Definitions

Classification of General Games

Department of Mathematics, Linköping University

Mikael Böörs
Tobias Wängberg

The number your thesis gets from administrators

Credits: **16 hp**

Level: **G2**

Supervisor: **Tom Everitt, Marcus Hutter**,
Computer Science Department, Australian National University

Examiner: **Martin Singull**,
Department of Mathematics, Linköping University

Linköping: **May 2017**

Abstract

A summary of the thesis, presenting the important results.

Keywords:
Keyword 1, keyword 2, etc.

URL for electronic version:
`Theurltothethesis`

Sammanfattning

You can write a swedish abstract in this environment to get correct (swedish) formatting and hyphenation

Nyckelord:
Nyckelord 1, nyckelord 2, etc.

URL för elektronisk version:
`Theurltothethesis`

Acknowledgements

Acknowledgements and thanks to people that have helped you.

Nomenclature

The notation used in the thesis.

Symbol	**Explanation**
$k, l \in \mathbb{N}$	index for natural numbers
$i, j \in \mathbb{N}$	index for players
u_i	payoff function for player i
U	set of payoff functions
A_i	set of actions available to player i
a	action profile
s_i	strategy of player i
S_i	set of strategies of player i
s	strategy profile
S	set of strategy profiles
P_i	payoff matrix of player i
G	game
$\blacksquare$	end of proof
$\diamondsuit$	end of example
$\mathbb{R}, \mathbb{N}, ...$	set of real,natural numbers
e	natural $e \doteq 2.71828...$
P	probability
$\vec{v}$	vector
$\sim$	strategically equivalent
NE	Nash equilibrium
HO	Huertas optimal
PO	Pareto optimal

Contents

0.0.1 Definitions i

1 Background 1

1.1 Purpose 1

1.2 Game Theory 2

1.2.1 Introduction to Game Theory 2

1.2.2 Strategic Games and Equilibria 4

1.2.3 Zero-Sum Game 7

1.2.4 Mixed Strategies 9

1.2.5 Pareto Optimality 12

1.2.6 2×2 Games 13

1.2.7 Standard 2×2 Games 18

1.3 Classification of Games 24

1.3.1 Classification of 2×2 games by Rapoport, Guyer, and Gordon 25

1.3.2 A geometric classification system for $2{\times}2$ inverval-symmetric games, (Harris, 1969) 29

1.3.3 Best response classification of 2×2 strategic games . . . 32

1.3.4 A Geometric classification of symmetric 2×2 games . . . 41

1.3.5 A Topologically-Based Classification of the 2×2 Ordinal Games 48

1.3.6 Summary 52

2 Results 55

2.1 Classification by Decomposition 55

2.1.1 Decomposition of symmetric 2×2 games 56

2.1.2 Stereographic projection 59

2.2 Analysis of the Regions 62

2.2.1 Stronger Zero-Sum 65

2.2.2 Weaker Zero-Sum 71

2.2.3 Mixed Strength 77
2.2.4 Comparison Between the Regions 83
2.2.5 Conclusions 83
2.3 Nash and Huertas Optimality 83
2.4 Experimental Analysis 90
2.4.1 Experiment design 91
2.4.2 Result Analysis 95
2.5 Comparative Literature Analysis 95
2.5.1 Analysis of the Prisoner's Dilemma Regions 96
2.5.2 Analysis of the Stag Hunt Regions 99
2.5.3 Analysis of the Chicken Regions 100
2.5.4 Analysis of the Civic Duty Game Regions 101
2.5.5 Comparison with Reviewed Classifications 102
2.6 Interesting Properties 103
2.7 Number of Games 104
2.7.1 Non-distinct players 110
2.8 Number of Nash equilibria 113

3 Discussion **119**

A Detailed class description **121**

Chapter 1

Background

This chapter will contain the game theoretical background needed for this thesis as well as a literature review of different methods of classifying 2×2 games.

Purpose

The purpose of this thesis is to investigate different methods of classification of 2×2 games with the ultimate goal of partitioning the space of $m \times n$ matrix games into interestingly different classes. To help us achieve this goal and to restrict the purpose to a manageable size we commit to investigating the answer to the following questions.

1. Does the classification proposed by Huertas-Rosero divide the space of symmetric 2×2 games into interesting different classes? If not, develop a formal criteria for why.

2. Does Nash equilibria and Pareto optimality capture all interesting properties of a game?

3. Are all (NE,PO)-different games different in an interesting way?

4. Estimate the number of (NE,PO)-different $m \times n$ games.

Game Theory

Introduction to Game Theory

The goal of game theory is to model strategic interaction between rational agents. An agent is *rational* if he always takes the action that best maximises his utility. More formally a game consists of players and a set of rules. These rules can include actions, information sets, payoffs, outcomes and strategies. The necessary rules that need to be included in order to define a game is a set of players, a set of strategies and a set of payoff functions.

A goal in game theory is for the modeller to be able to predict what will happen before the game is played out. In order to do this an important concept in game theory is the concept of equilibria. A game is in an equilibrium then the game is stable if no player has an incentive to deviate from his strategy. When there is a unique equilibrium you will have a pretty good idea what will happen when the game is played out. There are many games however where there is not an unique equilibrium, which means that one or more players are indifferent about their equilibrium strategies. There might not even exist an equilibrium i.e. at least one player would benefit from changing his strategy regardless of situation. This is a big challenge when studying games in game theory.

There are is a big difference between games where the players make their move simultaneously and where they make them sequentially. A simultaneous game can be considered to be a game where players have no information about each other.

Here is a list of fundamental definitions that will be used frequently throughout the report.

Definition 1.2.1 (Set of actions)**.** The *set of actions* $A_i = \{a_1, ...a_m\}$ of player i is the set of possible choices of actions available to him at a given point of the game.

Definition 1.2.2 (Action profile)**.** An *action profile* $a = \{a_1, ..., a_n\}$ is a combination of actions of players 1 to n.

Definition 1.2.3 (Strategy)**.** The *strategy* s_i of player i tells him what actions to take in every conceivable situation of the game. For each player i the set of all possible strategies s_i is denoted S_i.

Definition 1.2.4 (Strategy profile)**.** A *strategy profile* $s = \{s_1, ..., s_n\}$ is a combination of strategies of players 1 to n.

Definition 1.2.5 (Payoff)**.** The *payoff* player i gets is defined by his payoff function $u_i : \{s_1, ..., s_n\} \mapsto \mathbb{R}$ that takes a given strategy profile and maps it to a real number.

Definition 1.2.6 (Outcome)**.** The *outcome* includes every aspect of the game that the modeller of the game finds interesting. Examples could be the players payoff and what actions the players took.

Strategic form games, or normal form games, consists of a set of strategy profiles and a set of payoff functions mapping strategy profiles to payoffs. In the case when the set of strategies is finite and countable the game can be represented by a strategic form matrix with strategy profiles connected to payoffs.

Definition 1.2.7 (Strategic game)**.** Let P be a set of players with $|P| = n$ and for all $i \in P$ let S_i be the non-empty set of strategies of player i. Define the set of strategy profiles as $S \triangleq \prod_{i \in P} S_i$. $\forall i \in P$ let $u_i : S \to \mathbb{R}$ be the pay-off function of player i and let $U \triangleq \{u_1, u_2, ..., u_n\}$.
The triple $G = (S, U, P)$ is called a *n-player strategic game* for the set of players P. (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)

Strategic games with 2 players are often represented with a pair of matrices. Each matrix represents the payoffs for each player. The position i, j in payoff matrix P_1 corresponds to the payoff Player 1 receives from strategy profile $\{i, j\}$, i.e. $u_1(i, j) = P_{1_{i,j}}$. The payoff matrices for Player 1 and Player 2 are shown in Figure 1.1.

$$P_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad\qquad P_2 = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

Figure 1.1: Payoff Matrices

These payoff matrices are combined into a payoff bi-matrix which represents the game. An example is shown in Figure 1.2 where the strategy profile $s = \{0, 1\}$ results in payoff $u_1(0, 1) = b$ for Player 1 and payoff $u_2(0, 1) = f$ for Player 2.

$$P = \begin{bmatrix} (a,e) & (b,f) \\ (c,g) & (d,h) \end{bmatrix}$$

Figure 1.2: Payoff Bi-Matrix

Strategic Games and Equilibria

There are several kinds of different equilibria. We start by defining a few of them, which finally leads to defining the important Nash Equilibrium.

We start with the following definition of a dominant strategy, which then leads to the definition of a dominant strategy equilibrium. In order to simplify further reasoning we introduce the following notation.

Definition 1.2.8 (Notation S_{-i}, s_{-i})**.** Given a game with set of strategies S we define S_{-i} as $S_{-i} \triangleq \prod_{j \in P \setminus \{i\}} S_j$, for a strategy profile $s \in S$, we define (s_{-i}, s_i^*) as the strategy profile $(s_1, ..., s_{i-1}, s_i^*, s_{i+1}, ..., s_n) \in S$ (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010).

Definition 1.2.9 (Dominant strategy)**.** A strategy s_i for player i is called *dominant* if

$$\forall s_i' : s_i' \neq s_i \, \forall s_{-i} \; u_i(s_{-i}, s_i) > u_i(s_{-i}, s_i')$$

.

Definition 1.2.10 (Dominant strategy equilibrium)**.** A game is said to have a *dominant strategy equilibrium* if there exists strategy profile $s \in S$ such that all strategies in s are dominant.

The dominant strategy equilibrium is unique in every game which follows immediately from its definition.

A dominant strategy never exist if a players choice of action depends on the other players choice of action. This follows since if the strategy did depend on the strategy of another player, then that would imply that there is another strategy which is better in some situation and therefore it is not dominant.

A classic example of a game where a dominant strategy equilibrium exists is the so-called Prisoner's Dilemma presented in Table 1.8.

Another equilibrium concept is the so-called *iterated dominance equilibrium.* To reach this equilibrium in a game the players iteratively rule out the bad strategies, the so-called weakly dominated strategies, to finally have one best strategy remaining.

Definition 1.2.11 (Weakly dominated strategy)**.** A strategy s_i is called *weakly dominated* if

$$\exists s_i' \forall s_{-i} \;\; u_i(s_{-i}, s_i') \geqslant u_i(s_{-i}, s_i)$$

and

$$\exists s_{-i}' \;\; u_i(s_{-i}', s_i') > u_i(s_{-i}', s_i).$$

In other words this means that a strategy s_i for player i is weakly dominated if there is another strategy for that player which is always better or equal in any possible strategy profile and is strictly better in at least one strategy profile.

Now we are ready to define the concept of *iterated dominance equilibrium.*

Definition 1.2.12 (Iterated dominance equilibrium)**.** An *iterated dominance equilibrium* is a strategy profile $s \in S$ obtained by iteratively ruling out weakly dominated strategies from each player until only one strategy remains for each player. The remaining strategy profile is then said to be an iterated dominance equilibrium.

Important to note is that this kind of equilibrium does not have to be unique. Different strategy profiles can be obtained by removing the weakly dominated strategies in different order.

Nash Equilibrium is the standard form of equilibrium, abbreviated NE. Intuitively if a strategy profile is a NE this means that if a game is in NE then no player has an incentive to change his strategy given that no other player changes his strategy. A mathematical definition is as follows.

Definition 1.2.13 (Nash equilibrium)**.** Given a game with a set of strategy profiles S and a set of payoff functions U a strategy profile $s \in S$ is said to be a (weak) *Nash equilibrium* if

$$\forall i \; u_i(s_{-i}, s_i) \geqslant u_i(s_{-i}, s_i') \forall s_i', s_i' \neq s_i.$$

If the inequality in definition 1.2.13 is strict, then the Nash equilibrium is said to be strong. To save some time and ink we will use the abbreviation NE when referring to Nash equilibria.

A natural question is if every game has a NE. This is not the case. Consider for example the game of rock, paper, scissor. Given a strategy profile in that game, one player will always have an incentive to change his action to increase his payoff. In order for one player to win this game, the other player has to lose, so a Nash Equilibrium cannot exist. **insert proper example here, maybe use this with mixed extension**

In contrast to a dominant strategy equilibrium, a strategy in a NE only needs to be the best strategy against the other players strategies in the NE strategy profile, not against all other possible strategies the other players can make. When a dominant strategy equilibrium occurs i.e. when all players have dominant strategies, the dominant strategy equilibrium is also a unique Nash equilibrium.

Both a dominant strategy equilibrium and an iterated dominance equilibrium are Nash equilibria.

A concept closely related to NE is the concept of best response correspondences. For each player the best response correspondence defines the set of strategies that maximise the utility given that the other player use a fixed set of strategies.

Definition 1.2.14 (Best response correspondence)**.** Let $G = (S, U, P)$ be a strategic game such that $\forall i \in P \, \exists n_i \in \mathbb{N} : S_i \subset \mathbb{R}^{n_i}$, $S_i \neq \emptyset$ and S_i is compact. The correspondence $B_i : S_{-i} \to S_i$ is called the *best reply correspondence* for player i and is defined as

$$B_i(s_{-i}) \triangleq \{s_i^* \in S_i : u_i(s_{-i}, s_i^*) \geq u_i(s_{-i}, s_i') \forall s_i' \in S_i\}$$

for any given $s_{-i} \in S_{-i}$ (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010).

One quickly realises that if a strategy profile $s^* \in S$ is such that $s_i^* \in B_i(s_{-i}^*)$ for every player $i \in P$ then s^* is a Nash equilibrium. **No source!**

Add text about intuition behind Nash theorem

Theorem 1.2.15 (Nash theorem)**.** [Source: GGF] Given a strategic game $G = (S, U, P)$ let $\mu_{-i} : S_i \to \mathbb{R}$ be the function defined by $\mu_{-i}(s_i^*) = u_i(s_{-i}, s_i^*)$. If $\forall i \in P \, \exists n_i \in \mathbb{N} : S_i \subset \mathbb{R}^{n_i}$, $S_i \neq \emptyset$ and S_i is compact and if furthermore

$\forall i \in P\, u_i$ is continuous and μ_{-i} is quasi-concave, then $\exists s \in S : s$ is a Nash equilibrium (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010).

Note: A function $f : X \to \mathbb{R}$ where X is a convex set such that $\exists n \in \mathbb{N} : X \subset \mathbb{R}^n$ is *quasi-concave* if $\forall y \in [0,1]$ and $\forall x', x'' \in X f(yx' + (1-y)x'') \geq min\{f(x'), f(x'')\}$.

Zero-Sum Game

A zero sum game is a game where in every possible strategy profile of the game, the respective payoffs for the players add up to zero. Intuitively this means that what one player gains, the other players lose. (Rasmusen, 1989)

A mathematical definition is given in definition 1.2.16.

Definition 1.2.16 (Zero-sum game)**.** A game $G = (S, U, P)$ is called a *zero-sum game* if

$$\forall i, j \in P \;\; \forall s \in S \;\; u_i(s) + u_j(s) = 0, \;\; i \neq j.$$

Definition 1.2.17 (Lower and upper value)**.** If $G = (S, U, P)$ is a two-player zero sum game the *lower value* of G is defined as $V^- \triangleq \sup_{s_1 \in S_1} \inf_{s_2 \in S_2} u_1(s_1, s_2)$ and the *upper value* of G is defined as $V^+ \triangleq \inf_{s_2 \in S_2} \min_{s_1 \in S_1} u_1(s_1, s_2)$.
If $V^- = V^+$ then we define the value V of G as $V \triangleq V^+$. (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)

An interpretation of V^- is that it is the lowest pay-off that player one will gain given that he plays rationally, and V^+ is the highest pay-off that player one can receive given that both players play rationally. Note that if we denote player one's pay-off with p_1 then player two's pay-off $p_2 = -p_1$ and hence $V^- \leq p_1 \leq V^+ \Leftrightarrow -V^+ \leq p_2 \leq -V^-$.
Given a two-player zero-sum game G with value V we can define optimal strategies for the players.

