dependence games*, that lie between the individualistic approach of non-cooperative games and the fully cooperative approach of cooperative games.

Rationality and logic While Chapter 3 and Chapter 5 have dealt with a structural analysis of notions like coalitional strategy and agreement, Chapter 4 and Chapter 6 have analyzed those structures in terms of logical languages. The added value of such formulation lies in the simplicity of these languages, that capture the essential features of notions like undomination, preference lifting, agreements, in a modal language consisting of relatively few operators. Once the bridge is laid between formal languages and game-theoretical structures, results concerning the first (for instance decidability, final model property, succinctness of representation, reducibility etc.) can immediately be transferred to the second, shedding new light on the formal properties of game-theoretical structures.

Rationality and norms One of the most fascinating problems of modelling coalitional action is that of its regulation. Chapter 3 and Chapter 5 have come up with models of coalitional rationality, both for the classical account of coalitional power and for the one taking players' interdependence into account. In interaction however certain properties may be desirable and a language can be constructed to express this desirability. Chapter 4 and Chapter 6 have presented deontic languages to express what coalitions should do. First, interpreting norms as obligations (or forbiddance and permission) to act rationally (or forbiddance not to act irrationally); second evaluating coalitional rationality against a given labelling of outcomes, to be understood as undesirable ones.

All in all, our work has explored several aspects of coalitional rationality, where players choose together according to their mutual interests.

Appendix A

Representation Theorem

A.1 The Original Proof

The proof of correspondence in [54] is articulated in two main parts, corresponding to both directions:

- The easy part of the proof in [54] is checking that the α-effectivity function of strategic games is playable. This fact has already been noticed in the preliminaries.
- The difficult part of the proof is constructive: the idea is that from a playable effectivity function we can obtain a strategic game with the same α-effectivity function.

The focus here is on the latter part. The argument given in [54] discusses a procedure to construct from each playable effectivity function E a strategic game $\mathbb{G} = (N, S, \Sigma_i, o)$ such that $E^{\alpha}_{\mathbb{G}} = E$. The argument can be outlined in a few steps.

Step 1: the players and the domain remain the same The game (form) $\mathbb{G} = (N, S, \Sigma_i, o)$ inherits the set of outcomes and the set of players from the coalitional game (form) $\mathbb{G} = (N, S, E)$.

Step 2: coalitions choose a set from their effectivity function The second step concerns the construction of the sets of strategies for each player. To do this a family of functions is defined:

$$F_i = \{f_i : C_i \to 2^W \mid \text{ for all } C \text{ we have that } \ f_i(C) \in E(C)\}$$

where $C_i = \{C \subseteq N \mid i \in C\}$. Each function f_i assigns to the coalitions of which i is a member an arbitrary choice in the coalitional effectivity function. F_i simply collects all such functions.

The idea is to represents the strategies of a coalition C as the set $\bigcup\{F_i | i \in C\}$. Notice already that $F_\emptyset = \{\emptyset\}$, i.e. the empty coalition has only the empty strategy.

Step 3: coalitions are selected according to their choices The third step concerns the construction of coalitional choices, using the family of functions defined in Step 2.

Let $f = (f_i)_{i \in N}$, with $f_i \in F_i$, be a tuple of such assignments, one per player. We can now define the set $P_\infty(f)$ which results from iterative partitioning the coalitions in the coarsest possible way such that players in the same part are assigned same coalitional choices.

$$\begin{aligned} P_0(f) &= \langle N \rangle \\ P_1(f) &= P(f, N) = \langle C_1^1, \ldots, C_{k_1}^1 \rangle \\ P_2(f) &= \langle P(f, C_1^1), \ldots, P(f, C_{k_1}^1) \rangle = \langle C_2^2, \ldots, C_{k_2}^2 \rangle \\ &\ldots \\ P_\infty(f) &= P_r(f) \text{ such that } P_i(f) = P_{i+1}(f) \text{ for all } i \geq r, \end{aligned}$$

where each $P(f, C)$ returns the coarsest partitioning $\langle C_1, \ldots, C_m \rangle$ of coalition C such that for all $l \leq m$ and for all $i, j \in C_l$ it holds that $f_i(C) = f_j(C)$.

Notice that what happens is that a subset of C is part of the partition $P(f, C)$ if its members agree modulo f.

Step 4: an outcome is chosen in the intersection of coalitional choices Using the partition process it is possible to define the strategies and the outcome function in the game. Each player in N is given a set of strategies of the form (f_i, t_i, h_i) where $f_i \in F_i$, t_i is a player (possibly different from i), and $h_i : 2^W \setminus \emptyset \to W$ is a selector function that picks an arbitrary element from each nonempty subset of W.

Given the process of partition $P_\infty(f)$, the outcome function of game $\mathbb{G}$ is defined as follows:

$$o(\sigma_N) = h_{i_0}\Big(\bigcap_{l=1}^{k} f(C_l)\Big),$$

where i_0 is a uniquely chosen player, h_{i_0} is the selector function for player i_0, and C_l are partitions taken from $P_\infty(f)$.

That concludes the construction of game $\mathbb{G}$ which α-corresponds to the effectivity function E. The remaining two steps are supposed to prove that indeed $E = E^\alpha_{\mathbb{G}}$.

The proof: choices are preserved in the game First, an attempt to prove $E(C) \subseteq E^\alpha_{\mathbb{G}}(C)$ for arbitrary coalition C is presented, i.e. the proof that all choices in the original effectivity function are also choices in the derived game:

> For the inclusion from left to right, assume that $X \in E(C)$. Choose any C-strategy $\sigma_C = (f_i, t_i, h_i)_{i \in C}$ such that for all $i \in C$ and for all $C' \supseteq C$ we have $f_i(C') = X$.(*) By coalition monotonicity, such f_i exists.(**) Take now any $\overline{C}$-strategy, $\sigma_{\overline{C}} = (f_i, t_i, h_i)_{i \in \overline{C}}$. We need to show that $o(\sigma_C, \sigma_{\overline{C}}) \in X$. To

see this, note that C must be a subset of one of the partitions C_l in $P_\infty(f)$. Hence, $o(\sigma_N) = h_{i_0}(G(f)) = h_{i_0} \bigcap_{l=1}^{k} f(C_l) \in X$. [54, p.29]

The deduction of the last sentence is where the proof goes wrong. The problem is, for $C = \emptyset$, the only available strategy $\sigma_\emptyset$ is the empty strategy which vacuously satisfies condition (*).[1] And, for any player i, a choice assignment f_i satisfying the condition must exist. However, *there is no guarantee that any i will indeed choose f_i in its strategy* since the coalition C for which we can fix its strategy does not include any players. In consequence, we have no right to deduce that $h_{i_0}\Big(\bigcap_{l=1}^{k} f(C_l)\Big) \in X$: this could be only concluded if the intersection contains at least one player whose choice $f_i(C_l)$ is X (or a subset of X).

To see this more clearly, let us go back to the counterexample of Section 3.3.1. Note that $\sigma_{\overline{C}} = \sigma_{\{a\}} = (f_a, a, h_a)$ such that $f_a(\{a\}) \in E(\{a\})$. Let us now take $X = \mathbb{N} \setminus \{1\}$, $f_a(\{a\}) = \mathbb{N}$, and $h_a(\mathbb{N}) = 1$. Now, $o(\sigma_N) = o(\sigma_{\{a\}}) = 1 \notin X$, which invalidates the argument from [54] quoted above.

The proof: choices are not added in the game The proof of the other direction ($E^\alpha_{\mathbb{G}}(C) \subseteq E(C)$ i.e. that all choices in the derived game are also choices in the original effectivity function) fails too, because in order to establish the inclusion for $C = N$, it is reduced to inclusion (v) for $C = \emptyset$, and we have just shown that it does not necessarily hold.

This concludes the analysis of the proof of Representation Theorem from [54]. The construction of the strategic game corresponding to a given effectivity function fails because the game might endow the empty coalition and the grand coalition of players with inappropriate powers.

A.2 The New Proof

Proof *Given a strategic game $\mathbb{G}$ it is easy to see that its α-effectivity function $E^{\mathbb{G}}$ is truly playable (by Propositions 4 and 17).*

For the other direction, given a truly playable effectivity function E, we slightly change Pauly's procedure that has been previously outlined (steps 1–4). That is, we impose an additional constraint on players' strategies $\sigma_i = (f_i, t_i, h_i)$, namely, we require that $h_i(X) = x$ for some $\{x\} \in E(N)$. In other words, the selector functions only select the "jewels" in the crown.

Note that for $C \notin \{\emptyset, N\}$ the new procedure yields game $\mathbb{G}'$ with exactly the same $E^{\mathbb{G}}(C)$ as the original construction $\mathbb{G}$ from [54] because:

- *We do not add any new choice sets to $E^{\mathbb{G}}(C)$. Suppose that we do, then it can only happen because the selectors chosen by players outside C are restricted to $\{x \mid \{x\} \in E(N)\}$, and hence we can have that $X \cap \{x \mid \{x\} \in E(N)\} \in E^{\mathbb{G}'}(C)$ in the new construction for some $X \in E^{\mathbb{G}}(C)$ from the previous construction. However, by true*

[1] Notice the universal quantification over the members of the empty coalition.

playability of E and Proposition 17 we have that $\{x \mid \{x\} \in E(N)\} \in E(\emptyset)$*, and thus by superadditivity all the states* $y \notin \{x \mid \{x\} \in E(N)\}$ *can be removed from C's strategies that yielded X in* $\mathbb{G}$*. But then these states will also be removed from the intersection* $\bigcap_{l=1}^{k} f(C_l)$*, and so* $X \cap \{x \mid \{x\} \in E(N)\} \in E^{\mathbb{G}}(C)$ *already in the previous construction.*

- *We do not remove any choice sets from* $E^{\mathbb{G}}(C)$*. Suppose that we do, then it can be only because of removing an* $X \in E^{\mathbb{G}}(C)$ *which contains "superfluous" elements and replacing it with* $X \cap \{x \mid \{x\} \in E(N)\}$*. But then, X must also be in* $E^{\mathbb{G}'}(C)$ *because* $E^{\mathbb{G}'}(C)$ *is closed under supersets.*

It remains now to show that the procedure constructs a strategic game $\mathbb{G}$ *such that* $E(C) = E^{\mathbb{G}}(C)$ *for all* $C \subseteq N$*, that is, to show that both directions work well in case of truly playable structures.*

The proof of $E(C) \subseteq E^{\mathbb{G}}(C)$ *We show that* $E(C) \subseteq E^{\mathbb{G}}(C)$ *for* $C = \emptyset$ *and* $C = N$*, the only cases in which the original proof failed for playable structures.*

Assume that $X \in E(\emptyset)$*. We need to prove that* $X \in E^{\mathbb{G}}(\emptyset)$*. By true playability and Proposition 17 we know that there exists* $Y \in E(\emptyset)$ *such that* $Y \subseteq X$*,* $E^{nc}(\emptyset) = \{\{x \mid \{x\} \in E(N)\}\}$*. Now, consider any strategy profile* σ_N*. We have* $o(\sigma_N) = h_{i_0}\Big(\bigcap_{l=1}^{k} f(C_l)\Big) \in Y$ *because every* h_i *returns only elements in Y by construction.*

For the case $C = N$*, assume that* $X \in E(N)$*. We need to prove that* $X \in E^{\mathbb{G}}(N)$*. By true playability we have that there exists* $x \in X$ *such that* $\{x\} \in E(N)$*. Now, let,* σ_N *consist of strategies* $\sigma_i = (f_i, t_i, h_i)$ *such that* $f_i(N) = x$ *for every i. It is easy to see that* $o(\sigma_N) = x$*, and hence* $\{x\} \in E^{\mathbb{G}}(N)$*. Thus,* $X \in E^{\mathbb{G}}(N)$ *because* $E^{\mathbb{G}}(N)$ *is closed under supersets.*