Definition 1.2.18 (Optimal strategy)**.** Given a game G with the properties stated above we define an *optimal strategy* for player $i \in P$ as $s_i \in S_i : \inf_{s_{-i} \in S_{-i}} u_i(s_{-i}, s_i) = V_i$ where $V_1 \triangleq V$ and $V_2 \triangleq -V$.(Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)

Note that since a two-player zero-sum game does not always have a value there are no optimal strategies in every such game. However, in the cases where optimal strategies for both players exists the game has some important properties.

Theorem 1.2.19. Suppose $G = (S, U, P)$ is a two-player zero-sum game.If G has a Nash equilibrium $s \in S$, G also has a value $V = u_1(s_1, s_2)$ where s_1 and s_2 are optimal strategies for player one and player two respectively. (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)

Proof. Suppose $s \in S$ is a NE. Then the definition of NE states that

$$\forall s_1^* \in S_1 \, u_1(s_1^*, s_2) \leq u_1(s_1, s_2) \tag{1.1}$$

$$\forall s_2^* \in S_2 \, u_2(s_1, s_2^*) \leq u_2(s_1, s_2) \tag{1.2}$$

Now, equation 1.2 $\Leftrightarrow \forall s_2^* \in S_2 u_1(s_1, s_2) \leq u_1(s_1, s_2^*)$ and together with equation 1.1 we get that $s \in S$ is NE $\Leftrightarrow \forall s_1^* \in S_1, \forall s_2^* \in S_2$

$$u_1(s_1^*, s_2) \leq_{1.1} u_1(s_1, s_2) \leq_{1.2} u_1(s_1, s_2^*)$$

$V^- = \sup_{s_1^* \in S_1} \inf_{s_2^* \in S_2} u_1(s_1^*, s_2^*) = \{1.1, 1.2\} = \sup_{s_1^* \in S_1} u_1(s_1^*, s_2) = \{1.1\} = u_1(s_1, s_2) \Leftrightarrow V^- = u_1(s_1, s_2)$.

$V^+ = \inf_{s_2^* \in S_2} \sup_{s_1^* \in S_1} u_1(s_1^*, s_2^*) = \{1.1, 1.2\} = \inf_{s_2^* \in S_2} u_1(s_1, s_2^*) = \{1.2\} = u_1(s_1, s_2) \Leftrightarrow V^+ = u_1(s_1, s_2) = V^-$.

This proves that G has a value $V = V^+ = u_1(s_1, s_2)$.

s_1 is an optimal strategy for player one if $\inf_{s_2^* \in S_2} u_1(s_1, s_2^*) = V$.
$\inf_{s_2^* \in S_2} u_1(s_1, s_2^*) = \{1.2\} = u_1(s_1, s_2) = V$

In the same way s_2 is an optimal strategy for player two since
$\inf_{s_1^* \in S_1} u_2(s_1^*, s_2) = \inf_{s_1^* \in S_1} -u_1(s_1^*, s_2) = -\sup_{s_1^* \in S_1} u_1(s_1^*, s_2) = \{1.1\} = -u_1(s_1, s_2) = -V$. (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)
□

Theorem 1.2.20. Given a two-player zero-sum game $G = (S, U, P)$ such that G has a value V and both player one and player two have optimal strategies $s_1 \in S_1$ and $s_2 \in S_2$ respectively, then the game G has a NE and $V = u_1(s_1, s_2)$. (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)

Proof. Suppose that the conditions in the theorem are fulfilled. Since s_1 is an optimal strategy for player one it is true that
$\inf_{s_2^* \in S_2} u_1(s_1, s_2^*) = u_1(s_1, s_2') = V$ for some $s_2' \in S_2$ and since s_2 is an optimal strategy for player two it is also true that
$\inf_{s_1^* \in S_1} u_2(s_1^*, s_2) = -\sup_{s_1^* \in S_1} u_1(s_1^*, s_2) = -u_1(s_1', s_2) = -V \Leftrightarrow u_1(s_1', s_2) = V$ for some $s_1' \in S_1$.

It is trivially true that $\forall s_1^* \in S_1$ and $\forall s_2^* \in S_2$
$u_1(s_1^*, s_2) \leq V \leq u_1(s_1, s_2^*)$ and that we can choose $s_1' = s_1$ and $s_2' = s_2$ so that $V = u_1(s_1, s_2)$. Now
$u_1(s_1^*, s_2) \leq u_1(s_1, s_2) \leq u_1(s_1, s_2^*)$.

This proves that s is a NE and that $V = u_1(s_1, s_2)$. (Gonzalez-Diaz, García-Jurado, and Fiestras-Janeiro, 2010)

□

Mixed Strategies

It is common for games with a discrete strategy set to lack NE. In the chapter about Nash equilibrium we began by defining the *dominant strategy equilibrium* and the *iterative dominance equilibrium*, but since many games lack dominant strategies and weakly dominated strategies the definition was extended to NE, which exists in a far wider range of games. A further extension that can be made is called the *mixed strategy* extension. If a equilibrium exists in the case of a discrete strategy set we call it a *pure strategy equilibrium.* A *pure strategy* is defined as a deterministic function s_i such that

$$s_i : \Omega_i \mapsto a_i.$$

where Ω_i is player i's information set.

It is clear that many finite games, that is, games whose set of strategy profiles has finite cardinality, do not fulfil the conditions of Nash theorem. There are however methods to extend finite strategic games in a natural way to guarantee the existence of Nash equilibrium for every such game. The *mixed extension* of a game is on such method. To define the mixed extension of a game we first need to introduce some basic definition and concepts and we will start with a formal definition of a finite strategic game.

Definition 1.2.21 (Finite strategic game)**.** Given a game strategic game $G = (S, U, P)$, we say that G is *finite* if the cardinality of the set of strategy profile is finite, i.e. if $|S| < \infty$.

Obviously, finite games do not have convex sets of strategies and hence they are not included in Nash theorem.

Notation 1.2.22 (Σ)**.** Let S_i be a set of strategies for a player $i \in P$ in a finite game $G = (S, U, P)$. We denote the set of probability distributions over S_i by Σ_i. By probability distribution over S_i we mean a distribution p such that $p : X \to [0, 1]$ and $\sum_{x \in X} p(x) = 1$.
Furthermore we denote the Cartesian product over all Σ_i, where $i \in P$, by Σ.
We call an element $\sigma_i \in \Sigma_i$ *mixed strategy* and en element $\sigma \in \Sigma$ is called a *mixed strategy profile*. For the sake of convenience, we will henceforth call a strategy a *pure strategy* and a strategy profile a *pure strategy profile* whenever mixed strategies or mixed strategy profiles are involved.

Definition 1.2.23 (Mixed extension)**.** Given a finite game $G = (S, U, P)$ we define the *mixed extension* of G, denoted $\Gamma(G)$, as the strategic game $\Gamma(G) = (\Sigma, U, P)$. Σ is the set of mixed strategy profiles of G and for any pure strategy profile $s \in S$, $\sigma(s) \triangleq \prod_{i \in P} \sigma_i(s_i)$.
U is the set of pay-off functions u_i for every player i, where $\forall \sigma \in \Sigma$, $u_i(\sigma)$ is defined as $\sum_{s \in S} u_i(s) * \sigma(s)$.

This means that a player in some sense is indifferent to his actions but instead lets a stochastic device decide his action given his current information where the strategy is to choose the right device given his information set. If $f > 0$ in the definition above the game is called *completely mixed*. So if the game is completely mixed any action has probability of occurring and therefore a complete mix of all actions can be made. The set of mixed strategies becomes continuous assigning probabilities $p \in [0, 1]$ to actions.

An intuitive argument against the mixed strategy extension is that it feels unrealistic that a player would let chance decide the outcome of the game. Another way to look at it is to say that the players actions are only stochastic from an observers point of view, and not from their subjective point of view. So for example given a game with mixed strategies you could view it like you have a set of players P_i representing each player p_i in the game where all the players in P_i only have one action in the action set of p_i available to him. Then when replaying the game for all players in P_i the fraction of each action will correspond to the mixed strategy probability of these actions.

The main idea with defining the mixed extension of a finite game is that the extension will fulfil the conditions of Nash theorem, and hence the following theorem.

Theorem 1.2.24 (NE existence in $\Gamma(G)$)**.** The mixed extension of any finite strategic game has a Nash equilibrium.

The idea behind the proof of this theorem is that the mixed extension of a finite strategic game is constructed in such a way that for every player i Σ_i is convex, and that the rest of the conditions of Nash theorem are fulfilled as well. Hence Nash theorem guarantees the existence of at least one Nash equilibrium.

Example

Another topic is how to find the NE in a mixed strategy extension of a game. This is done by first constructing the *expected payoff functions*. Then an intuitive way of finding the equilibrium probabilities is equating the payoff functions for every player and solving for the probabilities. This makes sense because if the payoff of all strategies are the same the player has no incentive to change it, i.e. equilibrium. Another way is via taking the derivative of the payoff function with respect to each action probability solving for the other players probability of actions maximizing the current players payoff. This might not make sense at first glance but if you choose probabilities like this the current player will have no incentive to change strategy. So doing this for all players will lead to an Nash equilibrium in the game. Since the payoff function is always linear both methods will always give the same result.

A way to disprove the existence of mixed strategy equilibrium is to solve for probabilities using the methods above and showing that they don't belong to the interval $[0, 1]$.

If the payoff function is continuous then this means that a small alteration in strategy will also result in a small change in payoff. So when the discontinuous payoff function might 'miss' the optimal strategies, the continuous one is able to come much closer to the equilibrium and therefore likelihood of existence of NE is much larger with the mixed strategy extension. If the strategy set is compact the likelihood of NE also increases. For example if the set is open and the optimal strategy is at the boundary of the set, then NE cannot exist since the player will want to get arbitrarily close to that boundary. If the set is unbounded, then the player might want to go towards infinity if for example his action set are the natural numbers and the higher number he picks the higher payoff he gets. (Rasmusen, 1989)

If all players in a game are able to play with strategies mapping to the same distribution, then those strategies are called *correlated strategies.*

Pareto Optimality

In prevoius sections of this chapter we have talked a lot about different kinds of equilibrium, which in a sense means there is a stability in the game if equilibrium exists. But this says nothing about how desirable this outcome is for the players. In the example of the Prisoner's Dilemma, it would be much better for *both* players if they could agree on choosing Cooperate instead of Defect even though $\{\text{Defect}, \text{Defect}\}$ is a Nash equilibrium. We say that $\{\text{Defect}, \text{Defect}\}$ is *Pareto dominated* by $\{\text{Cooperate}, \text{Cooperate}\}$, see definition 1.2.25 and definition 1.2.26 (Rasmusen, 1989).

Definition 1.2.25 (Strong Pareto-dominance)**.** A strategy profile $s^\star$ *strongly Pareto-dominates* $s \neq s^\star$ if

$$\forall i \;\; u_i(s^\star) > u_i(s).$$

Definition 1.2.26 (Weak Pareto-dominance)**.** A strategy profile $s^\star$ *weakly pareto dominates* $s \neq s^\star$ if

$$\exists i \;\; u_i(s^\star) > u_i(s)$$

and

$$\nexists j \;\; u_j(s^\star) < u_j(s).$$

An example where there does not exist a pareto dominating strategy profile is a zero-sum game, because in order to make a player better off you have to make another worse off.

In games where players are allowed to discuss with one another they might want to agree on a strategy profile equilibria that is better for all the players, for example a pareto dominant one.

A concept commonly used in game theory is the concept of Pareto optimality. Intuitively an outcome is Pareto optimal if there is no way to change the allocation of resources (payoffs) to make all players simultaneously obtain more resources. The NE concept is used to predict strategic play, whereas Pareto optimality is used to measure the efficiency of an outcome.

Definition 1.2.27 (Pareto optimality)**.** A strategy profile $s^\star \in S$ is *Pareto optimal* if for all $i \in P$

$$u_i(s') \geq u_i(s^\star) \Rightarrow \exists j \in P: \; u_j(s') < u_j(s^\star), \; s' \neq s^\star.$$

Pareto optimality will sometimes be abbreviated with PO.

Pareto optimality means that it is not possible to improve any players payoff in strategy profile $s^\star$ by changing strategy profile without decreasing another players payoff. A good example, presented in Table 1.8 of a game where the equilibrium does not coincide with Pareto optimality is the *Prisoner's Dilemma.* The equilibrium outcome is $\{\text{Defect}, \text{Defect}\}$ but the Pareto optimal one is $\{\text{Cooperate}, \text{Cooperate}\}$, but in this case it is not an equilibrium. In a zero-sum game all strategy profiles are Pareto optimal since in order to increase a player's payoff another player's payoff must decrease. (Rasmusen, 1989)

In a sequential game with sequential rationality *pareto perfectness* means that given that the players have reached an agreement on how the game should be played out no player will want to re-negotiate in any future subgame. So the agreement must hold no matter what even if the players are allowed to re-discuss and deviate from it. So a strategy profile can be Pareto optimal but not Pareto perfect since the players might want to re-negotiate if they reached some other subgame. (Rasmusen, 1989)

2×2 Games

A 2×2 game is a game with two players with two strategies each. For example, a $2 \times 3 \times 4$ game is a game with three players with 2, 3 and 4 actions respectively. Since 2×2 games are the smallest interesting kind of game in game theory they are suitable for thorough study. (Rapoport, Guyer, and Gordon, 1978) An example is shown in Table 1.1.

		Player 2	
		0	1
Player 1	0	(a, x)	(b, y)
	1	(c, z)	(d, w)

Table 1.1: General 2×2 game

In this thesis we will for the sake of consistency denote the player choosing row as Player 1 and the player choosing column as Player 2 except in some cases when other names for the players are more suitable for the presentation.

An important subclass of the 2×2 games are the *symmetric* 2×2 games, defined in definition 1.2.28.

Definition 1.2.28 (Symmetric game)**.** A 2×2 game with payoff matrix P_1 for Player 1 and payoff matrix P_2 for Player 2 is called *symmetric* if $P_1 = P_2^t$.

A symmetric 2×2 game has the convenient property that the game looks exactly the same no matter what players perspective one chooses to look at it from. This means that the payoff matrix of one player equals the transpose of the opposing players payoff matrix. An example is shown in Table 1.2.

		Player 2	
		0	1
Player 1	0	(a, a)	(b, c)
	1	(c, b)	(d, d)

Table 1.2: Symmetric 2×2 game

It is common to restrict to ordinal payoffs, not taking into account the actual numerical values but only the ordinal structure between them and not counting the games where there are ties between one players payoffs. Using an ordinal scale means only ranking the payoffs from smallest to largest, not taking into account the ratio between them. In this thesis the payoffs will be named 1, 2, 3 and 4 when using an ordinal scale, where 4 is the highest ranked payoff and 1 the lowest. Properties like the Nash equilibrium and Pareto optimality only depends on the ordinal structure so this is a reasonable simplification to make.

		Player 2	
		0	1
Player 1	0	$(402, 402)$	$(-35, 88)$
	1	$(88, -35)$	$(33, 33)$

Table 1.3: Game with Numerical Values

		Player 2	
		0	1
Player 1	0	$(4, 4)$	$(1, 3)$
	1	$(3, 1)$	$(2, 2)$

Table 1.4: Game with Ordinal Scale

The game in Table 1.3 uses numerical values, whereas the game in Table 1.4 uses an ordinal scale. Note that these games have the same Nash equilibrium and Pareto optimal outcomes, so they can be regarded as the same game. This is of course a big simplification and some information about the game is lost.

When restricting to an ordinal scale however, the number of games goes from an infinite continuum of games to a finite number, which is a big advantage when studying 2×2 games. This means that it is now possible to exhaustively study these games, which was impossible before.

Another simplification that can be made is to regard games where the names of one or both players actions have been swapped, as the same game. This corresponds to row swapping and column swapping in the game matrix. The players can also be regarded as identical. This means swapping the players will not result in a new game, but a different representation of it.

For example, the game in Table 1.5 and the game in Table 1.6 is considered to be the same since the only difference is that the actions of Player 1 has been relabeled. We say that these games are *strategically equivalent*, see Definition 1.2.29. In the same way the game in Table 1.6 and the game in Table 1.7 are strategically equivalent because the only difference is that the positions of the players have been swapped.

		Player 2: 0	Player 2: 1
Player 1	0	$(4, 2)$	$(2, 3)$
	1	$(1, 2)$	$(3, 4)$

Table 1.5: Game 1

		Player 2: 0	Player 2: 1
Player 1	0	$(1, 2)$	$(3, 4)$
	1	$(4, 2)$	$(2, 3)$

Table 1.6: Game 2

		Player 2: 0	Player 2: 1
Player 1	0	$(2, 1)$	$(2, 4)$
	1	$(4, 3)$	$(3, 2)$

Table 1.7: Game 3

Given these simplifying assumptions, the total amount of strategically non-equivalent games can be calculated using combinatorial methods.

When using the ordinal scale, there are 4 payoffs for each player and 4 places to put them, so this means that there are $4! \times 4! = 576$ different 2×2 games with ordinal scale. If we restrict to the symmetric games, one players payoffs is given by transposing the other players payoff matrix, see Definition 1.2.28. Therefore there is a total of $4! = 24$ symmetric 2×2 games. By using the assumptions of strategic equivalence, see Definition 1.2.29, each symmetric game can be represented by 2 different matrices, by simultaneously relabeling both players actions. This means that the total number of strategically non-equivalent symmetric games are $\frac{24}{2} = 12$. Each symmetric game can however be represented by 4 different matrices because of the assumption of relabeling actions. But for the game to remain symmetric *both* the rows and the columns need to be permuted simultaneously and therefore a symmetric game can only be represented by 2 symmetric matrices. Since the game looks the same from both players perspective, swapping positions of the players will not result in a new matrix. An asymmetric 2×2 game can on the other hand be represented in 8 different ways. First 2 representations are obtained by swapping the players. For each of those games the actions can be relabeled in the same way as was done with the symmetric games. Therefore the total number of representations of asymmetric games are $2 \times 4 = 8$. Since we know the total number of 2×2 games are 576, given the assumptions, we can solve equation 2.7.1 for x, where x is the total number of asymmetric games. (Rapoport, Guyer, and Gordon, 1978)

$$576 = 12 \times 4 + x \times 8 \Leftrightarrow x = 66. \tag{1.3}$$

The result is that there are $66+12 = 78$ strategically non-equivalent 2×2 games, which is a very small number compared to infinity. This way of calculating 2×2 games is used by Rapoport, Guyer, and Gordon, (1978) with the same result.

Definition 1.2.29 (Strategically equivalent)**.** Games G and G' are said to be *strategically equivalent*, denoted by $G \sim G'$ if G' can be obtained by in any way permuting the rows or columns of G or transposing both players payoff matrices in G.

It follows that from this choice definition of strategically equivalence that strategic equivalence is an equivalence relation. To realise this suppose that $G \sim G'$ and $G' \sim G''$. Of course, $G \sim G$ since since the identity permutation can be applied to G so the relation is reflexive. $G' \sim G$ since the inverse to the permutation taking G to G' can always be applied and $G^T = G' \Leftrightarrow G'^T = (G^T)^T = G$, so therefore the relation is symmetric. If σ_1 and σ_2 are some combinations of permutations and transposes such that $\sigma_1(G) = G'$ and $\sigma_2(G') = G''$ then $\sigma_2(\sigma_1(G)) = \sigma_2(G') = G''$ and hence $\sim$ is transitive.

Definition 1.2.30 (Common interest game)**.** A 2×2 game with payoff matrix P_1 for Player 1 and P_2 for Player 2 is called a game of common interest if

$$P_1 = P_2.$$

An interesting yet simple property of games is that games can always be decomposed in a common interest game and a zero-sum game. This idea together with more complicated forms of decompositions was proposed and studied in more detail by Hwang and Rey-Bellet, (2016). The mathematical theory behind their various decompositions will not be discussed in detail in this thesis, the focus will be on the simple decomposition presented in Theorem 1.2.31 which is inspired by the work of Hwang and Rey-Bellet, (2016).

Theorem 1.2.31 (Decomposition)**.** A 2×2 game with payoff matrix P can always be decomposed into the sum of a zero-sum game with payoff matrix Z and a game of common interest with payoff matrix C such that

$$P = C + Z.$$

Proof. Let G be an arbitrary 2×2 game as in Table 1.1. G can be decomposed as follows:

$$P = \begin{bmatrix} (a,x) & (b,y) \\ (c,z) & (d,w) \end{bmatrix} = \begin{bmatrix} (\frac{a+x}{2}, \frac{a+x}{2}) & (\frac{b+y}{2}, \frac{b+y}{2}) \\ (\frac{c+z}{2}, \frac{c+z}{2}) & (\frac{d+w}{2}, \frac{d+w}{2}) \end{bmatrix} + \begin{bmatrix} (\frac{a-x}{2}, \frac{x-a}{2}) & (\frac{b-y}{2}, \frac{y-b}{2}) \\ (\frac{c-z}{2}, \frac{z-c}{2}) & (\frac{d-w}{2}, \frac{w-d}{2}) \end{bmatrix} = C + Z.$$

Where C is a common interest game according to Definition 1.2.29 and Z is a zero-sum game according to Definition 1.2.30.

□

The result in Theorem 1.2.31 shows that any game can always be divided into a common interest part and a conflict part. Depending on how strong the common interest is in relation to the conflict and how the conflict works with or against the common interest, different games are acquired.

Standard 2×2 Games

In this subsection some of the 2×2 that is usually refered to as the standard games, because the are so well studied, will be presented with an attempt to capture the essential properties of these games.

One of the perhaps most known games in game theory is the Prisoner's Dilemma and has been studied intensely by authors as Axelrod and Hamilton, (1981). The game is between two prisoners who have been arrested for a crime and are now being interrogated and must choose between cooperation and defection. If a prisoner chooses to cooperate, he remains silent. If he chooses to defect he will try to blame the other prisoner for the crime. In this game mutual cooperation is rewarded and mutual defection is punished. If both prisoners remain silent, they will both only receive a short prison sentence. If they defect and try to blame each other they will get a long prison sentence. The dilemma arises because of the possibility to exploit a cooperating player. If a prisoner chooses to defect while the other cooperates, the cooperating player will receive a lifetime in prison, often called the Sucker's Payoff, while the other prisoner is set free and receives his highest payoff. (Axelrod and Hamilton, 1981)

An example of the Prisoner's Dilemma in strategic form is shown in Table 1.8.

		Prisoner 2	
		Cooperate	Defect
Prisoner 1	Cooperate	$(3, 3)$	$(1, 4)$
	Defect	$(4, 1)$	$(2, 2)$

Table 1.8: Prisoners Dilemma

Interesting to note is that if both players stopped thinking of himself, departing from the axioms of rationality in game theory, and instead tried to maximize the *other* prisoner's payoff the more fortunate outcome would be that both players choose Cooperate which is also the only Huertas optimal outcome.

What can be said about the properties of prisoners dilemma? The most apparent properties are that both players have dominant strategies, but the unique Nash equilibrium is Pareto dominated. It seems that it is the fact that the NE is Pareto dominated that makes the Prisoner's Dilemma so interesting. The Prisoner's Dilemmas' cousin Deadlock (Gomes-Casseres, 1996) is the same game except for the fact that the NE is not Pareto dominated. It has not gotten nearly as much attention as Prisoner's Dilemma and Deadlock is by many considered an uninteresting game and it looks almost exactly the same as the Prisoner's Dilemma at first glance.

		Prisoner 2	
		Cooperate	Defect
Prisoner 1	Cooperate	$(3, 3)$	$(4, 1)$
	Defect	$(1, 4)$	$(2, 2)$

Table 1.9: Deadlock

Note that the game in Table 1.9 is obtained by transposing the prisoners payoff matrices in Table 1.8, i.e. Prisoner 1 play the game for Prisoner 2 and vice versa.

Another branch of games that seems to have caught the interest of many are the coordination games. These are games that, unlike the Prisoner's Dilemma and Deadlock have two Nash equilibria on the outcomes when the players coordinate on the same action.

An example of a simple form of a coordination game is a game between two groups of people living close to each other that have just started using automobiles. They have noticed complications however when the two groups are unable to coordinate on using the same rules. It is not important what rules they use, as long as they coordinate. This can be modelled as a game between two players, each representing one of the groups, that choose between two set of rules as shown in Table 1.10. This game is called the Coordination Game by Rasmusen, (1989).

		Player 2	
		Rule 1	Rule 2
Player 1	Rule 1	$(1, 1)$	$(-1, -1)$
	Rule 2	$(-1, -1)$	$(1, 1)$

Table 1.10: Coordination Game

Even though the game in Table 1.10 is an extremely simple game, it is still of some interest and appear in many real life situations.

A slightly more complicated game, similar to the one in Table 1.10, is the following.

		Player 2	
		Rule 1	Rule 2
Player 1	Rule 1	$(2, 1)$	$(-1, -1)$
	Rule 2	$(-1, -1)$	$(1, 2)$

Table 1.11: The Battle of the Sexes

This is almost the exact same game as the one in Table 1.10 but in the game in Table 1.11 Player 1 has a preference for Rule 1, since he already knows Rule 1 and it is difficult to learn a new rule. In the same way Player 2 has a preference for Rule 2. This game has received a lot of attention and is usually called *The Battle of The Sexes* (Rasmusen, 1989). The main difference between this game and the game in Table 1.10 is that there is conflict. Player 1 prefers that they coordinate on Rule 1 but Player 2 prefers Rule 2. Note that the one thing that seems to make this game more interesting is conflict, in the same way as Pareto dominance made Prisoner's Dilemma more interesting than Deadlock. In terms of Pareto optimality, Pareto dominance and Nash equilibria, these games are the same.

A third kind of coordination game is a game where both coordinating strategy profiles are NE, but one Pareto dominates the other. It is similar to Coordination Game, Table 1.10, except for the Pareto dominated NE. Imagine a game between a hunter and a tracker. The hunter is good at killing the animal once he found it but does not know how to track it down. The tracker can easily track animals down, but lacks the skill of killing them. They have to decide between hunting Stag or hunting Hare. If they manage to hunt down a Stag, it will give them higher reward than only hunting down a small Hare. So if they manage to coordinate on hunting Stag, both players will receive their highest possible payoff. A model of this situation is shown in Table 1.12. This game is called Ranked Coordination by Rasmusen, (1989).

		Tracker	
		Stag	Hare
Hunter	Stag	$(3,3)$	$(0,0)$
	Hare	$(0,0)$	$(1,1)$

Table 1.12: Ranked Coordination

What makes this game uninteresting is that both players can safely choose Stag because it gives them the highest payoff. So when playing this game it is safe to assume that the opposing player will also choose to hunt stag and choose the same yourself.

The game showed in Table 1.13 below adds a little more flavour to the game in Table 1.12. Picture the same situation as the one between the hunter and the tracker, but this time change the players to two hunters who has some limited experience of both hunting and tracking. In this game the highest possible payoff for both players is still from coordination on hunting Stag, but in this game both hunters are skilled enough to hunt Hare on their own. Because they have to share the profit when hunting together, it is more profitable to hunt Hare alone than coordinating on that action. This is another game that has caught a lot of interest and is commonly referred to as the *Stag Hunt* (Skyrms, 2004).

		Hunter 2	
		Stag	Hare
Hunter 1	Stag	$(3,3)$	$(0,2)$
	Hare	$(2,0)$	$(1,1)$

Table 1.13: Stag Hunt

Ranked Coordination and Stag Hunt are they same in that they share the same NE outcomes and of NE outcome Pareto dominates the other. The main difference is that the strategy profile $\{\text{Hare}, \text{Hare}\}$ is Pareto optimal in Ranked Coordination but not in Stag Hunt. In Stag Hunt, it is no longer risk free to hunt Stag, because the player can potentially gain nothing depending on the other players choice. So if you are unsure about the opposing players type, it might be a better choice to choose to hunt hare since you will at least never end up empty handed.

Note that all of the coordination games described above is no-conflict games, except for The Battle of the Sexes, so apparently conflict is not the only interesting aspect. The players in Stag Hunt do however receive opposite rewards when they fail to coordinate, so there is conflict in some sense, but it does not seem to be essential to these games.

Another kind of game are the *constant sum games* (Rapoport, Guyer, and Gordon, 1978). A constant sum game is a game where the sum of the payoffs in each cell of the payoff bimatrix has a constant sum. An example that most people have played themselves is the 3×3 game rock, paper, scissor. In this game the players interests are completely opposite. All outcomes are Pareto optimal, because it is impossible to make on player better of without hurting the other, but no outcome is a Nash equilibrium. A simpler version of rock, paper, scissor is the 2×2 game *Matching Pennies* shown in Table 1.14 (Robert Gibbons, 1992). In this game two players must choose between the numbers 0 or 1. If the sum of the numbers is even, then Player Even wins. If the sum is odd, then Player Odd wins.

		Player Odd 0	Player Odd 1
Player Even	0	$(1, -1)$	$(-1, 1)$
	1	$(-1, 1)$	$(1, -1)$

Table 1.14: Matching Pennies

What makes this game interesting could be the extremeness of it. The players interests could not be more mismatched, it is a game of *complete opposition* (Rapoport, Guyer, and Gordon, 1978).

The game of *Chicken* (Rasmusen, 1989), showed in Table 1.15, is also one of the more famous 2×2 games. A background story to this game could be as follows. Consider a classroom with pupils and a teacher. Two of the pupils are competing who whisper swear words loudest without getting caught. If both continue this competition however, the teacher will notice them and they will both get detention, which is the worst case scenario. If they both agree to stop, nothing will happen. If, on the other hand, one of the pupils decide to give up but the other continues he will be seen as a coward while the other gets all the glory for having the courage to defy the teacher. Likewise, if one of the pupils believe that the other will not give up, then he will choose to give up because he is more afraid of the teacher than of humiliation.

		Pupil 2 Continue	Pupil 2 Give Up
Pupil 1	Continue	$(-2, -2)$	$(1, -1)$
	Give Up	$(-1, 1)$	$(0, 0)$

Table 1.15: Chicken

This game is a symmetric 2×2 game with two Nash equilibria, similar to the coordination game. The big difference is that the equilibria are in the "uncoordinated" outcomes and none of them are Pareto optimal. There is however one Pareto optimal outcome in the symmetric strategy profile $\{\text{Give Up}, \text{Give Up}\}$. One could argue that the Pareto optimal outcome would be the most fair one and therefore players should aim for this outcome. But given that the players are rational, if one player suspects the other would give up he will decide to continue. If both players think in this way, it will lead to the worst possible outcome. Like the Prisoner's Dilemma, this game is also tragic in some sense, since conflict is the best option in this game and one player will receive high utility at the expense of the other players utility.

Classification of Games

In this chapter we will present a literature review over some of the more well-known approaches to classifying 2×2 games. The presentation of the different classifications are done with respect to similarities in the approaches of classifying games. The theory of each classification method is presented and followed by a discussion of differences, similarities, advantages and disadvantages. From this review we conclude that some of these classification methods lack motivation of why their classification conditions are interesting and some lack mathematical structure and sophistication. We believe that these are fundamental conditions that need to be fulfilled for a classification to be useful.

The earliest and perhaps most well known classification reviewed in this chapter is the classification done by Rapoport, Guyer, and Gordon, (1978). They base their classification on properties such as how aligned the interests of the players are, different types of equilibria, the number of dominant strategies and combinations of different kinds of pressures on the players. This is followed by the geometric classification done by Harris, (1969) who investigates a method of classifying symmetric games based on a two-parameter parameterisation of the interval scale payoffs, defining regions in the plane for different kinds of games. Harris classification includes games with ties, unlike Rapoport, Guyer, and Gordon who use a strict ordinal scale of payoffs.

Following the presentation of the classification method proposed by Huertas-Rosero, (2003), a method of classification introduced by Borm and Du, (1987) is presented. This method is based upon different types of combinations of best response correspondences of the two players in a mixed strategy game. This method of classification differ from the two methods introduced by Rapoport, Guyer, and Gordon and Harris in that it classifies mixed strategy games as well as pure strategy games. Borm's method is strictly analytic and unlike the two previous methods, as it does not take human irrationalities into account.