The proof of $E^{\mathbb{G}}(C) \subseteq E(C)$ *We show that* $E^{\mathbb{G}}(C) \subseteq E(C)$*, that is, we will assume that* $X \notin E(C)$*, and show that* $X \notin E^{\mathbb{G}}(N)$*. We do it by a slight modification of the original proof from [54].*

Suppose first that $C = N$*. Then,* $\overline{X} \in E(\emptyset)$ *by N-maximality, and by (v) we have* $\overline{X} \in E^{\mathbb{G}}(\emptyset)$*. Since* $E^{\mathbb{G}}$ *is truly playable, we have also that* $X \notin E^{\mathbb{G}}(N)$*.*

Assume now that $C \neq N$*, and let* $j_0 \in \overline{C}$*. Let* σ_C *be any strategy for coalition C. We must show that there is a strategy* $\sigma_{\overline{C}}$ *such that* $o(\sigma_C, \sigma_{\overline{C}}) \notin X$*. To show this, we take* $\sigma_{\overline{C}} = (f_i, t_i, h_i)_{i \in \overline{C}}$ *such that for all* $C' \supseteq \overline{C}$ *and for all* $i \in \overline{C}$ *we have* $f_i(C') = W$*. We also choose* t_{j_0} *such that* $((t_1 + \ldots + t_N) \bmod n) + 1 = j_0$*. Note that* $\overline{C}$ *must be an element of one of the partitions* C_l *in* $P_\infty(f)$*, say* C_{l_0}*. Moreover, there must be a partitioning* $\langle C_1, \ldots, C_k \rangle$ *of* $N \setminus C_{l_0}$ *such that* $G(f) = f(C_{l_0}) \cap \bigcap_{l=1}^{k} f(C_l) = \bigcap_{l=1}^{k} f(C_l)$*. Since* $f(C_l) \in E(C_l)$ *we get that* $G(f) \in N \setminus C_{l_0}$ *by superadditivity. By coalition-monotonicity and the fact that* $N \setminus C_{l_0} \subseteq C$*, we also have* $G(f) \in E(C)$*. Finally, by (*) and superadditivity we obtain* $G(f) \cap \{x \mid \{x\} \in E(N)\} \in E(C)$*.*

Since $X \notin E(C)$ *and* $E(C)$ *is closed under supersets, it must hold that* $G(f) \cap \{x \mid \{x\} \in E(N)\} \not\subseteq X$*. Thus, there is some* $s_0 \in W$ *such that:* $s_0 \in G(f)$*,* $\{s_0\} \in E(N)$*, and* $s_0 \notin X$*. Now we fix* h_{j_0} *so that* $h_{j_0}(G(f)) = s_0$*. Then,* $o(\sigma_C, \sigma_{\overline{C}}) = h_{j_0}(G(f)) = s_0 \notin X$ *which concludes the proof.*

Appendix B

Selected Proofs

B.1 The subgame operator: validities

$$[C \downarrow \xi]p \leftrightarrow ([C]\xi \rightarrow p)$$

Proof *Take an arbitrary tuple* M, w. $M, w \models [C \downarrow \xi]p \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \downarrow_{(C,\xi^M)} \models p \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \models p \Leftrightarrow M, w \models [C]\xi \rightarrow p$.

$$[C \downarrow \xi]\neg\varphi \leftrightarrow ([C]\xi \rightarrow \neg[C \downarrow \xi]\varphi)$$

Proof *Take an arbitrary tuple* M, w. $M, w \models [C \downarrow \xi]\neg\varphi \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \downarrow_{(C,\xi^M)} \models \neg\varphi \Leftrightarrow M, w \models [C]\xi$ *implies that* $(M, w \models [C]\xi$ *and* $M, w \downarrow_{(C,\xi^M,w)} \models \neg\varphi) \Leftrightarrow M, w \models [C]\xi$ *implies that not*$(M, w \models [C]\xi$ *implies* $M, w \downarrow_{(C,\xi^M)} \not\models \neg\varphi) \Leftrightarrow \models [C]\xi$ *implies that not*$(M, w \models [C]\xi$ *implies* $M, w \downarrow_{(C,\xi^M)} \models \varphi) \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \not\models [C \downarrow \xi]\varphi \Leftrightarrow M, w \models [C]\xi \rightarrow \neg[C \downarrow \xi]\varphi$

$$[C \downarrow \xi](\varphi \wedge \psi) \leftrightarrow ([C \downarrow \xi]\varphi \wedge [C \downarrow \xi]\psi)$$

Proof *Take an arbitrary tuple* M, w. $M, w \models [C \downarrow \xi](\varphi \wedge \psi) \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \downarrow_{(C,\xi^M)} \models \varphi \wedge \psi \Leftrightarrow M, w \models [C]\xi$ *implies that* $(M, w \downarrow_{(C,\xi^M)} \models \varphi$ *and* $M, w \downarrow_{(C,\xi^M)} \models \psi) \Leftrightarrow (M, w \models [C]\xi$ *implies that* $M, w \downarrow_{(C,\xi^M)} \models \varphi)$ *and* $(M, w \models [C]\xi$ *implies that* $M, w \downarrow_{(C,\xi^M)} \models \psi) \Leftrightarrow (M, w \models [C \downarrow \xi]\varphi)$ *and* $(M, w \models [C \downarrow \xi]\psi) \Leftrightarrow M, w \models ([C \downarrow \xi]\varphi \wedge [C \downarrow \xi]\psi)$

$$[C \downarrow \xi]A\varphi \leftrightarrow ([C]\xi \rightarrow A\varphi)$$

Proof *Take an arbitrary tuple*M, w. $M, w \models [C \downarrow \xi]A\varphi \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \downarrow_{(C,\xi^M)} \models A\varphi \Leftrightarrow M, w \models [C]\xi$ *implies that* $M, w \models A\varphi \Leftrightarrow M, w \models [C]\xi \rightarrow A\varphi$

$$[C \downarrow \xi][C']\varphi \leftrightarrow ([C]\xi \rightarrow [C'](\xi \rightarrow \varphi))(\text{ for } C' \cap C = \emptyset \text{ and } C' \neq \emptyset)$$

Proof $\Leftarrow$: *Suppose, for some* $C' \neq \emptyset$, *that* $[C]\xi \to [C'](\xi \to \varphi)$ *and* $M, w \not\models [C \downarrow \xi][C']\varphi$ *for some* C *such that* $(C \cap C') = \emptyset$. *The semantic clauses then tell us that (if* $\xi^M \in E(w)(C)$ *then* $(\xi \to \varphi)^M \in E(w)(C')$*) and* $\xi^M \in E(w)(C)$ *and* $\varphi^M \notin E'(w)(C')$. *[We write* E' *for* $E \downarrow_{(C,\xi^M)}$*.] By modus ponens* $(\xi \to \varphi)^M \in E(w)(C')$.

By the definition of update, $E'(w)(C') = (E(w)(C') \sqcap \xi^M)^{sup}$. *So,* $((\xi \to \varphi)^M \cap \xi^M) \in E'(w)(C')$. *By elementary set theory this just says that* $\varphi^M \in E'(w)(C')$. *Contradiction.*

$\Rightarrow$: *Suppose, for some* $C' \neq \emptyset$, *that* $M, w \models [C \downarrow \xi][C']\varphi$ *and* $M, w \not\models [C]\xi \to [C'](\xi \to \varphi)$ *for some* C *such that* $(C \cap C') = \emptyset$. *The semantic clauses then tell us that (if* $\xi^M \in E(w)(C)$ *then* $\varphi^M \in E'(w)(C')$*) and* $\xi^M \in E(w)(C)$ *and* $(\xi \to \varphi)^M \notin E(w)(C')$. *By modus ponens we are assuming that* $\varphi^M \in E'(w)(C')$ *and* $(\xi \to \varphi)^M \notin E(w)(C')$.

By the definition of update, $E'(w)(C') = (E(w)(C') \sqcap \xi^M)^{sup}$. *Because* $\varphi^M \in E'(w)(C')$, *there must be some* $X \in E(w)(C')$, *such that* $(X \cap \xi^M) \subseteq \varphi^M$. *By elementary set theory, it must be the case that* $X \subseteq (\xi \to \varphi)^M$.

Hence, by outcome monotonicity of E, *if* $X \in E(w)(C')$, *then* $(\xi \to \varphi)^M \in E(w)(C')$. *Contradiction.*

$$[C \downarrow \xi]([C']\varphi \leftrightarrow A(\xi \to \varphi))(\text{ for } C' \cap C \neq \emptyset)$$

Proof *Take arbitrary tuple* M, w, *and* $\xi^M \in E(w)(C)$. *Consider a coalition* C' *with* $C' \cap C \neq \emptyset$. *We have that* $E(w \downarrow_{(C,\xi^M)})(C') = (\xi^M)^{sup}$ *by semantics. This means that* $\xi^M \subseteq \varphi^M$ *if and only if* $\varphi^M \in E(w \downarrow_{(C,\xi^M)})(C')$. *In conclusion* $M, w \models [C \downarrow \xi]([C']\varphi \leftrightarrow A(\xi \to \varphi))$. *Notice that this also means* $M, w \models [C \downarrow \xi][C']\varphi \leftrightarrow A(\xi \to \varphi)$.

$$[C \downarrow \xi][C']\varphi \leftrightarrow ([C]\xi \to [C']\varphi)(\text{ for } C' = \emptyset)$$

Proof *It follows directly from the semantics of the update operator for the case of* $D = \emptyset$.

B.2 The switch operator: validities

$$[sw]p \leftrightarrow p$$

Proof *Take arbitrary* M, w. $M, w \models [sw]p \Leftrightarrow M, (sw, w) \models p \Leftrightarrow M, w \models p$.

$$[sw]\neg\varphi \leftrightarrow \neg[sw]\varphi$$

Proof *Take arbitrary* M, w. $M, w \models [sw]\neg\varphi \Leftrightarrow M, (sw, w) \models \neg\varphi \Leftrightarrow M, (sw, w) \not\models \varphi \Leftrightarrow M, w \not\models [sw]\varphi \Leftrightarrow M, w \models \neg[sw]\varphi$

$$[sw](\varphi \wedge \psi) \leftrightarrow ([sw]\varphi \wedge [sw]\psi)$$

Proof *Take arbitrary* M, w. $M, w \models [sw](\varphi \wedge \psi) \Leftrightarrow M, (sw, w) \models \varphi \wedge \psi \Leftrightarrow M, (sw, w) \models \varphi$ *and* $M, (sw, w) \models \psi \Leftrightarrow M, w \models [sw]\varphi$ *and* $M, w \models [sw]\psi \Leftrightarrow M, w \models [sw]\varphi \wedge [sw]\psi$

$$[sw]A\varphi \leftrightarrow A\varphi$$

Proof *Take arbitrary* M, w. $M, w \models [sw]A\varphi \Leftrightarrow M, (sw, w) \models A\varphi \Leftrightarrow \varphi^M = W \Leftrightarrow \varphi^M = W \Leftrightarrow M, w \models A\varphi$.

$$[sw]\Box_i^{\preceq}\varphi \leftrightarrow \Box_i^{\preceq}\varphi$$

Proof *Take arbitrary* M, w. $M, w \models [sw]\Box_i^{\preceq}\varphi \Leftrightarrow M, (sw, w) \models \Box_i^{\preceq}\varphi \Leftrightarrow M, sw(v) \models \varphi$ *for every* v *such that* $w \preceq_i v \Leftrightarrow M, v \models \varphi$ *for every* v *such that* $w \preceq_i v \Leftrightarrow M, w \models \Box_i^{\preceq}\varphi$.

$$[sw][k]\varphi \leftrightarrow [sw^{-1}(k)]\varphi$$

Proof *Take arbitrary* M, w. $M, w \models [sw][k]\varphi \Leftrightarrow M, (sw, w) \models [k]\varphi \Leftrightarrow \varphi^M \in E(sw(w))(k) \Leftrightarrow \varphi^M \in E(w)(j)$, *for* $sw(k) = j \Leftrightarrow M, w \models [sw^{-1}k]\varphi$.

$$[sw][k \downarrow \psi]\varphi \leftrightarrow [sw^{-1}(k) \downarrow \psi]\varphi$$

Proof *Take arbitrary* M, w. $M, w \models [sw][k \downarrow \psi]\varphi \Leftrightarrow M, (sw, w) \models [k \downarrow \psi]\varphi \Leftrightarrow M, (sw, w) \models [k]\psi$ *implies* $M \downarrow_{k,\psi,w}, (sw, w) \models \varphi \Leftrightarrow M \models [sw^{-1}(k)]\psi$ *implies* $M \downarrow_{sw^{-1}(k)(w),\psi^M} \models \varphi \Leftrightarrow M \models [sw^{-1}(k) \downarrow \psi]\varphi$.

B.3 Completeness for TPCL

We will prove completeness of Truly Playable Coalition Logic, using canonical model followed by filtration for monotone logics, partly using constructions from [23] and [54]. Thus, we will also obtain finite model property for TPCL. Here we only sketch the standard canonical model construction and refer the reader for further details to [23] and [54].

To shorten the notation we hereafter denote the logic TPCL by $\mathcal{L}$.

Given a TPCL $\mathcal{L}$ we write $\vdash_{\mathcal{L}} \varphi$ for $\varphi \in \mathcal{L}$ and $\Sigma \vdash_{\mathcal{L}} \varphi$ if there exists $\sigma_1, \sigma_2, ..., \sigma_n \in \Sigma$ with $\sigma_1 \wedge \sigma_2 \wedge ... \wedge \sigma_n \rightarrow \varphi \in \mathcal{L}$. As usual, omitting $\mathcal{L}$ is equivalent to considering the smallest TPCL. A set of formulas Σ is $\mathcal{L}$-inconsistent if $\Sigma_{\mathcal{L}} \vdash \bot$. A TPCL $\mathcal{L}$ is *sound* with respect to a class of TPCL models $\mathcal{K}$ if $\Sigma \vdash_{\mathcal{L}} \varphi$ implies $\Sigma \models_{\mathcal{K}} \varphi$, and *complete* if the converse holds. It is *weakly complete* if it is complete for $\Sigma = \emptyset$. As a consequence, if a language $\mathcal{L}$ is weakly complete with respect to a class of models $\mathcal{K}$ then every $\mathcal{L}$-consistent formula can be satisfied in a model $M \in \mathcal{K}$. Soundness and weak completeness are equivalent to the fact that $\mathcal{L} = \mathcal{L}_{\mathcal{K}}$, with respect to a class of models $\mathcal{K}$. If moreover $\mathcal{K}$ is a class of models with finite domain (or a class of finite models) then $\mathcal{L}$ is said to have the *finite model property*.

Using a well-known argument [23], every $\mathcal{L}$-consistent set of formulas Σ can be extended to a maximally consistent set $\Sigma^*_{\mathcal{L}}$ such that $\Sigma \subseteq \Sigma^*_{\mathcal{L}}$ and for every formula

$\varphi \in \mathcal{L}$ we have that either $\varphi \in \Sigma^*_{\mathcal{L}}$ or $\neg\varphi \in \Sigma^*_{\mathcal{L}}$; $\varphi \vee \psi \in \Sigma^*_{\mathcal{L}}$ if and only if $\varphi \in \Sigma^*_{\mathcal{L}}$ or $\psi \in \Sigma^*_{\mathcal{L}}$; if $\Sigma^*_{\mathcal{L}} \vdash_{\mathcal{L}} \varphi$ then $\varphi \in \Sigma^*_{\mathcal{L}}$.

We take now the set $W^{\mathcal{L}}$ of maximally consistent sets and we define $\varphi^* = \{s \in W^{\mathcal{L}} \mid \varphi \in s\}$ to be the *proof set* of φ.

Definition 78 (Canonical Model) *The canonical model for* TPCL *is* $M^{\mathcal{L}} = (W^{\mathcal{L}}, E^{\mathcal{L}}, R^{\mathcal{L}}, V^{\mathcal{L}})$ *where:*

- $w \in V^{\mathcal{L}}(p)$ *if and only if* $p \in w$;
- $X \in E^{\mathcal{L}}(w)(C)$ *if and only if* $\exists\psi^* \subseteq X : [C]\psi \in w$, *for* $C \neq N$
- $X \in E^{\mathcal{L}}(w)(C)$ *if and only if* $\forall\psi^*$ *if* $X \subseteq \psi^*$ *then* $[C]\psi \in w$, *for* $C = N$
- $wR^{\mathcal{L}}v$ *if and only if* $\forall\psi$, *if* $\psi \in v$ *then* $\langle O\rangle\psi \in w$.

Some remarks:

- That $E^{\mathcal{L}}$ is playable and well-defined is proved in [54].
- The canonical relation for N is defined in [54] in the following slightly different, but de facto equivalent, way: $X \in E^{\mathcal{L}}(w)(N)$ if and only if $[\emptyset]\psi \notin w$ for all ψ^* such that $\psi^* \subseteq \overline{X}$. The equivalence follows easily from the fact that $\vdash_{\mathcal{L}} [N]\varphi \leftrightarrow \neg[\emptyset]\neg\varphi$.
- The canonical relation for $\langle O\rangle$ is defined as a canonical relation for normal modal logics [23].

Proposition 59 (Truth Lemma) *For any* $w \in W^{\mathcal{L}}$ *we have that* $M^{\mathcal{L}}, w \models \varphi$ *if and only if* $\varphi \in w$.

Proof *By induction on the length of* φ*: standard for atomic propositions, boolean formulas, and formulas of the form* $\langle O\rangle\psi$*; proved in [54] for formulas of the form* $[C]\psi$.

The canonical model is an extended coalition model, however it is not standard, neither truly playable. The reason for that is the fact that for all $\psi \in \mathcal{L}, \psi \in v$ implies that $[N]\psi \in w$ is not sufficient to conclude that $\{v\} \in E^{\mathcal{L}}(w)(N)$ as states are not characterized by unique formulas of the language of $\mathcal{L}$. In order to obtain a standard and truly playable model satisfying the given $\mathcal{L}$-consistent formula δ we are going to filter the canonical model with the finite set $\Sigma(\delta)$ obtained by taking all subformulae of δ and closing under boolean operators, up to propositional equivalence.

Filtrations First, we define a general notion of filtration for extended coalition models and then a special filtration construction that preserves playability. Filtrations of coalition models are introduced in [36] for the purpose of axiomatizing Nash-consistent Coalition Logic. What we do here is to add the filtration for the relation corresponding to the modality $\langle O\rangle$.

Let $M = (W, E, R, V)$ be an extended coalition model and Σ a subformula-closed set of formulas over $\mathcal{L}$. The equivalence classes induced by Σ on M are defined as follows:

$$v \equiv_\Sigma w \Leftrightarrow \text{ for all } \varphi \in \Sigma : M, v \models \varphi \text{ if and only if } M, w \models \varphi.$$

We denote the equivalence class to which v belongs by $|v|$ and the set $\{|v| \mid v \in X\}$ by $|X|$ for any $v \in W$ and $X \subseteq W$.

Definition 79 (Filtration) *Let $M = (W, E, R, V)$ be an extended coalition model and Σ a subformula closed set of formulas over $\mathcal{L}$. A coalition model $M^f_\Sigma = (W^f_\Sigma, E^f_\Sigma, R^f_\Sigma, V^f_\Sigma)$ is a filtration of M through Σ whenever the following conditions are satisfied:*

- $W^f_\Sigma = |W|$.
- *For all $C \subseteq N$ and $\varphi \in \Sigma$, $\varphi^M \in E(w)(C)$ implies $\{|v| \mid M, v \models \varphi\} \in E^f_\Sigma(|w|)(C)$.*
- *For all $C \subseteq N$ and $Y \subseteq |W|$: $Y \in E^f_\Sigma(|w|)(C)$ implies that for all $\varphi \in \Sigma$ if $\varphi^M \subseteq \{v \mid |v| \in Y\}$ then $\varphi^M \in E(w)(C)$.*
- *If wRv then $|w|R|v|$.*
- *If $|w|R|v|$ then for all $\langle O\rangle\varphi \in \Sigma$, if $M, v \models \varphi$ then $M, w \models \langle O\rangle\varphi$.*
- *$V^f_\Sigma(p) = |V(p)|$ for all atoms $p \in \Sigma$.*

The conditions above are needed to ensure the Filtration Lemma, as showed in [36] for the neighbourhood functions and e.g. in [23] for the binary relation.

Proposition 60 (Filtration Lemma) *If $M^f_\Sigma = (W^f_\Sigma, E^f_\Sigma, R^f_\Sigma, V^f_\Sigma)$ is a filtration of M through Σ then for all $\varphi \in \Sigma$ we have that $M, w \models \varphi$ if and only if $M^f_\Sigma, |w| \models \varphi$.*

Definition 80 (Playable Filtration) *Let $M = (W, E, R, V)$ be an extended coalition model and $\Sigma(\delta)$ the boolean closure of the set of subformulas of δ, such that $\delta \in \mathcal{L}$, the language of TPCL. A coalition model $M^F_{\Sigma(\delta)} = (W^F_{\Sigma(\delta)}, E^F_{\Sigma(\delta)}, R^F_{\Sigma(\delta)}, V^F_{\Sigma(\delta)})$ is a* playable filtration *of M through $\Sigma(\delta)$ whenever the following conditions are satisfied:*

- $W^F_{\Sigma(\delta)} = |W|$.
- *For all $C \subset N, C \neq N$, and $Y \subseteq |W|$: $Y \in E^F_{\Sigma(\delta)}(|w|)(C)$ if and only if there exists $\varphi \in \Sigma(\delta)$ such that $\varphi^M \subseteq \{v \mid |v| \in Y\}$ and $\varphi^M \in E(w)(C)$.*
- *For all $Y \subseteq |W|$: $Y \in E^F_{\Sigma(\delta)}(|w|)(N)$ if and only if $\overline{Y} \notin E^F_{\Sigma(\delta)}(|w|)(\emptyset)$.*
- *$|w|R^F_{\Sigma(\delta)}|v|$ if and only if there exists $w' \in |w|$, $\exists v' \in |v|$ such that $w'Rv'$.*
- *$V^F_{\Sigma(\delta)}(p) = |V(p)|$ for all atoms $p \in \Sigma(\delta)$.*

That $M^F_{\Sigma(\delta)}$ is a filtration in the sense of Definition 79 is proved in [36] for the coalitional modalities. We have added to that a minimal filtration for modality $\langle O \rangle$. So $M^F_{\Sigma(\delta)}$ is a filtration in the sense of Definition 79. In [36] it is also shown that playability is preserved by that filtration and that every subset of $W^F_{\Sigma(\delta)}$ is definable by a formula of $\Sigma(\delta)$ as follows. First, for every state $|w| \in |W|$ we define

$$\chi^F_{\Sigma(\delta)}(|w|) := \bigwedge \{\varphi \in \Sigma(\delta) \mid M^F_{\Sigma(\delta)}, |w| \models \varphi\}.$$

Then, for every $Y \subseteq |W|$ we put

$$\chi^F_{\Sigma(\delta)}(Y) := \bigvee \{\chi^F_{\Sigma(\delta)}(|w|) \mid |w| \in Y\}.$$

It is straightforward to show, using the filtration lemma, that for every $Y \subseteq |W|$:

$$M^F_{\Sigma(\delta)}, |w| \models \chi^F_{\Sigma(\delta)}(Y) \text{ if and only if } |w| \in Y,$$

that is, $\chi^F_{\Sigma(\delta)}(Y)$ indeed characterizes the set y in $M^F_{\Sigma(\delta)}$.