A method similar to the one done by Harris, (1969) is the one introduced by Huertas-Rosero, (2003). They are similar since both classifies games geometrically and both restrict themselves to symmetric games, However Huertas-Rosero chooses a different parametrisation and projects the space of games onto the three-dimensional unit sphere. The sphere is divided into regions depending on inequalities defined by Huertas optimality and NE and the regions provides the basis for this classification system.

The final classification method presented is a topological classificaiton done by Robinson and Goforth, (2003). Based on a swapping operation swapping adjacent payoffs a topology is induced on the space 2×2 ordinal games. The result is a 144 game topological map, with different regions of the map representing games with different kinds of properties.

Classification of 2×2 games by Rapoport, Guyer, and Gordon

Rapoport, Guyer, and Gordon, (1978) use a strictly ordinal scale when defining the payoffs. The strictly ordinal scale means that they do not allow ties, i.e. none of the payoffs of a given player are allowed to be equal. With this notation for the payoffs, it is easy to calculate the number of 2×2 games as shown in the game theory subsection. Harris, (1969) extends this classification to an interval scale, where he classifies games geometrically by creating different regions in the plane for each symmetric 2×2 game using a certain set of parameters.

Rapoport and Guyer classifies all 2×2 games, not just the symmetric, which gives a more general classification than for example Huertas-Rosero, (2003). The problem with this typological approach is that is hard to generalize since they classify by a lot of different conditions which makes it difficult to get a complete intuitive sense of how the classification works and how they are actually related mathematically.

Rapoport, Guyer, and Gordon uses a system often used in biology for classifying. They divided the games first in *phyla*. Then he divides each phyla into classes, orders, genera and finally species. The species are divided in lexicographic order depending on *stability*.

The phyla consists of three different branches, games of *complete opposition*, *partial conflict* and *no conflict*.

A game of no conflict is described as a game where both they players interests are completely aligned on a single outcome, i.e. both players highest ranked payoffs is in the same cell of the matrix.

A game of complete opposition is a game where the players interests are completely opposite in each outcome. A good example of this is a *constant sum* *game*, which is a game where the sum of both players outcomes is constant in each cell of the payoff matrix. Observe for example the following constant sum game, which is also a game of complete opposition:

		Player 2: 0	Player 2: 1
Player 1	0	$(7, -5)$	$(3, -1)$
	1	$(-4, 6)$	$(12, -10)$

If we replace the values of the payoffs with their rank, it becomes more clear:

		Player 2: 0	Player 2: 1
Player 1	0	$(3, 2)$	$(2, 3)$
	1	$(1, 4)$	$(4, 1)$

Finally a game of partial conflict is a game where the players interests are aligned on some outcomes, but differ on other. To illustrate this consider the following game with ranked payoffs:

		Player 2: 0	Player 2: 1
Player 1	0	$(4, 3)$	$(2, 1)$
	1	$(1, 2)$	$(3, 4)$

Player 1 and Player 2 clearly both prefers the coordinated strategy outcomes on the diagonal over the non-coordinated ones, but Player 1 prefers outcome 00 over outcome 11 and vice versa. So in this respect it is a game of mixed conflict.

These phyla are the divided by if the *natural outcome* is NE or not, which makes up the classes. The class in which the natural outcome is NE is further divided into subclasses if it is also PO or not.

A natural outcome is intuitively described as an outcome which occurs with highest frequency if the game is replayed in an experiment (Rapoport, Guyer, and Gordon, 1978). The natural outcome has two separate definitions depending on the existence of dominant strategies.

Definition 1.3.1 (Natural outcome)**.** If at least one dominant strategy exist, then the *natural outcome* is defined as the unique NE. If no dominant strategy exist, then the *natural outcome* is defined as the intersection between both players *maximin* strategies.

The orders of the game are defined by if the game has 0, 1 or 2 dominant strategies. If the game has 1 or 2 dominant strategies then the game always has one unique NE, an example is the Prisoner's Dilemma where the number of dominant strategies is 2.

Lastly the genera is defined by if there exists different kinds of *pressures* in the game acting on the players given that the current strategy profile is NE. The concept of pressure could be interesting in cases of iterated play, where the players might actually be tempted (or forced) to shift from his equilibrium strategy. The pressures that are defined are *competitive*, *force* and *threat* pressures.

A game with competitive pressure is described as a game where a player values an outcome where he gets higher relative payoff compared to the opposing player, even if the payoff he gets in that outcome is actually smaller than in the equilibrium one.

The concept of forced pressure is best explained by an example. Rapoport, Guyer, and Gordon, (1978) gave the example in Table 1.16.

		Player 2	
		0	1
Player 1	0	$(2,4)$	$(4,1)$
	1	$(1,2)$	$(3,3)$

Table 1.16: Forced Pressure

In this game strategy profile $\{0,0\}$ is both a unique NE and Natural outcome, but Player 1 has reasons not to be satisfied with that outcome since he gets his second worst possible outcome. Consider now the case where he, in iterated play, chooses to change his strategy so that the strategy profile becomes $\{1,0\}$ instead. Then this results in Player 2 being forced to change his strategy so that the game ends up in $\{1,1\}$ which is gives Player 1 his second best payoff.

Rapoport, Guyer, and Gordon, (1978) gave the example in Table 1.17 to introduce the concept of threat pressure.

		Player 2	
		0	1
Player 1	0	$(2,4)$	$(4,3)$
	1	$(1,2)$	$(3,1)$

Table 1.17: Threat Pressure

Like in the example of the game with forced pressure the strategy profile $\{0,0\}$ is both NE and NO. Here, on the other hand, the sought after outcome for Player 1 is in $\{0,1\}$ which is only obtained by him not doing anything and Player 2 for some reason changing his strategy for a lower payoff. In iterated play you could however imagine that Player 1 in iterated play sometimes changes to action 1 giving both players a worse payoff. This induces a threat on Player 2 to instead change his strategy to 1 and giving Player 1 what he desires but still not losing as much as if he would do nothing while Player 1 changes his strategy.

Combinations of these kinds of pressures gives 8 different genera in total:

1. No pressures.
2. Competitive pressure only.
3. Force pressure only.
4. Threat pressure only.
5. Threat and Force pressure.
6. Threat and Competitive pressure.
7. Force and Competitive pressure.
8. Threat and Force and Competitive pressure.

The resulting genera is then divided into different species by if the games in the particular is strongly stable, stable, weakly stable or unstable.

A *strongly stable* game is a game of no-conflict with no pressures. A *stable* game is any game with no pressures. A *weakly stable* game is any game with a single pressure. An *unstable game* is a any game with more than 2 pressures.

The classification is represented by one graph for each class in each phyla, resulting in a total of 24 end nodes representing different kinds of lexicographically ordered species.

The classification by (Rapoport, Guyer, and Gordon, 1978) is typological in the sense that they classify games purely based on properties that they find interesting, instead of using purely mathematical conditions. This is a more experimental approach to classifying games in contrast to later approaches (Huertas-Rosero, 2003), (Robinson and Goforth, 2003), (Borm and Du, 1987). Rapoport, Guyer, and Gordon include aspects such as stability, meaning how frequent a certain outcome is when the game is played iteratively. One kind of iterated play could be to let two players play the game repeatedly and observe the frequencies of outcomes. Another could be to pick random players to play the game for the first time and observe how they play and what outcomes are more common than others. A lot of effort were put in by the authors to experimentally investigate the different kinds of games, leading up to their choice of classification.

This classification is interesting when considering real situations involving agents with human-like behaviour, meaning that the players involved might consider defecting from an equilibrium outcome to, for example, increase his relative payoff in comparison to the other players payoff. Since this behaviour is not strictly rational the classification is not as suitable for games involving strictly rational players. Because the classification is partly based on experimental results, it could be difficult to mathematically generalize the classification to larger games.

A geometric classification system for 2×2 inverval-symmetric games, (Harris, 1969)

Harris, (1969) classification of 2×2 symmetric games is similar to the one by Rapoport, Guyer, and Gordon but there are some fundamental differences. Harris extends the Rapoport, Guyer, and Gordon classification by using an interval scale instead of the ordinal one, investigating the possible ties that can occur between payoffs. Harris mentions that many concepts and comparisons become meaningless when using only an ordinal scale for the payoffs and it is therefore interesting to classify using interval scale instead. Some similarities are that he has an experimental view when classifying games and also takes various psychological aspects into account when discussing games that are played iteratively.

His approach to classifying games is better in the respect that it is easier to alter the way the classification is done by manipulating the parameters used, which cannot be done as easily in the classification system done by Rapoport, Guyer, and Gordon.

Harris uses a geometric approach to classifying the 2×2 interval-symmetric games, representing each of the classes of the interval-symmetric games by different regions in the plane. He investigates the *ties* by looking at what games are defined at the boundaries between regions as well as the corners at $(\pm\infty, \pm\infty)$ and (the trivial uninteresting game) in $(0,0)$. He also extends this classification in a later article by classifying all 2×2 games using a similar approach but adding more planes (Harris, 1972).

Harris uses a generalisation of symmetric games by allowing one players payoffs to be a positive linear transformation of the other players payoffs. To begin with the way that Harris constructs the payoff matrix as presented in Table 1.18 to satisfy the conditions below.

$$\begin{cases} A = k_1 a + k_2 \\ B = k_1 b + k_2 \\ C = k_1 c + k_2 \\ D = k_1 d + k_2. \end{cases}$$

		Player 2	
		0	1
Player 1	0	(d, D)	(c, A)
	1	(a, C)	(b, B)

Table 1.18: Payoff Matrix

Because of the way that the matrix is constructed it always represents a symmetric game in the ordinal sense, with the constraint that $k_1 \geq 0$ and for an arbitrary $k_2 \in \mathbb{R}$. The constraint that $a \geq c$ is also imposed. This is to make the inequalities for r_3 and r_4 in terms of a, b, c and d unambiguous and means that Player 1 gets smaller payoff for strategy profile $\{0,1\}$ and larger for $\{1,0\}$ and vice versa for player 2. No generality is lost because the other case, where Player 1 gets the higher payoff in $\{1,0\}$, can be obtained by transposing both players payoff matrices, so it does not add anything interesting to include both games.(Harris, 1969)

The games are represented in the (r_3, r_4)-plane where

$$\begin{cases} r_3 = \frac{b-c}{a-c} = \frac{B-C}{A-C} \\ r_4 = \frac{a-d}{a-c} = \frac{A-D}{A-C}. \end{cases} \tag{1.4}$$

Note that from r_3 and r_4 alone it is impossible to numerically reconstruct the payoff matrix, only the ordinal relations between them. In order to restore the payoff matrix the k_1 and k_2 values, and the payoff differences must be preserved, but doing so would result in more parameters. Note that the values of r_3 and r_4 do not depends on k_1 or k_2.

The r_3 and r_4 parameters define 12 different regions in the plane by defining regions with the following inequalities:

$$\begin{cases} r_3 \lesseqgtr 1 \Leftrightarrow a \lesseqgtr b \\ r_4 \lesseqgtr 0 \Leftrightarrow a \lesseqgtr d \\ r_3 \lesseqgtr 0 \Leftrightarrow b \lesseqgtr c \\ r_3 + r_4 \lesseqgtr 1 \Leftrightarrow b \lesseqgtr d \\ r_4 \lesseqgtr 1 \Leftrightarrow c \lesseqgtr d \end{cases} \tag{1.5}$$

which, together with the previously assumed inequality $a > c$ define the different regions of the plane.(Harris, 1969)

By investigating the boundaries between regions Harris discovers that they all belong to some category of games belonging to some region in the (r_3, r_4) plane. The resulting regions defines the following types of games:

1. A total of 6 regions defining different kinds of no-conflict games. A game is a no conflict game iff d or b is the highest ranked payoff.

2. Two regions defining games with *strongly stable equilibria.*

3. One region with the Apology game.

4. One region with the (Restricted) Battle of the Sexes game.

5. One region with the Chicken game. This is divided into two with the additional restriction that $a + c \geq 2d$.

6. one region with the Prisoner's Dilemma, which is divided into 3 restricted regions depending on if $a + c \geq 2d$ or $2b \geq a + c$ or $2b < a + c < 2d$.

See Harris, (1969) for the complete map over the resulting regions. The total amount of regions is 12, the same amount of regions that Huertas-Rosero, (2003) got from his geometrical classification and the same amount of 2×2 symmetric games as Rapoport, Guyer, and Gordon, (1978) defines in their classification. Huertas-Rosero however uses different parameters and classifies only by different cases of NE and PO.

An advantage of this classification is that it can easily be modified to other variations, by considering other inequalities on r_3 and r_4 and examining the resulting regions. It is therefore easier to modify than the one made by Rapoport, Guyer, and Gordon. Another advantage with a geometrical representation of games is that it is easily illustrated. If, however this method is to be generalized with a higher amount of parameters, then this illustrative advantage may be lost. Therefore it might not be the most suitable classification approach to use if the aim is to generalize it to a larger number of players for example.

Similar to the classification done by Rapoport, Guyer, and Gordon, Harris discusses many irrational and psychological factors that could affect how players choose to play the game. This makes the classification interesting when trying to understand real world situations but applies poorly when investigating purely rational players.

Best response classification of 2×2 strategic games

In the article *A classification of* $2{\times}2$ *bimatrix games* (Borm and Du, 1987) a classification of mixed general 2×2 bimatrix games is proposed. The classification is based on what types of best reply correspondences, defined in definition 1.3.2, the two players have and on the types of Nash equilibria the correspondences induce. Borm and Du proposes that the set of 2×2 games should be divided into 15 classes and he also provides the fraction of the space of such games that each class represents by calculating the probability that a random game belongs to the class.

The central concepts in this method of classification is the concept of best reply correspondences. The definition for a best reply correspondence in pure strategic games is presented under **section**. Now we will provide a generalisation of the concept that extends to mixed strategy games as well.

Definition 1.3.2 (Best reply correspondence for mixed strategies)**.** Given a strategy profile $\sigma = (\sigma_1, \sigma_2) \in \Sigma_i$ in a 2×2 mixed strategy game $\Gamma(G)$ we say that σ_i is a best response of player i to the mixed strategy σ_{-i} of player $-i$ if $\forall \sigma'_i \in \Sigma_i \; u_i(\sigma_{-i}, \sigma_i) \geq u_i(\sigma_{-i}, \sigma'_i)$. The set of all best responses, denoted $B_i(\sigma_{-i})$, is called the *best reply correspondence* of player i to the mixed strategy σ_{-i} of player $-i$.

One quickly realises that if σ is a mixed strategy profile such that $\sigma_i \in B(\sigma_{-i})$ for both players $i \in \{1, 2\}$ then σ is a Nash equilibrium since for both players, σ satisfies $\forall \sigma'_i \in \Sigma_i \; u_i(\sigma_{-i}, \sigma_i) \geq u_i(\sigma_{-i}, \sigma'_i)$, i.e. σ satisfies the Nash conditions. There is in fact an equivalence relationship between σ being a Nash equilibrium and σ satisfying $\sigma_i \in B(\sigma_{-i})$ for both players.

Now we will define what we mean by the *graph* of a best reply correspondence. Notice that since $B_i(\sigma_{-i})$ can contain more than one element it is not a function but a *correspondence*. A correspondence can be described as a map that maps an element in one set to an element in the powerset of a set. That is, given two sets A and B the correspondence f is a map such that $f : A \to \mathcal{P}(B)$.

Definition 1.3.3 (Best reply correspondence graph.)**.** Given a 2×2 mixed extension game $\Gamma(G)$, we define the *graph* of B_i, $i \in \{1, 2\}$, as the set $\gamma(B_i) = \{\sigma \in \Sigma : \sigma_i \in B_i(\sigma_{-i})\}$.

An important fact for this classification method is that a strategy profile is a Nash equilibrium if and only if it is an element of the intersection between $\gamma(B_1)$ and $\gamma(B_2)$. This is formalised in the proposition below.