Proposition 61 *$M^F_{\Sigma(\delta)}$ is standard and truly playable.*

Proof *To prove that $M^{\mathcal{L},F}_{\Sigma(\delta)}$ is standard we have to show that for each $w, v \in W$, $|v| R^{\mathcal{L},F}_{\Sigma(\delta)} |w|$ if and only if $\{|v|\} \in E^{\mathcal{L},F}_{\Sigma(\delta)}(|w|)(N)$. From right to left is straightforward. For the other direction, suppose $|v| R^{\mathcal{L},F}_{\Sigma(\delta)} |w|$. Then $M^{\mathcal{L},F}_{\Sigma(\delta)}, |v| \models \langle O \rangle \chi^F_{\Sigma(\delta)}(|w|)$ by definition of $R^{\mathcal{L},F}_{\Sigma(\delta)}$ and by the properties of filtrations. By the fact that $R^{\mathcal{L},F}_{\Sigma(\delta)}$ is a minimal filtration we have that $\exists w' \in |w|, \exists v' \in |v|$ such that $v' R^{\mathcal{L}} w'$. By definition of $R^{\mathcal{L}}$ and the Truth Lemma we have that $M^{\mathcal{L}}, v' \models \langle O \rangle \chi^F_{\Sigma(\delta)}(|w|)$. By the axioms of $\mathcal{L}$ and the Truth Lemma we have $M^{\mathcal{L}}, v' \models [N] \chi^F_{\Sigma(\delta)}(|w|)$, hence $M^{\mathcal{L}}, v' \models \neg[\emptyset] \neg \chi^F_{\Sigma(\delta)}(|w|)$. Then $(\neg \chi^F_{\Sigma(\delta)}(|w|))^{M^{\mathcal{L}}} \notin E^{\mathcal{L}}(v')(\emptyset)$ by the definition of $E^{\mathcal{L}}$. But, by Definition 79 $\{(\neg \chi^F_{\Sigma(\delta)}(|w|))^{M^{\mathcal{L},F}_{\Sigma(\delta)}}\} \notin E^{\mathcal{L},F}_{\Sigma(\delta)}(|v|)(\emptyset)$ and in turn $\{(\chi^F_{\Sigma(\delta)}(|w|))^{M^{\mathcal{L},F}_{\Sigma(\delta)}}\} \in E^{\mathcal{L},F}_{\Sigma(\delta)}(|v|)(N)$. Recall now that $(\chi^F_{\Sigma(\delta)}(|w|))^{M^{\mathcal{L},F}_{\Sigma(\delta)}} = |w|$.*

Now, to prove that $M^{\mathcal{L},F}_{\Sigma(\delta)}$ is truly playable, assume $Y \in E^{\mathcal{L},F}_{\Sigma(\delta)}(|w|)(N)$. Then, $(\neg \chi^F_{\Sigma(\delta)}(Y))^{M^{\mathcal{L},F}_{\Sigma(\delta)}} \notin E^{\mathcal{L}}(w)(\emptyset)$ by the definition of filtration, which means that for all $\varphi \in \Sigma(\delta)$, if $\{v \mid |v| \in (\neg \chi^F_{\Sigma(\delta)}(Y))^{M^{\mathcal{L},F}_{\Sigma(\delta)}}\} \subseteq \varphi^M$ then $\varphi^M \notin E^{\mathcal{L}}(w)(\emptyset)$. In particular $(\neg \chi^F_{\Sigma(\delta)}(Y))^{M^{\mathcal{L}}} \notin E^{\mathcal{L}}(w)(\emptyset)$. By the definition of $E^{\mathcal{L}}$ we have that $[\emptyset] \neg \chi^F_{\Sigma(\delta)}(Y) \notin w$ and by true playability that $\langle O \rangle \chi^F_{\Sigma(\delta)}(Y) \in w$. By the definition of canonical relation for $\langle O \rangle$ we have that there exists v with $w R^{\mathcal{L}} v$ such that $\chi^F_{\Sigma(\delta)}(Y) \in v$. By definition of filtration $|w| R^{\mathcal{L},F}_{\Sigma(\delta)} |v|$ and by the Filtration Lemma $M^{\mathcal{L},F}_{\Sigma(\delta)}, |v| \models \chi^F_{\Sigma(\delta)}(Y)$. Finally, $\{|v|\} \in E^{\mathcal{L},F}_{\Sigma(\delta)}(|w|)(N)$ since $M^{\mathcal{L},F}_{\Sigma(\delta)}$ is standard.

This completes the proof of the Completeness theorem 37.

Corollary 62 (Finite Model Property) *The logic TPCL has the finite model property with respect to the class of models TrulyPlay.*

Bibliography

[1] J. Abdou and H. Keiding. *Effectivity Functions in Social Choice*. Kluwer Academic Publishers, 1991.

[2] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. In *FOCS '97: Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, page 100, Washington, DC, USA, 1997. IEEE Computer Society.

[3] M. A. Armstrong. *Groups and Symmetry*. Springer Science, 1998.

[4] K.J. Arrow. *Social Choice and Individual Values*. Yale University Press, 1970.

[5] R. Aumann. On the state of the art in game theory. *Games and Economic Behavior*, 24:181–210, 1998.

[6] R. J. Aumann. What is game theory trying to accomplish?, 1985. Frontiers of Economics, edited by K.Arrow and S.Honkapohja.

[7] R. J. Aumann. War and peace. Levine's Bibliography 321307000000000332, UCLA Department of Economics, September 2006.

[8] N. Belnap, M. Perloff, and M. Xu. *Facing The Future: Agents And Choices In Our Indeterminist World*. Oxford University Press, Usa, 2001.

[9] K. Binmore. *Playing for Real*. Oxford University Press, 2007.

[10] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science, 2001.

[11] P. Blackburn and J. van Benthem. Modal logic: A semantic perspective. *ETHICS*, 98:501–517, 1988.

[12] G. Boella, L. Sauro, and L. van der Torre. Admissible agreements among goal-directed agents. In *Proceedings of 2005 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'05)*, pages 543–554. IEEE Computer Society, 2005.

[13] G. Boella, L. Sauro, and L. van der Torre. Strengthening admissible coalitions. In *Proceeding of the 2006 conference on ECAI 2006: 17th European Conference on Artificial Intelligence*, pages 195–199. ACM, 2006.

[14] G. Boella, L. Sauro, and L. W. N. van der Torre. Reducing coalition structures via agreement specification. In Frank Dignum, Virginia Dignum, Sven Koenig, Sarit Kraus, Munindar P. Singh, and Michael Wooldridge, editors, *AAMAS*, pages 1187–1188. ACM, 2005.

[15] E. Bonzon, M.-C. Lagasquie-Schiex, and J. Lang. Dependencies between players in boolean games. *International Journal of Approximate Reasoning*, 50:899–914, 2009.

[16] S. Borgo. Coalitions in action logic. In *IJCAI*, pages 1822–1827, 2007.

[17] C. Boutilier. Toward a logic for qualitative decision theory. In *KR*, pages 75–86, 1994.

[18] J. Broersen, R. Mastop, J.J. Ch. Meyer, and P. Turrini. A deontic logic for socially optimal norms. In Ron van der Meyden and Leendert van der Torre, editors, *DEON*, volume 5076 of *Lecture Notes in Computer Science*, pages 218–232. Springer, 2008.

[19] J. Broersen, R. Mastop, J.J. Ch. Meyer, and P. Turrini. Determining the environment: A modal logic for closed interaction. *Synthese, special section of Knowledge, Rationality and Action*, 169(2):351–369, 2009.

[20] C. Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence*, 103:157–182, 1998.

[21] C. Castelfranchi, A. Cesta, and M. Miceli. Dependence relations among autonomous agents. In E. Werner and Y. Demazeau, editors, *Decentralized A.I.3*. Elsevier, 1992.

[22] K. Chatterjee, T. Henzinger, and N. Piterman. Strategy logic. Technical Report UCB/EECS-2007-78, University of California, Berkeley, May 2007.

[23] B. Chellas. *Modal Logic: an Introduction*. Cambridge University Press, 1980.

[24] J. Coleman. *Foundations of Social Theory*. Belknap Harvard, 1990.

[25] R. Conte and C. Castelfranchi. *Cognitive and Social Action*. UCL Press, 1995.

[26] P. Dunne, W. van der Hoek, S. Kraus, and M. Wooldridge. Cooperative boolean games. In *Proceedings of AAMAS 2008*, pages 1015–1022. ACM, 2008.

[27] P. Gardenfors. Rights, games and social choice. *Nous*, 15:341–56, 1981.

[28] D. Goldrei. *Classic Set Theory*. Chapman and Hall, 1998.

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

[30] V. Goranko, W. Jamroga, and P. Turrini. Strategic games and truly playable effectivity functions. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2011); Taipei, Taiwan, May 2-6, 2011*, 2011.

[31] V. Goranko and S. Passy. Using the universal modality: Gains and questions. *Journal of Logic and Computation*, 2(1):5–30, 1992.

[32] V. Goranko and G. van Drimmelen. Complete axiomatization and decidability of alternating-time temporal logic. *Theor. Comput. Sci.*, 353(1-3):93–117, 2006.

[33] D. Grossi. Unifying preference and judgment aggregation. In Carles Sierra, Cristiano Castelfranchi, Keith S. Decker, and Jaime Simão Sichman, editors, *AAMAS (1)*, pages 217–224. IFAAMAS, 2009.

[34] D. Grossi and P. Turrini. Dependence in games and dependence games. In *Third International Workshop on Computational Social Choice (COMSOC 2010); Duesseldorf, September 13-16, 2010*, pages 295–306, 2010. online proceedings, http://ccc.cs.uni-duesseldorf.de/COMSOC-2010/proceedings.shtml.

[35] D. Grossi and P. Turrini. Dependence theory via game theory. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1 - Volume 1*, AAMAS '10, pages 1147–1154, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.

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

[37] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, 2000.

[38] P. Harrenstein, W. van der Hoek, J.-J.Ch. Meyer, and C. Witteveen. Boolean games. In J. van Benthem, editor, *Proceedings of TARK'01*, pages 287–298. Morgan Kaufmann, 2001.

[39] H.H.Hansen. *Monotonic Modal Logics*. Master Thesis, Universiteit van Amsterdam, 2003.

[40] P. Highsmith. *Strangers on a Train*. Nationwide Book Service, 1950.

[41] J. Horty. *Deontic Logic and Agency*. Oxford University Press, 2001.

[42] Y. Kannai and B. Peleg. A note on the extension of an order on a set to the power set. *Journal of Economic Theory*, 32:172–175, 1984.

[43] B. Kooi and A.Tamminga. Conflicting obligations in multi-agent deontic logic. In Lou Goble and John-Jules Ch. Meyer, editors, *8th International Workshop on Deontic Logic in Computer Science (DEON 2006)*, pages 175–186. LNCS 4048, 2006.

[44] B. Kooi and A. Tamminga. Moral conflicts between groups of agents. *Journal of Philosophical Logic*, 37(1):1–21, 2008.

[45] S. Kuhn. Prisoner's dilemma. In *Stanford Encyclopedia of Philosophy*. 2007.

[46] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. ILLC Dissertation Series, 2008.

[47] J.J. Ch. Meyer and R. J. Wieringa. Deontic logic: a concise overview. pages 3–16. John Wiley and Sons Ltd., Chichester, UK, UK, 1993.

[48] J.J.Ch. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame J. of Formal Logic*, 29(1):109–136, 1988.

[49] H. Moulin. *The Strategy of Social Choice*. Advanced Textbooks in Economics, North Holland, 1983.

[50] H. Moulin and B. Peleg. Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics*, 10:115–145, 1982.

[51] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[52] R. Parikh. The logic of games and its applications. In *Selected papers of the international conference on "foundations of computation theory" on Topics in the theory of computation*, pages 111–139, New York, NY, USA, 1985. Elsevier North-Holland, Inc.

[53] R. Parikh. Social software. *Synthese*, 132(3):187–211, 2002.

[54] M. Pauly. *Logic for Social Software*. ILLC Dissertation Series, 2001.

[55] M. Pauly and R. Parikh. Game logic - an overview. *Studia Logica*, 75(2):165–182, 2003.

[56] B. Peleg. Effectivity functions, game forms, games and rights. *Social Choice and Welfare*, 15:67–80, 1998.

[57] L. Sauro. Qualitative criteria of admissibility for enforced agreements. *Computational & Mathematical Organization Theory*, 12(2-3):147–168, 2006.

[58] L. Sauro, L. van der Torre, and S. Villata. Dependency in cooperative boolean games. In A Håkansson, N. Nguyen, R. Hartung, R. Howlett, and L. Jain, editors, *Proceedings of KES-AMSTA 2009*, volume 5559 of *LNAI*, pages 1–10. Springer, 2009.

[59] J. Sichman. Depint: Dependence-based coalition formation in an open multi-agent scenario. *Journal of Artificial Societies and Social Simulation*, 1(2), 1998.

[60] J. Sichman and R. Conte. Multi-agent dependence by dependence graphs. In *Proceedings of AAMAS 2002*, ACM, pages 483–490, 2002.

[61] P. Turrini, D. Grossi, J. Broersen, and J.J. Ch. Meyer. Forbidding undesirable agreements: A dependence-based approach to the regulation of multi-agent systems. In Guido Governatori and Giovanni Sartor, editors, *DEON*, volume 6181 of *Lecture Notes in Computer Science*, pages 306–322. Springer, 2010.

[62] J. van Benthem. In praise of strategies. Research Report, http://www.illc.uva.nl/Publications/ResearchReports/PP-2008-03.text.pdf, 2007.

[63] J. van Benthem. For the better or for the worse: dynamic logic of preference. Research Report, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.158.7545, 2008.

[64] J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 14, 2004.

[65] J. van Benthem, O. Roy, and P. Girard. Everything else being equal: A modal logic approach to ceteris paribus preferences. ILLC Report PP-2007-09, 2007.

[66] J. van Benthem, O. Roy, and P. Girard. Everything else being equal: A modal logic approach to ceteris paribus preferences. Journal of Philosophical Logic, Volume 38, Number 1, p. 83-125, 2009.

[67] D. van Dalen. *Logic and Structure*. Springer, 1980.

[68] W. van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 157–164, New York, NY, USA, 2005. ACM.

[69] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Synthese Library, 2007.

[70] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[71] G. H. von Wright. *The logic of preference*. Edimburgh University Press, 1963.

[72] G. H. von Wright. Is there a logic of norms? *Ratio Juris*, 4(59):265–283, 1991.

[73] G.H. von Wright. The logic of preference reconsidered. *Theory and Decision*, 3:140–169, 1972.

[74] D. Walther, W. van der Hoek, and M. Wooldridge. Alternating-time temporal logic with explicit strategies. In Dov Samet, editor, *TARK*, pages 269–278, 2007.

Strategic Reasoning in Interdependence:
Logical and Game-Theoretical Investigations — *Summary*

Game theory is the branch of economics that studies interactive decision making, i.e. how entities that can reasonably be described as players of a game (e.g. a company that needs to choose the price of a new product, a country that should decide whether to withdraw from an occupied country, a PhD student who is about to decide whether to apply for a Postdoc position etc.) should behave, given their preferences and their information (e.g. the company wanting to attract a large portion of population but knowing that the new product is not perceived as useful by many potential customers, the country wanting to cut military spending but not knowing whether the local government could alone secure its own territory, the PhD student wanting to work in a sunny country and being aware of the weather conditions of his future workplace). Game theory is usually divided into two main branches: *non-cooperative game theory*, that studies the strategies that individuals should employ to reach their own goals, and *cooperative game theory*, that studies instead the effects of individuals joining their forces and getting the most out of their collective strategies.

The present work lies somewhat in between the two sides of game theory and studies the relation between the behaviour of individuals and the behaviour of coalitions to which they belong. The first part of the thesis, called **Strategic Reasoning and Coalitional Games**, studies what it means for a coalition of players to choose the best among the available alternatives, in particular what it means for a coalition to *prefer* a strategy above another and in what circumstances are those strategies at a coalition's disposal. Think for instance of a chess player who is setting up an attack against the opposite king. He knows that each of its pieces has invidual strengths (e.g., the knight can go to a central square, the bishop can control an important diagonal), but he is also aware that their real power lies in their combined forces (e.g., the knight and the bishop can *together* control a central square on an important diagonal). His reasoning starts from an individual perspective but it suddenly shifts to a coalitional one, where notions such as preferences and strategies acquire a more elaborated meaning and display specific formal properties. The thesis investigates them adopting the standard tools of logic and game-theory. The second part of the thesis, called **Strategic Reasoning and Dependence Games**, elaborates further upon the study of coalitional reasoning, focusing on the *network of interdependence* underlying each collective decision. Consider once again the chess player who is

deciding what to move. He is perfectly aware that pieces do not always perfectly and harmoniously coordinate. At times they actually obstruct each other while at other times they may even need to sacrifice themselves for their king to survive a mating attack. Their interaction displays a thick network of dependence relations (i.e. what each piece can do for the others) which strongly influences the strategies that can be played. In the classical account of cooperative game theory however this important condition is simply not taken into account. The present work bridges this gap, constructing a theory of coalitional rationality based on the resolution of its underlying dependence relations. Concretely it studies the mathematical properties characterizing those coalitions that arise from their members taking mutual advantage of each other. Finally, it relates those properties to the classical study of collective decision making.

Strategische Redenering in Afhankelijkheid:
Logische en Speltheoretische Onderzoeken — *Samenvatting*

Speltheorie is de tak van economie die de interactieve besluitvorming bestudeert, d.w.z. hoe individuen die beschreven kunnen worden als spelers (bijv. een bedrijf dat de prijs moet bepalen voor een nieuw product, een land dat moet beslissen of het zich terugtrekt uit een bezet land, een AIO die op het punt staat een beslissing te nemen over de sollicitatie voor een Postdoc positie enz.) zich horen te gedragen, gegeven hun voorkeuren en hun informatie (bijv. het bedrijf dat een groot gedeelte van de bevolking wil aantrekken wetende dat dit nieuwe product niet ontvangen zal worden als zijnde nuttig door veel potentiële klanten, het land dat wil snijden in defensie uitgaven niet wetende of de locale bestuurders zelf hun grondgebied zouden kunnen beschermen, de AIO die wil werken in een zonnig land en zich bewust is van de weersomstandigheden van zijn toekomstige werkplek). Speltheorie is gewoonlijk verdeeld in twee hoofdtakken: *niet-coöperatieve speltheorie*, die de strategieën bestudeert die individuen moeten gebruiken om hun doelen te bereiken, en *coöperatieve speltheorie*, die daarentegen de effecten bestudeert van individuen die hun krachten bundelen en het beste halen uit hun collectieve strategieën.

Het huidige werk ligt enigszins in het midden van deze twee kanten van speltheorie en bestudeert de relatie tussen het gedrag van individuen en het gedrag van coalities waar ze toe behoren. Het eerste gedeelte van het proefschrift, genoemd **Strategische Redenering en Spellen van Coalities**, bestudeert wat het betekent voor een coalitie van spelers om de beste van de beschikbare alternatieven te kiezen, in het bijzonder wat het betekent voor de coalitie om *voorkeur te hebben* voor de ene strategie boven de andere en in wat voor omstandigheden deze strategieën tot de beschikking staan van de coalitie. Denk bijvoorbeeld aan een schaakspeler die een aanval aan het opzetten is tegen de koning van de tegenpartij. Hij weet dat elk van de stukken individuele sterktes hebben (bijv., het paard kan naar een centraal veld gaan, de loper kan een belangrijke diagonaal controleren), maar hij is zich ook bewust dat hun echte sterkte in hun gecombineerde sterktes ligt (bijv., het paard en de loper kunnen *samen* een centraal veld op een belangrijke diagonaal controleren). Zijn redenering begint vanuit een individueel perspectief maar verschuift plotseling naar een coalitioneel perspectief, waar noties zoals voorkeuren en strategieën een meer uitgewerkte betekenis verwerven en specifieke formele eigenschappen ten toon spreiden. Het proefschrift onderzoekt deze, en maakt gebruik van de standaard voorwerpen van logica en speltheorie. Het tweede gedeelte van dit

proefschrift, genoemd **Strategische Redenering en Spellen van Afhankelijkheid**, gaat verder in op de studie van redenering van coalities, en focust op het *netwerk van onderlinge afhankelijkheid* dat een collectieve beslissing onderligt. Denk opnieuw aan de schaakspeler die nadenkt over zijn zetten. Hij is zich er van bewust dat stukken niet altijd perfect en harmonieus samenwerken. Nu en dan belemmeren ze elkaar, terwijl ze zich in andere gevallen wellicht moeten opofferen om een mat te voorkomen. Hun interacties tonen een dicht netwerk van afhankelijkheidrelaties (d.w.z. wat elk stuk kan doen voor de andere stukken) die de te spelen strategieën sterk beïnvloed. In de klassieke modellen van speltheorie wordt deze belangrijke voorwaarde simpelweg niet meegenomen. Het huidige werk overbrugt dit en bouwt een theorie van coalitionele rationaliteit, gebaseerd op de oplossing van de onderliggende afhankelijkheid relaties. Concreet gezien bestudeert het de wiskundige eigenschappen van coalities die zijn ontstaan uit hun leden, welke wederzijds voordeel halen uit hun samenwerking. Tenslotte relateert het deze eigenschappen aan de klassieke modellen van collectieve besluitvorming.

Ragionamento Strategico in Interdipendenza:
Ricerche di Logica e Teoria dei Giochi — *Riassunto*

La Teoria dei Giochi è un ramo dell'Economia che studia le decisioni in interazione, cioè il modo in cui tutti gli individui che possono ragionevolmente essere descritti come dei giocatori (per esempio un'impresa che sceglie il prezzo di un nuovo prodotto, uno Stato che deve decidere sul ritiro da uno stato occupato, uno studente di dottorato che sta per decidere se fare domanda per un post-dottorato etc.) si debbano comportare, date le loro preferenze e le loro informazioni (per esempio, l'impresa che vuole attrarre una larga parte della popolazione ma che sa che il nuovo prodotto non è percepito come utile da molti potenziali compratori, lo Stato che vuole risparmiare sulle spese militari ma che non sa se il governo locale può da solo riuscire a garantire la sicurezza del suo territorio, lo studente di dottorato che vuole lavorare in uno luogo caldo e che sa quali sono le condizioni del meteo presso il suo futuro posto di lavoro). La Teoria dei Giochi è di solito divisa in due rami: la *Teoria dei Giochi non cooperativa*, che studia le strategie che gli individui dovrebbero adottare per raggiungere i loro scopi; e la *Teoria dei Giochi cooperativa*, che studia invece che cosa gli individui potrebbero raggiungere se unissero le forze e adottassero strategie collettive.