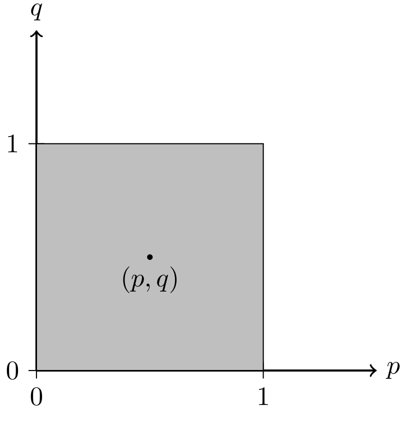
Proposition 1.3.4. Given a 2×2 mixed extension game $\Gamma(G)$, a mixed strategy profile $\sigma \in \Sigma$ is a Nash equilibrium if and only if $\sigma \in \gamma(B_1) \cap \gamma(B_2)$.

See Borm and Du, (1987) for a proof of proposition 1.3.4.

The reason for why this proposition is important is because it provides a method for finding all Nash equilibria of a 2×2 game. It states that one only has to determine the intersection of the graphs $\gamma(B_1)$ and $\gamma(B_2)$ to find every equilibria of a game. As we will see, there are only a finite number plausible of types of graphs and this allows us to categorize the games on the basis of what types of graphs they have.

Since we only consider 2×2 games at the moment a mixed strategy of a player $i \in \{1, 2\}$ is a pair $(p, 1-p)$ where $p \in [0, 1]$ is the probability of choosing the pure strategy s_1 and $1-p$ is the probability of choosing the strategy s_2. We can consider a pure strategy to be a special case of mixed strategies since $p = 1$ and $p = 0$ represents the pure strategies s_1 and s_2 respectively. Using this notation every mixed strategy profile is strictly determined by the pair $(p, q) \in [0,1]^2$, letting p define the strategy of player one and q the strategy of player two.

By associating a strategy profile with a pair (p, q) we can consider any strategy profile to be a point in the unit rectangle $[0,1]^2 \subset \mathbb{R}$.



Now that we have a way of describing strategy profiles as two-dimensional numbers, the following theorem provides some interesting results that makes it plausible to draw the graph of every type of best reply correspondence in the unit square.

Theorem 1.3.5 (Types of best reply correspondences)**.** Given a 2×2 mixed extension game $\Gamma(G)$ and a player $i \in P$, the best reply correspondence B_i has exactly one of the following properties:

(i) $\forall \sigma_{-i} \in \Sigma_{-i}\ B_i(\sigma_{-i}) = [0,1]$.

(ii) $\exists x \in \{0,1\} : \forall \sigma_{-i} \in \Sigma_{-i}\ B_i(\sigma_{-i}) = x$.

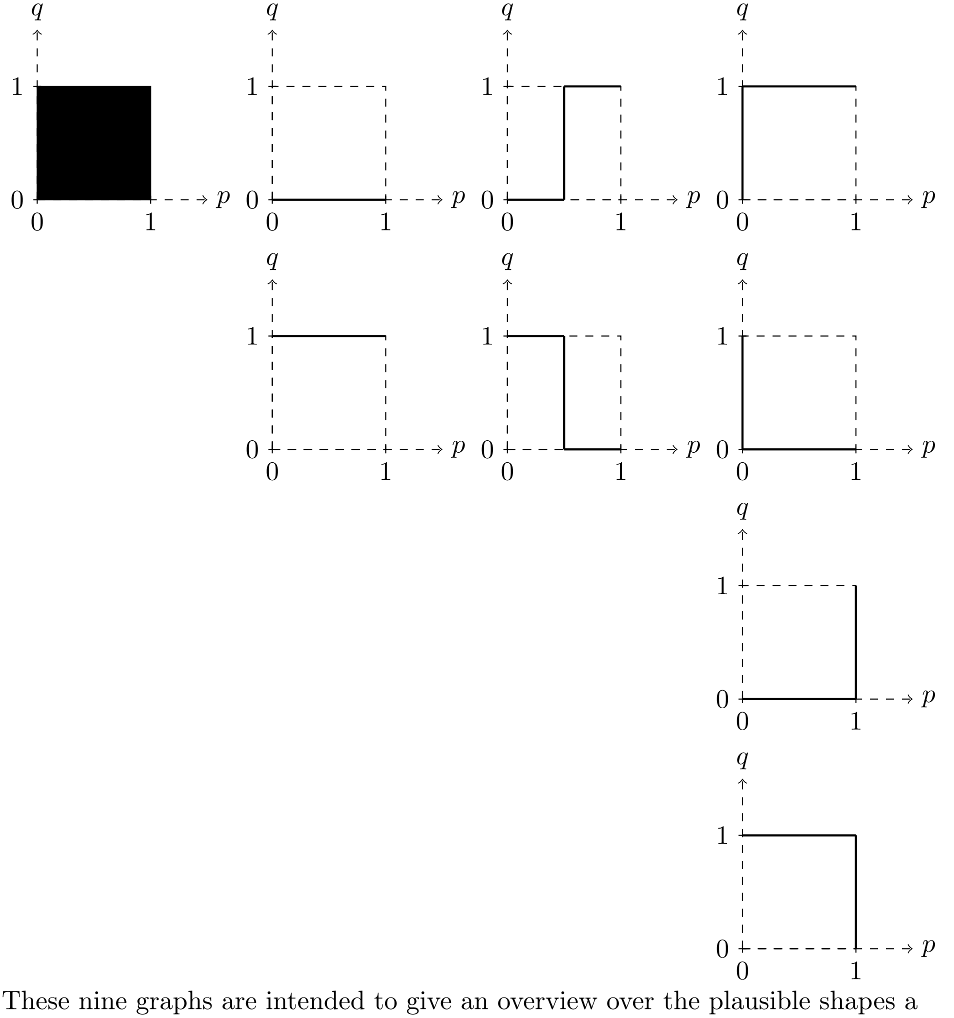
(iii) $\exists x, y \in \{0,1\}, z \in (0,1) : x \neq y$ and

$$B_i(\sigma_{-i}) = \begin{cases} x & \text{if } \sigma_{-i} \in (0, z), \\ [0,1] & \text{if } \sigma_{-i} = z, \\ y & \text{if } \sigma_{-i} \in (z, 1). \end{cases}$$

(iv) $\exists x, y \in \{0,1\}$:

$$B_i(\sigma_{-i}) = \begin{cases} [0,1] & \text{if } \sigma_{-i} = x, \\ y & \text{otherwise.} \end{cases}$$

That is, there are four types of best reply correspondences and hence four types of graphs. However the look of different graphs within one of these types can vary and it turns out that the four basic graphs can be drawn in nine different ways. Illustrations of these nine shapes are presented in **figure XXX**.



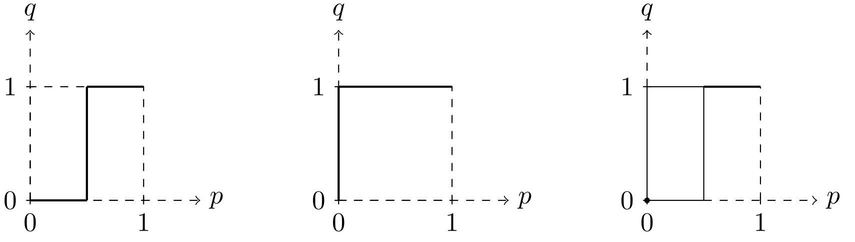
These nine graphs are intended to give an overview over the plausible shapes a BRC graph can have and to show the features of the four types of graphs.

We are soon ready to provide a classification of all 2×2 strategic games. The classification will partly be based upon the plausible combinations of best reply correspondences of the two players. We already stated that for each player there are four types of best reply correspondences. One might think that we would consider all 16 combinations of these correspondences but that is not the case since we do not wish to differ between the two players. That is, we consider for example the case where player one has a BRC of type (*ii*) and player two has a BRC of type (*iv*) to be equivalent to the case where player one has a BRC of type (*iv*) and player two has a BRC of type (*ii*). This is a reasonable assumption since the games in both cases has the same features.

To make further reasoning about the different combinations of BRC more clear we introduce the notation used by Borm.

Notation 1.3.6. Given a 2×2 game where player one has a best reply correspondence of type $a \in \{(i),(ii),(iii),(iv)\}$ and player two has a best reply correspondence of type $b \in \{(i),(ii),(iii),(iv)\}$, we denote the best response combination with $[a, b]$.

The consequence on the best response combinations of the fact that we do not distinguish between the two players can now be described by stating that $\forall a, b \in \{(i),(ii),(iii),(iv)\}\ [a,b] = [b,a]$. This implies that there are $\binom{4}{2} + 4 = 10$, not 16, best reply combinations in total. An example of what the graphs of a best reply combination can look like are illustrated in the figure below. The first graph is the graph of player one (type (*iii*)), the second is the graph of player two (type (*iv*)) and the third graph is the combination of the two where the intersection is highlighted representing all Nash strategy profiles of the game. **SHOW ALL GRAPHS IN APPENDIX?**



Now, we need to connect the type of graph a player has with the structure of the pay-off matrix. It turns out that this can be done by looking at the differences between some of the elements in the pay-off matrix. Suppose that the pay-off matrix of a game looks according to the matrix below

$$P_A = \begin{bmatrix} (a_{11}, b_{11}) & (a_{12}, b_{12}) \\ (a_{21}, b_{21}) & (a_{22}, b_{22}) \end{bmatrix}$$

where for example a_{12} is the pay-off for player one given that player one plays strategy $(1,0)$ and player two plays strategy $(0,1)$. If we denote the difference between a_{11} and a_{21} D_{A1} and in the same fashion define the variables $D_{A2} = a_{12} - a_{22}$, $D_{B1} = b_{11} - b_{21}$ and $D_{B2} = b_{12} - b_{22}$, we can conveniently describe the, to our purpose, important differences according to the system presented below. Note that D_{A1} denotes the differences between the element in the first column in player ones pay-off matrix, D_{A2} the denotes the differences between the element in the second column in player ones pay-off matrix and so on.

Notation 1.3.7. For player one we denote the differences D_{A1} and D_{A2} by $A[i,j]$, $i,j \in \{-1,0,1\}$ where $i = -1, i - 0$ and $i = 1$ means that $D_{A1} < 0, D_{A1} = 0$ and $D_{A1} > 0$ respectively and j has an analogue meaning for D_{A2}. The differences for player two are denoted by $B[i,j]$ according to the same system.

This notation is useful in the sense that it provides instant information of what action a player should play given that he knows what strategy the other player uses.

Example 1.3.8. Suppose player one has payoff matrix

$$\begin{bmatrix} 1 & 3 \\ 4 & 3 \end{bmatrix}.$$

Then the differences D_{A1} and D_{A2} are described by $A[-1,0]$ since $D_{A1} = 1-4 = -3 < 0$ and $D_{A2} = 3 - 3 = 0$.
Now if player two plays strategy $(1,0)$, player one should play the strategy $(0,1)$ since $A[-1,0]$ tells him that playing strategy $(0,1)$ results in higher payoff then strategy $(1,0)$. In the same way he can also tell that if player two instead chooses to play $(0,1)$ it does not matter if he plays strategy $(0,1)$ or strategy $(1,0)$, since both of them results in the same payoff.

Now we present a theorem connecting the differences $D_{pi}, i \in \{1,2\}$ for a player $p \in \{A,B\}$ with one of the four $(i),...,(iv)$ types of best reply correspondences presented in theorem 1.3.5.

Theorem 1.3.9. Given a 2×2 mixed strategy game $\Gamma(G)$ the type of best response correspondence $(i),...,(iv)$ described in theorem 1.3.5 of a player are determined by the differences $A[i,j]$ for player one and $B[i,j]$ for player two according to the table below.

Apart from basing the classification system on the best response combination Borm and Du also bases it upon the games *payoff combination* which is defined

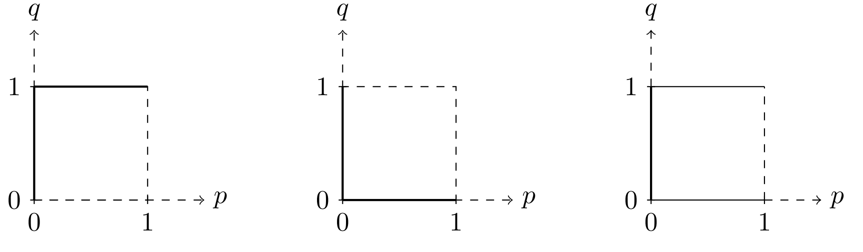
Type of BRC	Player one	Player two
(i)	$A[0,0]$	$B[0,0]$
(ii)	$A[-1,-1], A[1,1]$	$B[-1,-1], B[1,1]$
(iii)	$A[1,-1]$, $A[-1,1]$	$B[1,-1]$, $B[-1,1]$
(iv)	$A[1,0]$, $A[0,1]$, $A[-1,0]$, $A[0,-1]$	$B[1,0]$, $B[0,1]$, $B[-1,0]$, $B[0,-1]$

Table 1.19: Table of Classification of 2×2 symmetric non-zero sum games

as the pair $([i,j],[k,l])$ where i,j,k and l comes from the differences $A[i,j]$ and $B[k,l]$. Because we do not distinguish between different players, the payoff combination $([i,j],[k,l])$ is equivalent to the payoff combination $([k,l],[i,j])$. That means that there are $\binom{9}{2} + 9 = 45$ payoff combinations in total. The reason for why Borm and Du chooses to divide the ten classes defined by the best response combinations into more classes based on different types of payoff combination is simple. Within a few of the ten classes there are great variations of the types of Nash equilibria they can have. Consider for example the two cases of games with the same best response combinations below. Clearly they have very different sets of Nash equilibria.

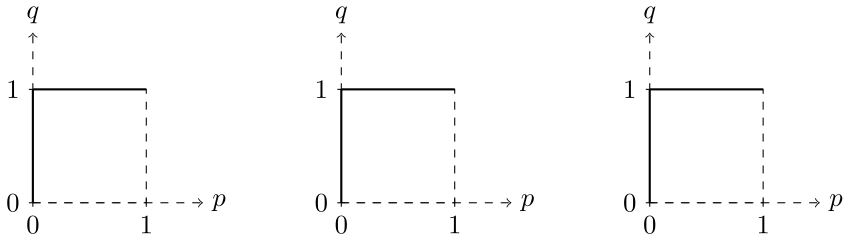
Example 1.3.10. Here we present two examples of games that have the same best response combination $[(iv),(iv)]$ but different types of Nash equilibria sets. The first two graphs in each case is the graphs of the BRC of each player and the third is the graph of their union with highlighted intersection (Nash set).

Case one:



Here, the Nash set is $\{0\} \times [0,1]$.

Case two:



In this case the Nash set is $\{0\} \times [0,1] \cup [0,1] \times \{1\}$.

This highlights the variations that can occur within a type of best response combination.

To avoid these huge variations in Nash sets within a best response combination it is necessary to make further distinctions within some om the combinations. These distinctions are made by dividing some of the ten classes into several classes by taking the payoff combinations into account in such a way that the structure of the Nash set within each class are the same. The final classification is presented in the table below and it contains a total of 15 classes.

This method of classification differs from the ones previously presented in several ways. It is, for example, a classification system that applies to all kinds of 2×2 games, not only strictly ordered or pure ones. It also differs in that it classifies games based on fewer concepts than the others, essentially just taking Nash equilibria into account. This might be considered a disadvantage since there are other important concepts, for example Pareto optimality, that are interesting to study in many situations and therefore might be a good aspect for a classification. However the simplicity of this method is also a strength since it allows us to divide every possible 2×2 game into just 15 classes, which is a relatively small number compared to what one would get if more aspects were taken into account.

Class	Best response combination	Payoff combination
I	$[(i),(i)]$	$([0,0],[0,0])$
II	$[(i),(ii)]$	$([0,0],[1,1])$ $([0,0],[-1,-1])$
III	$[(i),(iii)]$	$([0,0],[1,-1])$ $([0,0],[-1,1])$
IV	$[(i),(iv)]$	$([0,0],[1,0])$ $([0,0],[-1,0])$ $([0,0],[0,1])$ $([0,0],[0,-1])$
V	$[(ii),(ii)]$	$([1,1],[1,1])$ $([1,1],[-1,-1])$ $([-1,-1],[-1,-1])$
VI	$[(ii),(iii)]$	$([1,1],[1,-1])$ $([1,1],[-1,1])$ $([-1,-1],[1,-1])$ $([-1,-1],[-1,1])$
VII	$[(ii),(iv)]$	$([1,1],[1,0])$ $([1,1],[-1,0])$ $([-1,-1],[0,1])$ $([-1,-1],[0,-1])$
VIII	$[(ii),(iv)]$	$([1,1],[0,1])$ $([1,1],[0,-1])$ $([-1,-1],[1,0])$ $([-1,-1],[-1,0])$
IX	$[(iii),(iii)]$	$([1,-1],[1,-1])$ $([-1,1],[-1,1])$
X	$[(iii),(iii)]$	$([1,-1],[-1,1])$
XI	$[(iii),(iv)]$	$([1,0],[1,-1])$ $([-1,0],[-1,1])$ $([0,1],[-1,1])$ $([0,-1],[1,-1])$
XII	$[(iii),(iv)]$	$([1,0],[-1,1])$ $([-1,0],[1,-1])$ $([0,1],[1,-1])$ $([0,-1],[-1,1])$
XIII	$[(iv),(iv)]$	$([1,0],[1,0])$ $([-1,0],[0,1])$ $([0,-1],[0,-1])$
XIV	$[(iv),(iv)]$	$([1,0],[0,1])$ $([-1,0],[1,0])$ $([0,-1],[-1,0])$ $([0,1],[0,-1])$
XV	$[(iv),(iv)]$	$([1,0],[0,-1])$ $([-1,0],[-1,0])$ $([0,1],[0,1])$

A Geometric classification of symmetric 2×2 games

In this section we will present a way of classifying 2×2 symmetric non-zero sum games represented in the strategic form constructed by Huertas-Rosero, (2003). These games are completely defined by the payoff matrix of one of the players, since the other players payoff matrix is the transpose of the others matrix, see definition 1.2.28. This means that the number of parameters is 4.

Let $G = (S, U, P)$ be a game where $S = \{0, 1\} \times \{0, 1\}$ and $P = \{1, 2\}$ with U being a set of payoff functions. Player 1's payoff matrix is defined by

$$P_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and $P_2 = P_1^t$.

The parameter vector $\vec{v} = [a, b, c, d]^t$ defines a 4-dimensional parameter space that represents the properties of G. The reasoning below excludes the trivial cases where $a = b = c = d$ in which case each player is completely indifferent about his choice of action.

By using the properties of multiplicative and additive invariance followed by an isometry Huertas-Rosero reduces the number of parameters to 3 and projects the parameter space onto the 3-dimensional unit sphere. Using the property of additive invariance Huertas-Rosero translates the parameter space by subtracting the mean from each parameter. The new parameter space is the defined by

$$\vec{v}' = [a', b', c', d']^t = \vec{v} - \frac{1}{4}(a + b + c + d)[1, 1, 1, 1]^t. \tag{1.6}$$

By using the multiplicative invariance property Huertas-Rosero normalize the vector $\vec{v}'$ with the euclidean norm

$$\vec{v}'' = [a'', b'', c'', d'']^t = \frac{1}{||\vec{v}'||}\vec{v}' \tag{1.7}$$

making the parameter vector unitary.

By using the isometry described by

$$\Delta\vec{\mathbb{E}} = \begin{bmatrix} E_0 \\ \Delta E_A \\ \Delta E_B \\ \Delta E_{AB} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} a'' \\ b'' \\ c'' \\ d'' \end{bmatrix} \tag{1.8}$$

one can observe that $E_0 = 0$ and hence the parameter space is projected onto the 3-dimensional unit sphere with all the essential properties of G preserved. This implies that the game can be described by the vector

$$\Delta\vec{\mathbb{E}}^* = [\Delta E_A, \Delta E_B, \Delta E_{AB}].$$

In the vector above ΔE_A can be interpreted as the expected payoff difference between the player choosing action 0 or 1 given that the other player plays 0 or 1 with probability $\frac{1}{2}$. ΔE_B can be interpreted as the expected payoff difference of the opposing player given that the first player chooses 0 or 1 with probability $\frac{1}{2}$. Since the game is symmetric ΔE_A and ΔE_B is the same for both players. ΔE_{AB} is the expected payoff difference between the players choosing the same action versus different actions.

Note that $\Delta E_A, \Delta E_B$ and ΔE_{AB} are all pairwise orthogonal unit vectors, i.e. they form an orthonormal basis in $\mathbb{R}^3$. This is a crucial fact for the theory presented below.

Huertas-Rosero, (2003) claims to classify by Pareto optimality. **d**efinition of Pareto optimality is however not equivalent to the standard definition used. We will refer to his version of Pareto optimality as *Huertas optimality*.

Definition 1.3.11 (Huertas optimality)**.** A strategy profile s is called *Huertas optimal* if

$$\forall i \forall j : j \neq i \;\; u_j(s_{-i}, s_i) \geqslant u_j(s_{-i}, s_i'), \forall s_i', s_i' \neq s_i.$$

Huertas optimality, henceforth denoted as HO, is explained by Huertas-Rosero, (2003) as an outcome obtained when each player tries to maximize his opponents payoff, i.e. the HO conditions is equivalent to the NE conditions on the transposed payoff matrix.

To realise that this definition is not equivalent to the standard definition of Pareto optimality consider the game in Table 1.20.

		Player 2: 0	Player 2: 1
Player 1	0	$(3,3)$	$(4,1)$
	1	$(1,4)$	$(2,2)$

Table 1.20: Game 1

		Player 2: 0	Player 2: 1
Player 1	0	$(3,3)$	$(1,4)$
	1	$(4,1)$	$(2,2)$

Table 1.21: Game 2

To solve for HO we first transpose the payoff matrix of Game 1, which results in Game 2. The NE strategy profile in Game 2 is $\{1,1\}$ and hence the HO strategy profile in Game 1 is $\{1,1\}$. This is clearly not Pareto optimal since $\{0,0\}$ Pareto dominates $\{1,1\}$.

We interpret Huertas optimality as a *altruistic* Nash equilibrium, since it is obtained by all players maximising their opponents payoff functions instead of their own.

The proposed method for classifying G is to categorize it by different conditions on its Nash equilibria and Huertas optimality. Every game considered here is guaranteed to have at least one NE (Huertas-Rosero, 2003). Because of the symmetry Huertas optimality can be solved for by interchanging the players payoff matrices and solving for NE, therefore Huertas optimality is always guaranteed to exist.

Now expressing the inequalities for NE in the new parameters Huertas-Rosero get the following conditions where i represents the action of Player 1 and j the action of player 2. Huertas-Rosero, (2003) claims the following proposition without proof.

Proposition 1.3.12 (Huertas NE conditions)**.** An outcome $\{i,j\}$ is a NE iff

$$(-1)^i(\Delta E_A + (-1)^j \Delta E_{AB}) \geqslant 0,$$

$$(-1)^j(\Delta E_A + (-1)^i \Delta E_{AB}) \geqslant 0.$$

This proposition holds and to realise this we provide the proof below.

Proof. To show this we first note that given $\{0,0\}$ a NE then this is equivalent to $a \geq c$. Likewise $\{1,1\}$ NE $\Leftrightarrow d \geq b$, $\{1,0\}$ NE $\Leftrightarrow b \geq d, c \geq a$ and $\{0,1\}$ NE $\Leftrightarrow c \geq a, b \geq d$. We would now like to express these inequalities in terms of the combinations of i and j as in the inequalities above. Expressing the inequalities as

$$\begin{cases} \alpha' a + \beta' b + \delta' c + \gamma' d \geq 0 \\ \alpha'' a + \beta'' b + \delta'' c + \gamma'' d \geq 0 \end{cases} \tag{1.9}$$

and then determining the coefficients expressed in terms of $i$and j will result in the following. For example if $\{0,0\}$ is NE then $a \geq c$ so that means that $\alpha' = \alpha'' = 1$ and $\delta' = \delta'' = -1$ and the other coefficients are 0. By analogous reasoning for the other cases of reasoning for NE you get the following coefficients expressed in i and j:

$$\begin{cases} \alpha' = (-1)^i(1+(-1)^j) \\ \beta' = (-1)^i(1+(-1)^{j+1}) \\ \delta' = (-1)^i(-1+(-1)^{j+1}) \\ \gamma' = (-1)^i(-1+(-1)^j) \end{cases} \tag{1.10}$$

$$\begin{cases} \alpha'' = (-1)^j(1+(-1)^i) \\ \beta'' = (-1)^j(1+(-1)^{i+1}) \\ \delta'' = (-1)^j(-1+(-1)^{i+1}) \\ \gamma'' = (-1)^j(-1+(-1)^i) \end{cases} \tag{1.11}$$

By putting $i = j = 0$ we obtain the desired values $\alpha' = \alpha'' = 1$, $\delta' = \delta'' = -1$ and $\beta' = \beta'' = \gamma' = \gamma'' = 0$ which means $a \geq c$ and similarly for the other cases of NE.

By substituting the coefficients above for the ones expressed in i and j we get the following inequalities:

$$\begin{cases} (-1)^i(1+(-1)^j)a + (-1)^i(1+(-1)^{j+1})b+ \\ +(-1)^i(-1+(-1)^{j+1})c + (-1)^i(-1+(-1)^j)d \geq 0 \\ \\ (-1)^j(1+(-1)^i)a + (-1)^j(1+(-1)^{i+1})b+ \\ +(-1)^j(-1+(-1)^{i+1})c + (-1)^j(-1+(-1)^i)d \geq 0 \end{cases} \tag{1.12}$$

$$\Leftrightarrow$$

$$\begin{cases} (-1)^i(a+b-c-d) + (-1)^j(a-b-c+d) \geq 0 \\ \\ (-1)^j(a+b-c-d) + (-1)^i(a-b-c+d) \geq 0 \end{cases} \tag{1.13}$$

$$\Leftrightarrow$$

$$\begin{cases} (-1)^i\left(\frac{a+b-(c+d)}{2} + (-1)^j\,\frac{a+d-(b+c)}{2}\right) \geq 0 \\ \\ (-1)^j\left(\frac{a+b-(c+d)}{2} + (-1)^i\,\frac{a+d-(b+c)}{2}\right) \geq 0 \end{cases} \tag{1.14}$$

$$\Leftrightarrow$$

$$\begin{cases} (-1)^i(\Delta E_A + (-1)^j \Delta E_{AB}) \geq 0, \\ (-1)^j(\Delta E_A + (-1)^i \Delta E_{AB}) \geq 0. \end{cases} \tag{1.15}$$

□

This means that the NE condition can be expressed using only two parameters. The intuitive reason for why ΔE_B is not included in the inequalities above is because the definition of NE implies that neither player cares about the other players payoff. Similarly is Huertas optimality neither player cares about his own payoff, he just tries to maximize the payoff for his opponent, hence to obtain the inequalities for Huertas optimality ΔE_A is replaced by ΔE_B in the inequalities for NE. To express the inequalities for Huertas optimality we once again express them on the form introduced by Huertas-Rosero, (2003).

Proposition 1.3.13 (Huertas HO conditions)**.** The strategy profile (i, j) is Huertas optimal if it satisfies the following inequalities:

$$(-1)^i(\Delta E_B + (-1)^j \Delta E_{AB}) \geqslant 0,$$

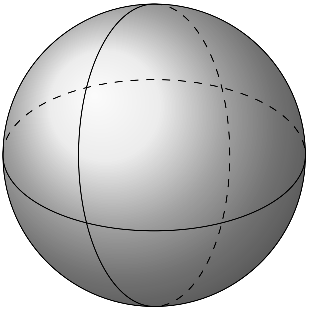
$$(-1)^j(\Delta E_B + (-1)^i \Delta E_{AB}) \geqslant 0.$$

The proof of proposition 1.3.13 is analogous to the proof of proposition 1.3.12 .

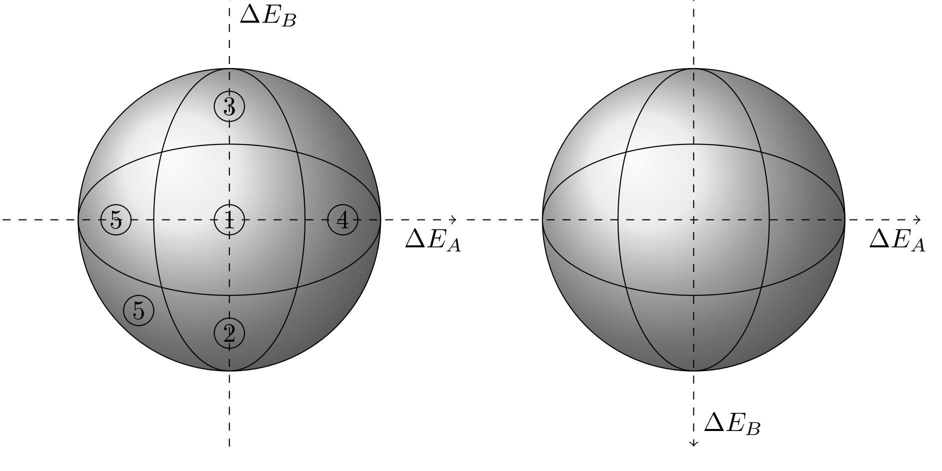
The inequalities for NE and HO defines two pairwise orthogonal planes each in $\mathbb{R}^3$ through the origin. These planes divides the unit sphere into fourteen different pieces, each defined by the set of inequalities they satisfy. (Huertas-Rosero, 2003)

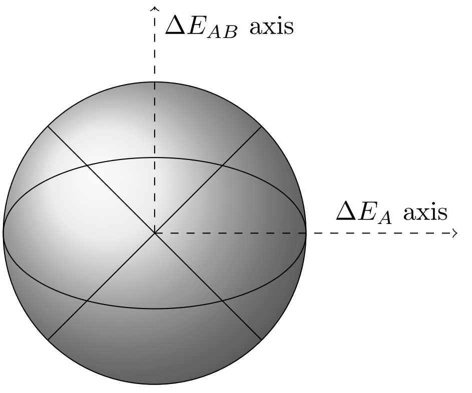
One might consider classifying games only based on which of the pieces of the unit sphere they belong to and that would define 14 classes of games. Note however that you can always interchange the order of the actions in the action set of a game and it will still be the same game. That is, if you interchange the rows and columns in the payoff matrix, i.e. transposing both players payoff matrices, you will still end up with the same game. This causes a problem when dividing the sphere into the fourteen pieces since some of those fourteen games are *strategically equivalent.* It makes sense to put strategically equivalent games into the same class, so this results in a total of 8 non-equivalent games. By further distinguishing games by if the payoff of the NE is higher than the payoff of HO, Huertas-Rosero get a total of 12 classes of games. The classification is illustrated in the **graph/ table** below.(Huertas-Rosero, 2003)

The planes defined by the Nash equilibria inequalities are orthogonal and will pass through the origin of the game sphere (Huertas-Rosero, 2003). This results in two great circles on the sphere dividing it into four different regions according to the illustration below.



In the same fashion the inequalities for Huertas optimality will define two orthogonal planes such that the line defined by these planes and the lines defined by the Nash planes are orthogonal (Huertas-Rosero, 2003). The regions of the sphere defined by the four planes are illustrated in the figures below.