Questo lavoro si trova nel mezzo delle due facce della Teoria dei Giochi e studia la relazione tra il comportamento degli individui e quello delle coalizioni di cui questi individui sono membri. La prima parte, che si intitola **Ragionamento Strategico e Giochi di Coalizione**, studia cosa significa per una coalizione il fatto di dover scegliere la migliore delle strategie disponibili, in particolare cosa significa *preferire* una strategia ad un'altra, e in quali circostanze queste strategie sono disponibili. Si pensi per esempio a un giocatore di scacchi che sta per attaccare il re nemico. Lui sa che ogni suo pezzo ha dei punti di forza (per esempio, il cavallo può controllare una casella centrale, l'alfiere può controllare un'importante diagonale), ma sa anche bene che la loro vera forza sta nell'uso combinato delle loro abilita' (per esempio, il cavallo e l'alfiere possono *insieme* controllare una casella centrale in un'importante diagonale). Il suo ragionamento inizia da una prospettiva individuale ma di colpo si sposta a una prospettiva di coalizione, dove nozioni come preferenze e strategie acquistano significati più elaborati e mostrano specifiche proprietà formali. La tesi le discute, facendo uso di strumenti classici della Logica e della Teoria dei Giochi. La seconda parte della Tesi, che si intitola **Ragionamento Strategico e Giochi di Dipendenza**, elabora più a fondo lo studio della razionalità collettiva,

concentrandosi sulla *rete di interdipendenza* che si trova al di sotto di ogni decisione delle coalizioni. Si consideri di nuovo il giocatore di scacchi che sta decidendo cosa muovere. Lui sa che non sempre i suoi pezzi sono perfettamente e armoniosamente coordinati. A volte si ostruiscono a vicenda, altre volte addirittura devono essere sacrificati per consentire al loro re di sopravvivere. La loro interazione mostra una spessa rete di relazioni di dipendenza (ciò che ogni pezzo può fare per gli altri) che influenza le strategie che possono essere giocate. Nella visione classica della Teoria dei Giochi cooperativa però questa condizione non è assolutamente considerata. Il presente lavoro colma questa mancanza, costruendo una teoria della razionalità di coalizione basata sulla risoluzione delle sottostanti relazioni di dipendenza. Concretamente, studia le proprietà matematiche che caratterizzano quelle coalizioni che si formano dal muto vantaggio dei propri membri. In fine, collega queste proprietà allo studio classico della teoria delle decisioni collettive.

Curriculum Vitae

Paolo Turrini

03-02-1980 Born in Cagliari, Italy;

1994-1999 High School Studies, *Liceo Classico don Bosco*, Cagliari, Italy;

1999-2005 University Studies, *Communication Sciences Department*, University of Siena, Italy;

2005-2007 PhD student, *Cognitive Science Department*, University of Siena, Italy;

2007-2011 PhD student, *Center for Artificial Intelligence*, Utrecht University, The Netherlands;

2011-2013 AFR-Marie Curie Postdoctoral Fellow, *Faculté des Sciences, de la Technologie et de la Communication*, University of Luxembourg, Luxembourg.

SIKS Dissertation Series

1998

1998-1 | **Johan van den Akker** (CWI), DEGAS - An Active, Temporal Database of Autonomous Objects.

1998-2 | **Floris Wiesman** (UM), Information Retrieval by Graphically Browsing Meta-Information.

1998-3 | **Ans Steuten** (TUD), A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective.

1998-4 | **Dennis Breuker** (UM), Memory versus Search in Games.

1998-5 | **E.W. Oskamp** (RUL), Computerondersteuning bij Straftoemeting.

1999

1999-1 | **Mark Sloof** (VU), Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products.

1999-2 | **Rob Potharst** (EUR), Classification using decision trees and neural nets.

1999-3 | **Don Beal** (UM), The Nature of Minimax Search.

1999-4 | **Jacques Penders** (UM), The practical Art of Moving Physical Objects.

1999-5 | **Aldo de Moor** (KUB), Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems.

1999-6 | **Niek J.E. Wijngaards** (VU), Re-design of compositional systems.

1999-7 | **David Spelt** (UT), Verification support for object database design.

1999-8 | **Jacques H.J. Lenting** (UM), Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation.

2000

2000-1 | **Frank Niessink** (VU), Perspectives on Improving Software Maintenance.

2000-2 | **Koen Holtman** (TUE), Prototyping of CMS Storage Management.

2000-3 | **Carolien M.T. Metselaar** (UVA), Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief.

2000-4 | **Geert de Haan** (VU), ETAG, A Formal Model of Competence Knowledge for User Interface Design.

2000-5 | **Ruud van der Pol** (UM), Knowledge-based Query Formulation in Information Retrieval.

2000-6 | **Rogier van Eijk** (UU), Programming Languages for Agent Communication.

2000-7 | **Niels Peek** (UU), Decision-theoretic Planning of Clinical Patient Management.

2000-8 | **Veerle Coup** (EUR), Sensitivity Analyis of Decision-Theoretic Networks.

2000-9 | **Florian Waas** (CWI), Principles of Probabilistic Query Optimization.

2000-10 | **Niels Nes** (CWI), Image Database Management System Design Considerations, Algorithms and Architecture.

2000-11 | **Jonas Karlsson** (CWI), Scalable Distributed Data Structures for Database Management.

2001

2001-1 | **Silja Renooij** (UU), Qualitative Approaches to Quantifying Probabilistic Networks.

2001-2 | **Koen Hindriks** (UU), Agent Programming Languages: Programming with Mental Models.

2001-3 | **Maarten van Someren** (UvA), Learning as problem solving.

2001-4 | **Evgueni Smirnov** (UM), Conjunctive and

Disjunctive Version Spaces with Instance-Based Boundary Sets.

2001-5 | **Jacco van Ossenbruggen** (VU), Processing Structured Hypermedia: A Matter of Style.

2001-6 | **Martijn van Welie** (VU), Task-based User Interface Design.

2001-7 | **Bastiaan Schonhage** (VU), Diva: Architectural Perspectives on Information Visualization.

2001-8 | **Pascal van Eck** (VU), A Compositional Semantic Structure for Multi-Agent Systems Dynamics.

2001-9 | **Pieter Jan 't Hoen** (RUL), Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes.

2001-10 | **Maarten Sierhuis** (UvA), Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design.

2001-11 | **Tom M. van Engers** (VUA), Knowledge Management: The Role of Mental Models in Business Systems Design.

2002

2002-01 | **Nico Lassing** (VU), Architecture-Level Modifiability Analysis.

2002-02 | **Roelof van Zwol** (UT), Modelling and searching web-based document collections.

2002-03 | **Henk Ernst Blok** (UT), Database Optimization Aspects for Information Retrieval.

2002-04 | **Juan Roberto Castelo Valdueza** (UU), The Discrete Acyclic Digraph Markov Model in Data Mining.

2002-05 | **Radu Serban** (VU), The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents.

2002-06 | **Laurens Mommers** (UL), Applied legal epistemology; Building a knowledge-based ontology of the legal domain.

2002-07 | **Peter Boncz** (CWI), Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications.

2002-08 | **Jaap Gordijn** (VU), Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas.

2002-09 | **Willem-Jan van den Heuvel** (KUB), Integrating Modern Business Applications with Objectified Legacy Systems.

2002-10 | **Brian Sheppard** (UM), Towards Perfect Play of Scrabble.

2002-11 | **Wouter C.A. Wijngaards** (VU), Agent Based Modelling of Dynamics: Biological and Organisational Applications.

2002-12 | **Albrecht Schmidt** (UVA), Processing XML in Database Systems.

2002-13 | **Hongjing Wu** (TUE), A Reference Architecture for Adaptive Hypermedia Applications.

2002-14 | **Wieke de Vries** (UU), Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems.

2002-15 | **Rik Eshuis** (UT), Semantics and Verification of UML Activity Diagrams for Workflow Modelling.

2002-16 | **Pieter van Langen** (VU), The Anatomy of Design: Foundations, Models and Applications.

2002-17 | **Stefan Manegold** (UVA), Understanding, Modeling, and Improving Main-Memory Database Performance.

2003

2003-01 | **Heiner Stuckenschmidt** (VU), Onotology-Based Information Sharing In Weakly Structured Environments.

2003-02 | **Jan Broersen** (VU), Modal Action Logics for Reasoning About Reactive Systems.

2003-03 | **Martijn Schuemie** (TUD), Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy.

2003-04 | **Milan Petkovic** (UT), Content-Based Video Retrieval Supported by Database Technology.

2003-05 | **Jos Lehmann** (UVA), Causation in Artificial Intelligence and Law - A modelling approach.

2003-06 | **Boris van Schooten** (UT), Development and specification of virtual environments.

2003-07 | **Machiel Jansen** (UvA), Formal Explorations of Knowledge Intensive Tasks.

2003-08 | **Yongping Ran** (UM), Repair Based Scheduling.

2003-09 | **Rens Kortmann** (UM), The resolution of visually guided behaviour.

2003-10 | **Andreas Lincke** (UvT), Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture.

2003-11 | **Simon Keizer** (UT), Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks.

2003-12 | **Roeland Ordelman** (UT), Dutch speech

recognition in multimedia information retrieval.

2003-13 | **Jeroen Donkers** (UM), Nosce Hostem - Searching with Opponent Models.

2003-14 | **Stijn Hoppenbrouwers** (KUN), Freezing Language: Conceptualisation Processes across ICT-Supported Organisations.

2003-15 | **Mathijs de Weerdt** (TUD), Plan Merging in Multi-Agent Systems.

2003-16 | **Menzo Windhouwer** (CWI), Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses.

2003-17 | **David Jansen** (UT), Extensions of Statecharts with Probability, Time, and Stochastic Timing.

2003-18 | **Levente Kocsis** (UM), Learning Search Decisions.

2004

2004-01 | **Virginia Dignum** (UU), A Model for Organizational Interaction: Based on Agents, Founded in Logic.

2004-02 | **Lai Xu** (UvT), Monitoring Multi-party Contracts for E-business.

2004-03 | **Perry Groot** (VU), A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving.

2004-04 | **Chris van Aart** (UVA), Organizational Principles for Multi-Agent Architectures.

2004-05 | **Viara Popova** (EUR), Knowledge discovery and monotonicity.

2004-06 | **Bart-Jan Hommes** (TUD), The Evaluation of Business Process Modeling Techniques.

2004-07 | **Elise Boltjes** (UM), Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes.

2004-08 | **Joop Verbeek** (UM), Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise.

2004-09 | **Martin Caminada** (VU), For the Sake of the Argument; explorations into argument-based reasoning.

2004-10 | **Suzanne Kabel** (UVA), Knowledge-rich indexing of learning-objects.

2004-11 | **Michel Klein** (VU), Change Management for Distributed Ontologies.

2004-12 | **The Duy Bui** (UT), Creating emotions and facial expressions for embodied agents.

2004-13 | **Wojciech Jamroga** (UT), Using Multiple Models of Reality: On Agents who Know how to Play.

2004-14 | **Paul Harrenstein** (UU), Logic in Conflict. Logical Explorations in Strategic Equilibrium.

2004-15 | **Arno Knobbe** (UU), Multi-Relational Data Mining.

2004-16 | **Federico Divina** (VU), Hybrid Genetic Relational Search for Inductive Learning.

2004-17 | **Mark Winands** (UM), Informed Search in Complex Games.

2004-18 | **Vania Bessa Machado** (UvA), Supporting the Construction of Qualitative Knowledge Models.

2004-19 | **Thijs Westerveld** (UT), Using generative probabilistic models for multimedia retrieval.

2004-20 | **Madelon Evers** (Nyenrode), Learning from Design: facilitating multidisciplinary design teams.

2005

2005-01 | **Floor Verdenius** (UVA), Methodological Aspects of Designing Induction-Based Applications.

2005-02 | **Erik van der Werf** (UM), AI techniques for the game of Go.

2005-03 | **Franc Grootjen** (RUN), A Pragmatic Approach to the Conceptualisation of Language.

2005-04 | **Nirvana Meratnia** (UT), Towards Database Support for Moving Object data.

2005-05 | **Gabriel Infante-Lopez** (UVA), Two-Level Probabilistic Grammars for Natural Language Parsing.

2005-06 | **Pieter Spronck** (UM), Adaptive Game AI.

2005-07 | **Flavius Frasincar** (TUE), Hypermedia Presentation Generation for Semantic Web Information Systems.

2005-08 | **Richard Vdovjak** (TUE), A Model-driven Approach for Building Distributed Ontology-based Web Applications.

2005-09 | **Jeen Broekstra** (VU), Storage, Querying and Inferencing for Semantic Web Languages.

2005-10 | **Anders Bouwer** (UVA), Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments.

2005-11 | **Elth Ogston** (VU), Agent Based Matchmaking and Clustering - A Decentralized Approach to Search.

2005-12 | **Csaba Boer** (EUR), Distributed Simulation in Industry.

2005-13 | **Fred Hamburg** (UL), Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen.

2005-14 | **Borys Omelayenko** (VU), Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics.

2005-15 | **Tibor Bosse** (VU), Analysis of the Dynamics of Cognitive Processes.

2005-16 | **Joris Graaumans** (UU), Usability of XML Query Languages.

2005-17 | **Boris Shishkov** (TUD), Software Specification Based on Re-usable Business Components.

2005-18 | **Danielle Sent** (UU), Test-selection strategies for probabilistic networks.

2005-19 | **Michel van Dartel** (UM), Situated Representation.

2005-20 | **Cristina Coteanu** (UL), Cyber Consumer Law, State of the Art and Perspectives.

2005-21 | **Wijnand Derks** (UT), Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics.

2006

2006-01 | **Samuil Angelov** (TUE), Foundations of B2B Electronic Contracting.

2006-02 | **Cristina Chisalita** (VU), Contextual issues in the design and use of information technology in organizations.

2006-03 | **Noor Christoph** (UVA), The role of metacognitive skills in learning to solve problems.

2006-04 | **Marta Sabou** (VU), Building Web Service Ontologies.

2006-05 | **Cees Pierik** (UU), Validation Techniques for Object-Oriented Proof Outlines.

2006-06 | **Ziv Baida** (VU), Software-aided Service Bundling – Intelligent Methods & Tools for Graphical Service Modeling.

2006-07 | **Marko Smiljanic** (UT), XML schema matching – balancing efficiency and effectiveness by means of clustering.

2006-08 | **Eelco Herder** (UT), Forward, Back and Home Again – Analyzing User Behavior on the Web.

2006-09 | **Mohamed Wahdan** (UM), Automatic Formulation of the Auditor's Opinion.

2006-10 | **Ronny Siebes** (VU), Semantic Routing in Peer-to-Peer Systems.

2006-11 | **Joeri van Ruth** (UT), Flattening Queries over Nested Data Types.

2006-12 | **Bert Bongers** (VU), Interactivation – Towards an e-cology of people, our technological environment, and the arts.

2006-13 | **Henk-Jan Lebbink** (UU), Dialogue and Decision Games for Information Exchanging Agents.

2006-14 | **Johan Hoorn** (VU), Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change.

2006-15 | **Rainer Malik** (UU), CONAN: Text Mining in the Biomedical Domain.

2006-16 | **Carsten Riggelsen** (UU), Approximation Methods for Efficient Learning of Bayesian Networks.

2006-17 | **Stacey Nagata** (UU), User Assistance for Multitasking with Interruptions on a Mobile Device.

2006-18 | **Valentin Zhizhkun** (UVA), Graph transformation for Natural Language Processing.

2006-19 | **Birna van Riemsdijk** (UU), Cognitive Agent Programming: A Semantic Approach.

2006-20 | **Marina Velikova** (UvT), Monotone models for prediction in data mining.

2006-21 | **Bas van Gils** (RUN), Aptness on the Web.

2006-22 | **Paul de Vrieze** (RUN), Fundaments of Adaptive Personalisation.

2006-23 | **Ion Juvina** (UU), Development of Cognitive Model for Navigating on the Web.

2006-24 | **Laura Hollink** (VU), Semantic Annotation for Retrieval of Visual Resources.

2006-25 | **Madalina Drugan** (UU), Conditional log-likelihood MDL and Evolutionary MCMC.

2006-26 | **Vojkan Mihajlovic** (UT), Score Region Algebra: A Flexible Framework for Structured Information Retrieval.

2006-27 | **Stefano Bocconi** (CWI), Vox Populi: generating video documentaries from semantically annotated media repositories.

2006-28 | **Borkur Sigurbjornsson** (UVA), Focused Information Access using XML Element Retrieval.

2007

2007-01 | **Kees Leune** (UvT), Access Control and Service-Oriented Architectures.

2007-02 | **Wouter Teepe** (RUG), Reconciling Information Exchange and Confidentiality: A Formal Approach.

2007-03 | **Peter Mika** (VU), Social Networks and the Semantic Web.

2007-04 | **Jurriaan van Diggelen** (UU), Achieving Semantic Interoperability in Multi-agent Systems: A Dialogue-based Approach.

2007-05 | **Bart Schermer** (UL), Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance.

2007-06 | **Gilad Mishne** (UVA), Applied Text Analytics for Blogs.

2007-07 | **Natasa Jovanovic** (UT), To Who It May Concern - Addressee Identification in Face-to-Face Meetings.

2007-08 | **Mark Hoogendoorn** (VU), Modeling of Change in Multi-Agent Organizations.

2007-09 | **David Mobach** (VU), Agent-Based Mediated Service Negotiation.

2007-10 | **Huib Aldewereld** (UU), Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols.

2007-11 | **Natalia Stash** (TUE), Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System.

2007-12 | **Marcel van Gerven** (RUN), Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty.

2007-13 | **Rutger Rienks** (UT), Meetings in Smart Environments; Implications of Progressing Technology.

2007-14 | **Niek Bergboer** (UM), Context-Based Image Analysis.

2007-15 | **Joyca Lacroix** (UM), NIM: a Situated Computational Memory Model.

2007-16 | **Davide Grossi** (UU), Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems.

2007-17 | **Theodore Charitos** (UU), Reasoning with Dynamic Networks in Practice.

2007-18 | **Bart Orriens** (UvT), On the development an management of adaptive business collaborations.

2007-19 | **David Levy** (UM), Intimate relationships with artificial partners.

2007-20 | **Slinger Jansen** (UU), Customer Configuration Updating in a Software Supply Network.

2007-21 | **Karianne Vermaas** (UU), Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005.

2007-22 | **Zlatko Zlatev** (UT), Goal-oriented design of value and process models from patterns.

2007-23 | **Peter Barna** (TUE), Specification of Application Logic in Web Information Systems.

2007-24 | **Georgina Ramï¿½rez Camps** (CWI), Structural Features in XML Retrieval.

2007-25 | **Joost Schalken** (VU), Empirical Investigations in Software Process Improvement.

2008

2008-01 | **Katalin Boer-Sorbuon** (EUR), Agent-Based Simulation of Financial Markets: A modular,continuous-time approach.

2008-02 | **Alexei Sharpanskykh** (VU), On Computer-Aided Methods for Modeling and Analysis of Organizations.

2008-03 | **Vera Hollink** (UVA), Optimizing hierarchical menus: a usage-based approach.

2008-04 | **Ander de Keijzer** (UT), Management of Uncertain Data - towards unattended integration.

2008-05 | **Bela Mutschler** (UT), Modeling and simulating causal dependencies on process-aware information systems from a cost perspective.

2008-06 | **Arjen Hommersom** (RUN), On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective.

2008-07 | **Peter van Rosmalen** (OU), Supporting the tutor in the design and support of adaptive e-learning.

2008-08 | **Janneke Bolt** (UU), Bayesian Networks: Aspects of Approximate Inference.

2008-09 | **Christof van Nimwegen** (UU), The paradox of the guided user: assistance can be counter-effective.

2008-10 | **Wauter Bosma** (UT), Discourse oriented summarization.

2008-11 | **Vera Kartseva** (VU), Designing Controls for Network Organizations: A Value-Based Approach.

2008-12 | **Jozsef Farkas** (RUN), A Semiotically Oriented Cognitive Model of Knowledge Representation.

2008-13 | **Caterina Carraciolo** (UVA), Topic Driven Access to Scientific Handbooks.

2008-14 | **Arthur van Bunningen** (UT), Context-Aware Querying; Better Answers with Less Effort.

2008-15 | **Martijn van Otterlo** (UT), The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains.

2008-16 | **Henriette van Vugt** (VU), Embodied agents from a user's perspective.

2008-17 | **Martin Op 't Land** (TUD), Applying Architecture and Ontology to the Splitting and Allying of Enterprises.

2008-18 | **Guido de Croon** (UM), Adaptive Active Vision.

2008-19 | **Henning Rode** (UT), From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search.

2008-20 | **Rex Arendsen** (UVA), Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven.

2008-21 | **Krisztian Balog** (UVA), People Search in the Enterprise.

2008-22 | **Henk Koning** (UU), Communication of IT-Architecture.

2008-23 | **Stefan Visscher** (UU), Bayesian network models for the management of ventilator-associated pneumonia.

2008-24 | **Zharko Aleksovski** (VU), Using background knowledge in ontology matching.

2008-25 | **Geert Jonker** (UU), Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency.

2008-26 | **Marijn Huijbregts** (UT), Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled.

2008-27 | **Hubert Vogten** (OU), Design and Implementation Strategies for IMS Learning Design.

2008-28 | **Ildiko Flesch** (RUN), On the Use of Independence Relations in Bayesian Networks.

2008-29 | **Dennis Reidsma** (UT), Annotations and Subjective Machines - Of Annotators, Embodied Agents, Users, and Other Humans.

2008-30 | **Wouter van Atteveldt** (VU), Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content.

2008-31 | **Loes Braun** (UM), Pro-Active Medical Information Retrieval.

2008-32 | **Trung H. Bui** (UT), Toward Affective Dialogue Management using Partially Observable Markov Decision Processes.

2008-33 | **Frank Terpstra** (UVA), Scientific Workflow Design; theoretical and practical issues.

2008-34 | **Jeroen de Knijf** (UU), Studies in Frequent Tree Mining.

2008-35 | **Ben Torben Nielsen** (UvT), Dendritic morphologies: function shapes structure.

2009

2009-01 | **Rasa Jurgelenaite** (RUN), Symmetric Causal Independence Models.

2009-02 | **Willem Robert van Hage** (VU), Evaluating Ontology-Alignment Techniques.

2009-03 | **Hans Stol** (UvT), A Framework for Evidence-based Policy Making Using IT.

2009-04 | **Josephine Nabukenya** (RUN), Improving the Quality of Organisational Policy Making using Collaboration Engineering.

2009-05 | **Sietse Overbeek** (RUN), Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality.

2009-06 | **Muhammad Subianto** (UU), Understanding Classification.

2009-07 | **Ronald Poppe** (UT), Discriminative Vision-Based Recovery and Recognition of Human Motion.

2009-08 | **Volker Nannen** (VU), Evolutionary Agent-Based Policy Analysis in Dynamic Environments.

2009-09 | **Benjamin Kanagwa** (RUN), Design, Discovery and Construction of Service-oriented Systems.

2009-10 | **Jan Wielemaker** (UVA), Logic programming for knowledge-intensive interactive applications.

2009-11 | **Alexander Boer** (UVA), Legal Theory, Sources of Law & the Semantic Web.

2009-12 | **Peter Massuthe** (TUE, Humboldt-Universitaet zu Berlin), Perating Guidelines for Services.

2009-13 | **Steven de Jong** (UM), Fairness in Multi-Agent Systems.

2009-14 | **Maksym Korotkiy** (VU), From ontology-enabled services to service-enabled ontologies. making ontologies work in e-science with ONTO-SOA

2009-15 | **Rinke Hoekstra** (UVA), Ontology Representation - Design Patterns and Ontologies that Make Sense.