Example 1.3.14. The piece of the sphere surrounding the north pole, which we call N, is defined by the following inequalities;

$$\begin{cases} \Delta E_{AB} \geq \Delta E_B \\ \Delta E_{AB} \geq -\Delta E_B \\ \Delta E_{AB} \geq \Delta E_A \\ \Delta E_{AB} \geq -\Delta E_A \end{cases}$$

which describes games where the strategy profiles $\{0,0\}$ and $\{1,1\}$ are NE and HO. To show this, we express the inequalities in the original parameters a'',b'',c'' and d''. The first inequality above can re-written as follows;

$$\frac{1}{2}(a'' + d'' - (b'' + c'')) \geq \frac{1}{2}(a'' + c'' - (b'' + d'')) \Leftrightarrow d'' \geq c'' \Leftrightarrow d \geq c.$$

By analogous calculations we get the following NE and HO conditions:

$$\begin{cases} a \geq c \\ d \geq b \end{cases} \qquad \begin{cases} a \geq b \\ d \geq c \end{cases}.$$

This method of classification has some similarities with the method proposed by Harris since they are both geometrical approaches that introduces parameters derived from the payoff matrices and uses these parameters to create a geometrical representation of the space of symmetrical games. Both methods also restrict themselves to symmetric games, making them less general than the other classification systems discussed in this chapter. Huertas-Rosero and Harris chooses to divide the space of games into regions based on different types of inequalities and they choose different geometrical representations. Huertas-Rosero three dimensional representation preserves more information about the original game in comparison to the two dimensional representation proposed by Harris.

Since this classification is derived from strictly theoretical concepts it might be easier to generalise than the classifications that are based partly on experimental concepts. In the case of generalisation one would however, as in the case with Harris classification, lose the illustrative advantages of the method.

A Topologically-Based Classification of the 2×2 Ordinal Games

Some of previous classifications discussed have been typological Rapoport, Guyer, and Gordon, (1978). This means that you identify some relations that you think are important and then classify with respect to those relations. A good example is the classification of plants (Robinson and Goforth, 2003). They can be classified for example by if they are edible or not. But by using this kind of classification plants that have no "deep" relation can end up in the same category. The classification can then be difficult to extend to other properties that might seem interesting. Thus it might be wiser to classify by deeper biological relations such as similarities in genes. This naturally requires deeper understanding about plants and thus the typological approach is more often adopted when a science is young and therefore the typological approach is considered more shallow (Robinson and Goforth, 2003).

In the article *A Topologically-Based Classification of the* 2×2 *Ordinal Games* by Robinson and Goforth the topological basis for typological classification is investigated. A total of 144 2×2 games is found, to be compared Rapoport, Guyer, and Gordon who claimed it to be 78 different kinds, because they do not distinguish between the players. The topology created by the games consists of 4 torii which connects into a multi-layered torus which has some interesting properties. They assume strictly ordinal relations between payoffs.

A topology is generated from a set of abstract objects called points. An operation is introduced by which a notion of neighbouring points can be defined. The relations between the objects is then studied based on the operation and neighbourhood defined. In this case the points will be the 2×2 games and the neighbourhood all swaps of adjacent payoffs.

When considering what definition to use for neighbouring games the natural way is the smallest possible change you can make in the payoff functions by changing the preference order defined by the current payoff function (Robinson and Goforth, 2003). This could be done by swapping the i:th ranked payoff for player X for the $i+1$:th. Borrowing notation from Robinson and Goforth, this operation will be denoted by X_{ij} where X belongs to the set of players and denotes what players payoff is being changed. Furthermore $i = k \in \{1,..,3\}, j = k+1$ means that the k:th ranked payoff is being swapped with the $k+1$:th. For example, using A_{12} on the following game

		Player A	
		0	1
Player B	0	$(2,4)$	$(4,3)$
	1	$(1,2)$	$(3,1)$

results in this game

		Player A	
		0	1
Player B	0	$(2,4)$	$(4,3)$
	1	$(1,1)$	$(3,2)$

where the lowest and second lowest payoff for player 2 have been swapped.

Following the reasoning above these games should also be neighbours. The set of neighbours $N(G)$ of a game $G(S,U,P)$ is defined by Robinson and Goforth, (2003) as follows

$$N(G) = \{X_{ij} : i \in \{1,..,3\}, j = i+1, X \in P\}.$$

That means that a game G is a neighbour of G' if $X_{ij}(G) = G'$ where $j = i+1$ for some $i \in \{1,..,3\}$. Note that every game then has a total of 6 neighbours because i can be assigned 3 values and the number of players is two.

An interesting feature of the topological approach is that by starting with an arbitrary game you can construct every game in the topology only by using the X_{ij} operation. Depending on which game is chosen to start with, different configurations of the map will be created. In the article by Robinson and Goforth they start with a particular symmetric game to get a structure on the topological map with the symmetric games on the diagonal on each layer and the games in each quadrant having different number of dominant strategies in the first layer. This leads to a particular configuration of the topological map which Robinson and Goforth call the *Dominant strategy configuration.*

First by observing some properties of the X_{ij} operation we can start to derive the 144 2×2 games. To begin with note that the operation is of order 2, i.e.

$$X_{ij}(X_{ij}(G)) = X_{ij}^2(G) = G$$

so it creates a loop of size 2, which is the smallest cycle that can be created. Note that

$$A_{12}(A_{23}(G)) \neq A_{23}(A_{12}(G)),$$

i.e. they do not commute, because they affect the same payoffs. On the other hand

$$A_{12}(B_{23}(G)) = B_{23}(A_{12}(G))$$

since they affect different payoffs. Therefore it does not matter in which order the B_{ij} and A_{ij} operations are performed but it does matter in which order operations affecting the same payoff are performed, except of course in the X_{ij}^2 case.

Note that

$$(A_{12}(A_{23}(G)))^3 = G.$$

The sequence of operations above forms a cycle of length 6. We will borrow the notation of a cycle from Robinson and Goforth which for the 6-cycle above is $[A_{12} + A_{23}]^\star$, which means the cycle obtained from using the swap-operations A_{12} and A_{23}.

The cycle created by the X_{12} operations create a cycle of length four, which is the authors call *tiles*. These are the smallest possible subspaces of the topological space and can be connected by either X_{23} swaps or X_{34} swaps, spanning the entire space.

$$(A_{12}(B_{12}(G)))^2 = G.$$

Another interesting property is that given that the game G is symmetric, then the compososition of either the X_{12} swaps or the X_{23} swaps, will lead to another symmetric game. Therefore, when starting with a symmetric game, the symmetric games will end up in the diagonal of each grid, which is an advantageous property of the map created.

By restricting yourself to fewer operations than the 6 defined above you can create *subspaces* of the topological space. However only 5 of the swap operations are needed to span the entire space (Robinson and Goforth, 2003). With fewer operations allowed the smaller the subspace gets. Depending on the number of neighbouring games a game has in a subspace a subspace can be more or less dense. The more neighbours the games in the subspace has, the denser the subspace is.

To begin creating the topological map Robinson and Goforth, (2003) start with the game which they call game 1 and restrict to the operations X_{12} and X_{23}. Keeping the properties above in mind, these operations creates the 6-cycles $[A_{12} + A_{23}]^\star$ and $[B_{12} + B_{23}]^\star$. The games defined by these operations can be represented in a grid created as follows. First games in the $[B_{12} + B_{23}]^\star$ cycle can be placed in a first row of the grid, where by changing columns and moving along the row, player A's payoffs are unchanged. Then by starting a $[A_{12} + A_{23}]^\star$ cycle from each game in the $[B_{12} + B_{23}]^\star$ cycle, the columns of the grid can be formed. When switching rows, and thereby moving along a column, player B's payoffs are invariant. Because the B_{ij} and A_{ij} operations commute, it does not matter in which order this is done. This creates a closed subspace with $6 \times 6 = 36$ games, which Robinson and Goforth calls layer 1. Note that the highest ranked payoff is left unaffected in this definition of a layer. You could imagine creating the layer by restricting to the $[X_{12}+X_{34}]^\star$ operations for example. The authors motivates this choice of layer by that the highest ranked payoff has the most impact on what kind of NE the game has. In this case, game 1 has the highest ranked payoff in diagonally opposite corners of the payoff matrix.

Some properties of layer 1 is that the in the games in the top three rows player 1 has one dominant strategy, and in the 3 left most columns player 2 has one dominant strategy. So this divides the grid into 4 quadrants representing the different orders defined by Rapoport, Guyer, and Gordon, (1978) in their classification. This grid in defines a *torus* because the loops connects the right most column to the left most column, and the upper most row to the lower most row.

Because the configuration of the highest payoff can be placed in 4 different ways, four layers will be defined, connected by X_{34} operations. In this way the result of 144 2×2 games is gained, because $36 \times 4 = 144$.

The torii defined by the different layers are in turn connected by other torii by using the $[X_{12} + X_{34}]^\star$ loops. These loops generate games that are connected from one layer to one or two other layers, but that are within the same *stack*. A stack is defined as the set of games that align vertically within a layer. An interesting property to observe about a game in a particular stack is that the corresponding one in a different layer is just a relabelling of the column-players actions, given that he is unaware of the row-players payoffs.

The equivalent of the typologically classified 78 2×2 games found by Rapoport and Guyer can be found in the topological map by removing the games in layer 2 and all games below the diagonal in layer 2 and layer 3. (Robinson and Goforth, 2003)

This method distinguishes from the others by that begins by only investigating the purely topological relations between games and only after establishing the topological map it is divided into regions depending on the game theoretical conditions thought important. It is therefore in a sense the most general classification compared to the others discussed in this chapter. An important difference compared to all the other classifications is that he distinguish between the two players. This is the reason for why he considers 144 2×2 games instead of the previously established 78 2×2 games. Since this is a purely mathematical approach it is more suitable for generalisation than most of the other classifications covered here.

Summary

The classification done by Rapoport, Guyer, and Gordon, (1978) has many interesting qualities, but due to the fact that the relations chosen are more experimental than mathematical it is not as suitable for mathematical generalisation compared to the other methods discussed. This does not mean that the classification is meaningless, on the contrary it brings up a lot of interesting results that are applicable in many real world situations.

The geometrical classifications by Harris, (1969) and Huertas-Rosero, (2003) are more promising for generalisation. Huertas-Rosero, (2003) classification has simpler classification conditions and could therefore be a better choice for generalisation. A problem with the geometrical approach is that the number of parameters increase quickly in bigger games. The dimensionality of the geometrical representation would therefore quickly grow beyond three dimensions and it would therefore not be as illustrative, as is the advantage when classifying 2×2 games geometrically.

The classification using best response correspondence is a mathematical classification system that might be appropriate to generalise. However, we consider the fact that it only takes Nash equilibria into account a limitation in the usefulness of the classification since other important game theoretical concepts are not represented. This makes it interesting to investigate if the method is extendable and if it could be modified to include more aspects, for example Huertas optimality as in the classification done by Huertas-Rosero, (2003).

The topological approach is the most general and elegant in its simplicity. It shows promise for generalisation because of the mathematical nature of the classification that is not limited by the dimensionality of the problem, which is a big advantage comparing with the other classification approaches. A critique to this classification method is that the measure of similarity used to define neighbouring games are not necessarily that similar. They have less focus on analysing what properties make a game interesting and more focus on mathematical structures. This makes us suspect that this approach might not be the best when trying to classify games into interestingly different classes. Another problem is the rate that the total amount of games grows fast with higher dimensionality, making it impossible to construct a complete useful map of the 3×3 games for example. We will present a formula for the number of $m \times n$ games in Chapter 2.

When analysing these classifications we notice that most approaches captures some interesting properties, but fails to capture other interesting properties that another approach instead manages to include. Rapoport, Guyer, and Gordon, (1978) base their classification on experimentally verified interesting properties, but they do not provide any mathematical basis making it harder to compare classes theoretically and less suitable to generalise to higher dimensions. The topological approach by Robinson and Goforth, (2003) is in a way an opposite approach to the one by Rapoport, Guyer, and Gordon. They focus a lot on creating mathematical structure but instead lose too much focus on answering what properties actually make a game interesting. A strength of the classification by Borm and Du, (1987) is also its mathematical structure and simplicity. Since they use best reply correspondence as the basis for the classification it is interesting in a strategic sense, it tells you how a certain game is supposed to be played. The weakness of this approach is that it has less focus on what games are interesting from a game theoretic point of view. The geometrical approach by Huertas-Rosero, (2003) is also more mathematical. There is however no motivation provided for why it is interesting to classify based on Huertas optimality. The HO and NE conditions do seem to capture something interesting about the 2×2 symmetric games, but it needs to be investigated further to determine if these concepts actually captures what is interesting about these games. Harris, (1969) combines the experimental motivation of what properties make a game interesting with a mathematical parameterization and geometrical illustration. Harris can therefore give a strong motivation for why the resulting classes are interesting and also have a good mathematical structure that can be generalised and analysed.

The conclusion of this analysis is that for a classification to be useful it needs have the following properties:

1. conditions that has some motivation of why they are interesting,

2. mathematical structure so that it is more easily analysed, be possible to generalise to higher dimensions and extendable by adding more conditions,

3. include all standard games,

4. have mathematical sophistication so that no conditions need to be added ad hoc.

Our impression is that the classifications reviewed in this thesis fail to find the right balance between these aspects. We summarise which of the aspects we find important is and is not included in each of the classification methods in the table below.

Classification Author	Mathematical Structure	Interesting Conditions	Mathematical Sophistication	All Standard Games
Rapoport, Guyer, and Gordon	No	Yes	No	No
Robinson and Goforth	Yes	No	Yes	No
Harris	Yes	Yes	No	No
Borm and Du	Yes	No	Yes	No
Huertas-Rosero	Yes	Yes	No	No

Table 1.22: Summary

With mathematical structure we mean that classification is based on mathematical conditions. With interesting conditions we mean that the conditions of which the classification is made has some motivation of why these conditions separate games in an interesting way. With mathematical sophistication we mean that no restriction on the mathematically defined regions are added ad hoc. So a classification is mathematically sophisticated if the initial mathematical conditions capture everything the author finds interesting to classify by. With the final condition concerning standard games we mean if the classification captures all the, by us known, standard games.

Chapter 2

Results

In this chapter the main results will be presented. The first section is an analysis of the literature review in Chapter 1 where the strengths and weaknesses between the classifications are compared. In the following section we discuss what properties make a game interesting. To answer this we decompose the standard games and analyse the result. In the succeeding section we present our classification based on the decomposition of 2×2 symmetric games. The two following sections are devoted to analysing this classification theoretically and experimentally. The two final sections are separate from the classification presented. In the section Number of Games we estimate the number of $m \times n$ games. The final section contains a discussion about the interesting properties the NE concept captures.

Classification by Decomposition

Based on the discussion on the standard games in the previous section, it seems that a big aspect of what makes a game interesting is conflict of interest between the players. The particular combination between direction and strength of conflict and common interest seem to decide what type of game it is. We therefore propose a classification of the symmetric 2×2 games based on the decomposition into common interest and conflict. We will use Stereographic projection to project the regions on the sphere onto the complex plane, making the map more illustrative.