2009-16 | **Fritz Reul** (UvT), New Architectures in Computer Chess.

2009-17 | **Laurens van der Maaten** (UvT), Feature Extraction from Visual Data.

2009-18 | **Fabian Groffen** (CWI), Armada, An

Evolving Database System.

2009-19 | **Valentin Robu** (CWI), Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets.

2009-20 | **Bob van der Vecht** (UU), Adjustable Autonomy: Controling Influences on Decision Making.

2009-21 | **Stijn Vanderlooy** (UM), Ranking and Reliable Classification.

2009-22 | **Pavel Serdyukov** (UT), Search For Expertise: Going beyond direct evidence.

2009-23 | **Peter Hofgesang** (VU), Modelling Web Usage in a Changing Environment.

2009-24 | **Annerieke Heuvelink** (VUA), Cognitive Models for Training Simulations.

2009-25 | **Alex van Ballegooij** (CWI), "RAM: Array Database Management through Relational Mapping".

2009-26 | **Fernando Koch** (UU), An Agent-Based Model for the Development of Intelligent Mobile Services.

2009-27 | **Christian Glahn** (OU), Contextual Support of Social Engagement and Reflection on the Web.

2009-28 | **Sander Evers** (UT), Sensor Data Management with Probabilistic Models.

2009-29 | **Stanislav Pokraev** (UT), Model-Driven Semantic Integration of Service-Oriented Applications.

2009-30 | **Marcin Zukowski** (CWI), Balancing vectorized query execution with bandwidth-optimized storage.

2009-31 | **Sofiya Katrenko** (UVA), A Closer Look at Learning Relations from Text.

2009-32 | **Rik Farenhorst and Remco de Boer** (VU), Architectural Knowledge Management: Supporting Architects and Auditors.

2009-33 | **Khiet Truong** (UT), How Does Real Affect Affect Affect Recognition In Speech?.

2009-34 | **Inge van de Weerd** (UU), Advancing in Software Product Management: An Incremental Method Engineering Approach.

2009-35 | **Wouter Koelewijn** (UL), Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling.

2009-36 | **Marco Kalz** (OUN), Placement Support for Learners in Learning Networks.

2009-37 | **Hendrik Drachsler** (OUN), Navigation Support for Learners in Informal Learning Networks.

2009-38 | **Riina Vuorikari** (OU), Tags and self-organisation: a metadata ecology for learning resources in a multilingual context.

2009-39 | **Christian Stahl** (TUE, Humboldt-Universitaet zu Berlin), Service Substitution – A Behavioral Approach Based on Petri Nets.

2009-40 | **Stephan Raaijmakers** (UvT), Multinomial Language Learning: Investigations into the Geometry of Language.

2009-41 | **Igor Berezhnyy** (UvT), Digital Analysis of Paintings.

2009-42 | **Toine Bogers** (UvT), Recommender Systems for Social Bookmarking.

2009-43 | **Virginia Nunes Leal Franqueira** (UT), Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients.

2009-44 | **Roberto Santana Tapia** (UT), Assessing Business-IT Alignment in Networked Organizations.

2009-45 | **Jilles Vreeken** (UU), Making Pattern Mining Useful.

2009-46 | **Loredana Afanasiev** (UvA), Querying XML: Benchmarks and Recursion.

2010

2010-01 | **Matthijs van Leeuwen** (UU), Patterns that Matter.

2010-02 | **Ingo Wassink** (UT), Work flows in Life Science.

2010-03 | **Joost Geurts** (CWI), A Document Engineering Model and Processing Framework for Multimedia documents.

2010-04 | **Olga Kulyk** (UT), Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments.

2010-05 | **Claudia Hauff** (UT), Predicting the Effectiveness of Queries and Retrieval Systems.

2010-06 | **Sander Bakkes** (UvT), Rapid Adaptation of Video Game AI.

2010-07 | **Wim Fikkert** (UT), A Gesture interaction at a Distance.

2010-08 | **Krzysztof Siewicz** (UL), Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments.

2010-09 | **Hugo Kielman** (UL), Politiële gegevensverwerking en Privacy, Naar een effectieve waarborging.

2010-10 | **Rebecca Ong** (UL), Mobile Communication and Protection of Children.

2010-11 | **Adriaan Ter Mors** (TUD), The world according to MARP: Multi-Agent Route Planning.

2010-12 | **Susan van den Braak** (UU), Sensemaking software for crime analysis.

2010-13 | **Gianluigi Folino** (RUN), High Performance Data Mining using Bio-inspired techniques.

2010-14 | **Sander van Splunter** (VU), Automated Web Service Reconfiguration.

2010-15 | **Lianne Bodenstaff** (UT), Managing Dependency Relations in Inter-Organizational Models.

2010-16 | **Sicco Verwer** (TUD), Efficient Identification of Timed Automata, theory and practice.

2010-17 | **Spyros Kotoulas** (VU), Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications.

2010-18 | **Charlotte Gerritsen** (VU), Caught in the Act: Investigating Crime by Agent-Based Simulation.

2010-19 | **Henriette Cramer** (UvA), People's Responses to Autonomous and Adaptive Systems.

2010-20 | **Ivo Swartjes** (UT), Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative.

2010-21 | **Harold van Heerde** (UT), Privacy-aware data management by means of data degradation.

2010-22 | **Michiel Hildebrand** (CWI), End-user Support for Access to Heterogeneous Linked Data.

2010-23 | **Bas Steunebrink** (UU), The Logical Structure of Emotions.

2010-24 | **Dmytro Tykhonov** (), Designing Generic and Efficient Negotiation Strategies.

2010-25 | **Zulfiqar Ali Memon** (VU), Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective.

2010-26 | **Ying Zhang** (CWI), XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines.

2010-27 | **Marten Voulon** (UL), Automatisch contracteren.

2010-28 | **Arne Koopman** (UU), Characteristic Relational Patterns.

2010-29 | **Stratos Idreos** (CWI), Database Cracking: Towards Auto-tuning Database Kernels.

2010-30 | **Marieke van Erp** (UvT), Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval.

2010-31 | **Victor de Boer** (UVA), Ontology Enrichment from Heterogeneous Sources on the Web.

2010-32 | **Marcel Hiel** (UvT), An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems.

2010-33 | **Robin Aly** (UT), Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval.

2010-34 | **Teduh Dirgahayu** (UT), Interaction Design in Service Compositions.

2010-35 | **Dolf Trieschnigg** (UT), Proof of Concept: Concept-based Biomedical Information Retrieval.

2010-36 | **Jose Janssen** (OU), Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification.

2010-37 | **Niels Lohmann** (TUE), Correctness of services and their composition.

2010-38 | **Dirk Fahland** (TUE), From Scenarios to components.

2010-39 | **Ghazanfar Farooq Siddiqui** (VU), Integrative modeling of emotions in virtual agents.

2010-40 | **Mark van Assem** (VU), Converting and Integrating Vocabularies for the Semantic Web.

2010-41 | **Guillaume Chaslot** (UM), Monte-Carlo Tree Search.

2010-42 | **Sybren de Kinderen** (VU), Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach.

2010-43 | **Peter van Kranenburg** (UU), A Computational Approach to Content-Based Retrieval of Folk Song Melodies.

2010-44 | **Pieter Bellekens** (TUE), An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain.

2010-45 | **Vasilios Andrikopoulos** (UvT), A theory and model for the evolution of software services.

2010-46 | **Vincent Pijpers** (VU), e3alignment: Exploring Inter-Organizational Business-ICT Alignment.

2010-47 | **Chen Li** (UT), Mining Process Model Variants: Challenges, Techniques, Examples.

2010-48 | **Milan Lovric** (EUR), Behavioral Finance and Agent-Based Artificial Markets.

2010-49 | **Jahn-Takeshi Saito** (UM), Solving difficult game positions.

2010-50 | **Bouke Huurnink** (UVA), Search in Audiovisual Broadcast Archives.

2010-51 | **Alia Khairia Amin** (CWI), Understanding and supporting information seeking tasks in multiple sources.

2010-52 | **Peter-Paul van Maanen** (VU), Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention.

2010-53 | **Edgar Meij** (UVA), Combining Concepts and Language Models for Information Access.

2011

2011-01 | **Botond Cseke** (RUN), Variational Algorithms for Bayesian Inference in Latent Gaussian Models.

2011-02 | **Nick Tinnemeier** (UU), Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language.

2011-03 | **Jan Martijn van der Werf** (TUE), Compositional Design and Verification of Component-Based Information Systems.

2011-04 | **Hado van Hasselt** (UU), Insights in Reinforcement Learning - Formal analysis and empirical evaluation of temporal-difference learning algorithms.

2011-05 | **Base van der Raadt** (VU), Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

2011-06 | **Yiwen Wang** (TUE), Semantically-Enhanced Recommendations in Cultural Heritage.

2011-07 | **Yujia Cao** (UT), Multimodal Information Presentation for High Load Human Computer Interaction.

2011-08 | **Nieske Vergunst** (UU), BDI-based Generation of Robust Task-Oriented Dialogues.

2011-09 | **Tim de Jong** (OU), Contextualised Mobile Media for Learning.

2011-10 | **Bart Bogaert** (UvT), Cloud Content Contention.

2011-11 | **Dhaval Vyas** (UT), Designing for Awareness: An Experience-focused HCI Perspective.

2011-12 | **Carmen Bratosin** (TUE), Grid Architecture for Distributed Process Mining.

2011-13 | **Xiaoyu Mao** (UvT), Airport under Control; Multiagent Scheduling for Airport Ground Handling.

2011-14 | **Milan Lovric** (EUR), Behavioral Finance and Agent-Based Artificial Markets.

2011-15 | **Marijn Koolen** (UVA), The Meaning of Structure: the Value of Link Evidence for Information Retrieval.

2011-16 | **Maarten Schadd** (UM), Selective Search in Games of Different Complexity.

2011-17 | **Jiyin He** (UVA), Exploring Topic Structure: Coherence, Diversity and Relatedness.

2011-18 | **Mark Ponsen** (UM), Strategic Decision-Making in complex games.

2011-19 | **Ellen Rusman** (OU), The Mind's Eye on Personal Profiles.

2011-20 | **Qing Gu** (VU), Guiding service-oriented software engineering - A view-based approach.

2011-21 | **Linda Terlouw** (TUD), Modularization and Specification of Service-Oriented Systems.

2011-22 | **Junte Zhang** (UVA), System Evaluation of Archival Description and Access.

2011-23 | **Wouter Weerkamp** (UVA), Finding People and their Utterances in Social Media.

2011-24 | **Herwin van Welbergen** (UT), Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior.

2011-25 | **Syed Waqar ul Qounain Jaffry** (VU), Analysis and Validation of Models for Trust Dynamics.

2011-26 | **Matthijs Aart Pontier** (VU), Virtual Agents for Human Communication.

2011-27 | **Aniel Bhulai** (VU), Dynamic website optimization through autonomous management of design patterns.

2011-28 | **Rianne Kaptein** (UVA), Effective Focused Retrieval by Exploiting Query Context and Document Structure.

2011-29 | **Faisal Kamiran** (TUE), Discrimination-aware Classification.

2011-30 | **Egon van den Broek** (UT), Affective Signal Processing (ASP): Unraveling the mystery of emotions.

2011-31 | **Ludo Waltman** (EUR), Computational and Game-Theoretic Approaches for Modeling Bounded Rationality.

2011-32 | **Nees Jan van Eck** (EUR), Methodological Advances in Bibliometric Mapping of Science.

2011-33 | **Tom van der Weide** (UU), Arguing to Motivate Decisions.

2011-34 | **Paolo Turrini** (UU), Strategic Reasoning

in Interdependence: Logical and Game-theoretical Investigations.