Decomposition of symmetric 2×2 games

Theorem 1.2.31 describes how a general 2×2 game matrix can be decomposed into one common interest matrix and one conflict matrix. In the special case of 2×2 symmetric games the decomposition of the payoff matrix has the form shown in 2.1.

$$\begin{bmatrix} (a,a) & (b,c) \\ (c,b) & (d,d) \end{bmatrix} = \begin{bmatrix} (a,a) & (\frac{b+c}{2}, \frac{b+c}{2}) \\ (\frac{b+c}{2}, \frac{b+c}{2}) & (d,d) \end{bmatrix} + \begin{bmatrix} (0,0) & (\frac{b-c}{2}, -\frac{b-c}{2}) \\ (-\frac{b-c}{2}, \frac{b-c}{2}) & (0,0) \end{bmatrix} \tag{2.1}$$

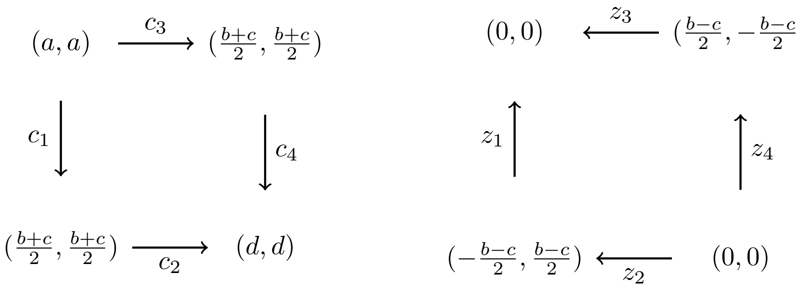
Notice that since the game is symmetric, the first player's payoff matrix is the transpose of the other player's payoff matrix. This means that the game, and hence the decomposition of it, is determined by one of the player's payoff matrix. The decomposed symmetric 2×2 payoff matrix is determined by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & \frac{b+c}{2} \\ \frac{b+c}{2} & d \end{bmatrix} + \begin{bmatrix} 0 & \frac{b-c}{2} \\ -\frac{b-c}{2} & 0 \end{bmatrix}.$$

The main reason for why we want to decompose games is to find out if the conflict part of the game is a counter effect to the common interest part of the game and if so, how strong the counter effect is. For example if $a > \frac{b+c}{2}$ player one prefers outcome 00 in the common interest matrix over outcome 10 and so on. These preferences can be described by four arrows according to below. The same concept is used to describe the type of conflict of a game and in the right picture the possible preference arrows of the conflict matrix is shown.

$$\begin{array}{ccc} (a,a) & \longleftrightarrow & (\frac{b+c}{2}, \frac{b+c}{2}) \\ \updownarrow & & \updownarrow \\ (\frac{b+c}{2}, \frac{b+c}{2}) & \longleftrightarrow & (d,d) \end{array} \qquad \begin{array}{ccc} (0,0) & \longleftrightarrow & (\frac{b-c}{2}, -\frac{b-c}{2}) \\ \updownarrow & & \updownarrow \\ (-\frac{b-c}{2}, \frac{b-c}{2}) & \longleftrightarrow & (0,0) \end{array}$$

However it is not only the direction of the arrows, that is the preference relations, that matter since we also need to know how strong the arrows are in order to know how strong the common interest preferences are in comparison to the conflict preferences. Therefore we represent the arrows and their strength with one dimensional vectors. The vertical arrows represents the preferences of Player 1 and the horizontal arrows represents the preferences of Player 2.



In the figure above $c_1 \triangleq -a + \frac{b+c}{2}$, $c_2 \triangleq d - \frac{b+c}{2}$, $z_1 \triangleq 0 - (-\frac{b-c}{2}) = \frac{b-c}{2}$ and because of symmetry $c_1 = c_3, c_4 = c_2$ and $z_1 = z_2 = z_3 = z_4$.

Since the transformation matrix taking $[a, b, c, d]^t$ to $[c_1, c_2, z_1]^t$ is not invertible we define the variable $x = \frac{a+b+c+d}{2}$ and get the invertible linear transformation shown in equation 2.2.

$$\begin{bmatrix} x \\ c_1 \\ c_2 \\ z_1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -2 & 1 & 1 & 0 \\ 0 & -1 & -1 & 2 \\ 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = A \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \tag{2.2}$$

The transformation matrix A in equation 2.2 is an isometry since $|A| = 1$.

Since we assume additive invariance we can translate the $[a, b, c, d]^t$ vector such that $a + b + c + d = 0$ as described in equation 2.3.

$$\begin{bmatrix} a' \\ b' \\ c' \\ d' \end{bmatrix} \triangleq \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} - \frac{a+b+c+d}{4} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \tag{2.3}$$

This translation makes $x = 0$ for every a', b', c' and d' and thus if we define a vector space

$$\begin{bmatrix} x' \\ c_1' \\ c_2' \\ z_1' \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + z_1 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \text{ where } x', c_1', c_2', z_1' \in \mathbb{R}. \tag{2.4}$$

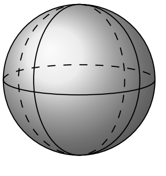
Since $x = 0$ every symmetric 2×2 game can be represented by a vector $\begin{bmatrix} 0 & c_1' & c_2' & z_1' \end{bmatrix}^t$, denoted $\begin{bmatrix} c_1 & c_2 & z_1 \end{bmatrix}^t$, in the three-dimensional subspace of $\mathbb{R}^4$ where $x = 0$. Since the vectors defining x, c_1, c_2 and z_1 in equation 2.2 are linearly independent, we know that the subspace defined by the vectors $\begin{bmatrix} c_1 & c_2 & z_1 \end{bmatrix}^t$ spans $\mathbb{R}^3$.

Here we do not consider the trivial games where $a = b = c = d$ and hence we define the space of games as $H = \mathbb{R}^3 \setminus \{\vec{0}\}$. Since we assume **scale invariance** we can, without loss of important information, make every vector $\vec{v'} \in H$ unitary by dividing with the Eucledian norm as shown in equation 2.5.

$$\vec{v} = \begin{bmatrix} c_1 \\ c_2 \\ z_1 \end{bmatrix} \triangleq \frac{1}{|\vec{v'}|} \begin{bmatrix} c_1' \\ c_2' \\ z_1' \end{bmatrix} \tag{2.5}$$

Now every symmetric 2×2 game is represented as a vector $\vec{v} \in G$ where $G = \{[c_1, c_2, z_3]^t \in \mathbb{R}^3 : c_1^2 + c_2^2 + z_1^2 = 1\}$ is the three dimensional unit sphere.

The interesting properties of the decomposition of a symmetric game is the sign of c_1, c_2 and z_1 and their relative length. We are therefore interested in the size relationship between $|c_1|, |c_2|$ and $|z_1|$. In order to distinguish regions by the sign of the arrows we introduce the planes $c_1 = 0$, $c_2 = 0$ and $z_1 = 0$ which divide the sphere into 8 peices of equal size. Since they are planes through the origin, the intersection of the unit sphere and the planes are great circles. An illustration of these circles are presented below.



Since we are also interested in the relationships between the lengths $|c_1|$, $|c_2|$ and $|z_1|$ we define the planes $z_1 = \pm c_1$, $z_1 = \pm c_2$ and $c_1 = \pm c_2$. These planes are, like the previous planes, also planes through the origin. Therefore their intersection with the unit sphere are also great circles on the sphere.

Proposition 2.1.1. The planes $c_1 = 0, c_2 = 0, z_1 = 0, z_1 = \pm c_1, z_1 = \pm c_2$ and $c_1 = \pm c_2$ divides the sphere into 48 different regions.

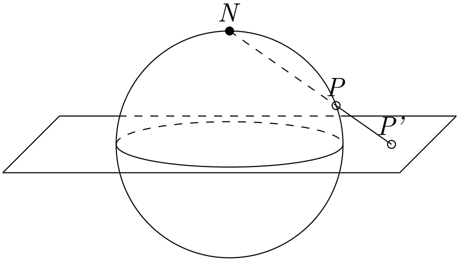
Proof. The first three planes defines 2 regions each and since they are orthogonal this results in $2^3 = 8$ regions in total. The lengths $|c_1|, |c_2|$ and $|z_1|$ can be ordered in $3! = 6$ ways independantly of the signs of c_1, c_2 and z_1. Hence the planes divide the sphere into $2^3 \times 3! = 48$ regions in total. □

Stereographic projection

The 48 regions of the sphere consists of games with different types of decomposition. These regions will be the basis of our proposed classification of symmetric 2×2 games. To get an overview of these regions we project the sphere onto the c_1c_2 plane using the stereographic projection. This is done by considering $[0, 0, 1]^t$ to be the north pole of the game sphere (that is the unit sphere) $G = \{[c_1, c_2, z_3]^t \in \mathbb{R}^3 : c_1^2 + c_2^2 + z_1^2 = 1\}$ and using by the stereographic projection

$$(c_1', c_2') = (\frac{c_1}{1 - z_1}, \frac{c_2}{1 - z_1})$$

where $z_1 \neq 1$, to project G onto the c_1c_2 plane. The game $[0, 0, 1]^t$ is projected to the point at infinity, that is the horizon of the plane.



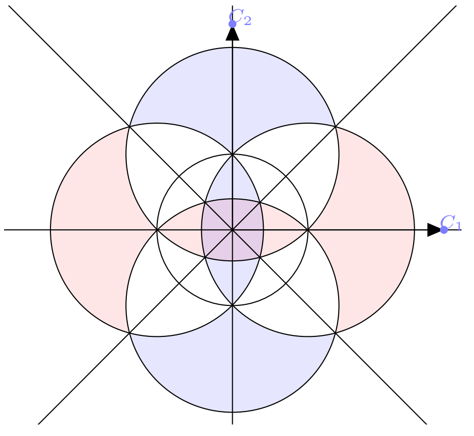
We know that circles on G that does not contain the north pole are projected onto circles in the plane and that great circles that does contain the north pole are projected onto lines through the origin in the plane under the stereographic projection. We also know that the projection is conformal, that is we know that it preserves angles and intersection points. This makes it simple to determine the image of the projection.(Coxeter and Greitzer, 1967)

The great circles defined by the planes $c_1 = 0$ and $c_2 = 0$ both contains the north pole and therefore they will be projected onto lines though the origin. The intersection of the great circle defined by $c_1 = 0$ and the c_1c_2 plane is $\{(0, -1), (0, 1)\}$ and these points are mapped to themselves under the projection and hence we know that the image of the circle is a line through the origin and through the point $(0, 1)$. Of course this line is the c_2-axis and in the same way it is easy to confirm that the circle defined by $c_2 = 0$ is mapped onto the c_1-axis.

Similarly the great circles defined by $c_1 = \pm c_2$ intersect each other at the north pole (and at the south pole) and they intersect the equator at $\{(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})\}$ and $\{(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})\}$ respectively. This means that the great circle defined by the plane $c_1 = c_2$ is mapped to the line $c_1' = c_2'$ and the great circle defined by the plane $c_1 = -c_2$ is mapped to the line $c_1' = -c_2'$ in the plane.

The points $[0, \pm 1, 0]^t$ are both on the circle defined by $z_1 = c_1$ and these are mapped to $(0, \pm 1)$ and since we know that this great circle is mapped onto a circle in the plane we only need to find the image of one more point of the great circle to determine the image circle. For example the point $[\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}]^t$ is mapped to $(\frac{1}{\sqrt{2}-1}, 0)$ and hence the circle in the plane is the circle with center $(1, 0)$ and radius $\sqrt{2}$. In the same way the great circles defined by $z_1 = -c_1$, $z_1 = c_2$ and $z_1 = -c_2$ are mapped onto the circles with radius $\sqrt{2}$ and centers $(-1, 0)$, $(0, 1)$ and $(0, -1)$ respectively.

Lastly the great circle defined by $z_1 = 0$ is the equator and is mapped onto itself, that is it is mapped onto the unit circle in the c_1c_2 plane. The resulting circles and lines are presented in the illustration below. Red colour means that $|z_1| \geq |c_2|$ and blue colour means that $|z_1| \geq |c_1|$. The white area outside of all of the circle should be seen as both red and blue, just like the middle square is both red and blue.



In the picture one can clearly see all 48 regions. It is clear in what regions $c_1 > 0$ and in what regions $c_2 > 0$ since the cartesian structure of the c_1c_2 plane immediately yeilds that $c_1 > 0$ in the first and fourth quadrant and $c_2 > 0$ in the first and the second quadrant. Because the stereographic projection projects all points of the sphere with $z_1 < 0$ to points inside the unit circle in the c_1c_2 plane, we can find all such point inside the middle circle in the image. These insights together with the meaning of the colours of the graph makes it easy to understand what decomposition properties each region defines.

So far we have distinguished between games based on their decomposition. However we do not wish to distinguish between games that are permutations of each other, that is, if a game can be constructed by permuting the rows or columns of another game's payoff matrix we consider them to be the same game. Because of the symmetry in the payoff matrix in a symmetric 2×2 game, each game has exactly two equivalent permutations and each of those games are constructed by permuting the rows and columns of the other game's payoff matrix. This permutation is shown in the matrices below. We use the symbol " $\sim$ " to write that two matrices are equivalent in the sense that we consider them to be strategically equivalent games. For example we consider the following two matrices equivalent.

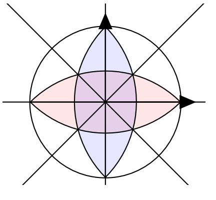
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \sim \begin{bmatrix} d & c \\ b & a \end{bmatrix}$$

These matrices have the following decomposition.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & \frac{b+c}{2} \\ \frac{b+c}{2} & d \end{bmatrix} + \begin{bmatrix} 0 & \frac{b-c}{2} \\ -\frac{b-c}{2} & 0 \end{bmatrix}$$

$$A' = \begin{bmatrix} d & c \\ b & a \end{bmatrix} = \begin{bmatrix} d & \frac{b+c}{2} \\ \frac{b+c}{2} & a \end{bmatrix} + \begin{bmatrix} 0 & -\frac{b-c}{2} \\ \frac{b-c}{2} & 0 \end{bmatrix}$$

That is $c_1 = -c_2'$, $c_2 = -c_1'$ and $z_1 = -z_1'$ and from this we can tell that the two equivalent permutations of a game will appear in exactly two regions in the game map presented above. In fact, since $z_1 = -z_1'$ one of the permutations will be found in a region outside the unit circle and the other permutation will be found in a region inside the unit circle. Since we do not wish to distinguish between these games, we can reduce the number of regions from 48 to $\frac{48}{2} = 24$ and consider the regions inside the unit circle to represent both themselves and their permutation equivalent region outside of the unit circle. A complete map of all of the regions and their properties can be found in the appendix and there we also present what pairs of regions are strategically equivalent. A map of the 24 non-equivalent regions are presented below and the colours have the same meaning as in the previous map.



Analysis of the Regions

In this section we will analyze the different regions obtained from our classification. See Appendix B for a map all the 48 regions in the projective plane and a table of what regions are strategically equivalent to each other, according to definition 1.2.29. See Figure 2.1 for a map of the regions analysed here.

From this analysis we concluded that all the interesting standard games have the common property that the conflict arrows points in the opposite direction as the common interest arrows. The arrows added in the figures in this thesis are the same as the arrows used by Rasmusen, (1989). Their purpose to illustrate the preferences of the players. Below is a figure showing a general decomposition of a 2×2 game.

$$\begin{array}{ccc}(a,x) & \longleftrightarrow & (b,y)\\ \updownarrow & & \updownarrow \\ (c,z) & \longleftrightarrow & (d,w)\end{array} = \begin{array}{ccc}(\frac{a+x}{2},\frac{a+x}{2}) & \longleftrightarrow & (\frac{b+y}{2},\frac{b+y}{2})\\ \updownarrow & & \updownarrow\\ (\frac{c+z}{2},\frac{c+z}{2}) & \longleftrightarrow & (\frac{d+w}{2},\frac{d+w}{2})\end{array} + \begin{array}{ccc}(\frac{a-x}{2},\frac{x-a}{2}) & \longleftrightarrow & (\frac{b-y}{2},\frac{y-b}{2})\\ \updownarrow & & \updownarrow\\ (\frac{c-z}{2},\frac{z-c}{2}) & \longleftrightarrow & (\frac{d-w}{2},\frac{w-d}{2})\end{array}$$

The map is divided into 3 different categories of regions, each category containing 4 different regions, depending on the strength of the conflict in relation to the common interest. In the middle of our map there are 4 different regions where $|z_1| > |c|$, i.e. the conflict is stronger than the common interest. We label these regions N_k, $k = 1, ..., 4$. Each of these regions are divided into 2 smaller sub-regions by the lines $c_1 = c_2$ and $-c_1 = c_2$. In Appendix A we call these regions Stronger Zero-Sum regions since the zero-sum part is stronger. The N_k regions contain the standard games Prisoner's Dilemma and Deadlock.

There are 4 regions where, in contrast to the N_k regions, $|z_1| < |c|$. This means that in these regions the conflict is weaker than the common interest. We label these regions K_k , $k = 1, ..., 4$. Every one of these regions are also divided into 2 smaller sub-regions by the $c_1 = c_2$ and $-c_1 = c_2$ lines. We call these regions Weaker Zero-Sum regions in Appendix A because the zero-sum part is weaker in these regions. The Weaker Zero-Sum regions contain the Civic Duty Game.

The 4 remaining regions are a mix of the N_k and K_k regions in the sense that $|z_1| > |c_i|$ and $|z_1| < |c_j|$, $i, j \in \{1, 2\}$, $i \neq j$. They are called Mixed Conflict regions in Appendix A. We label them T_k, $k = 1, ..., 4$. The T_k regions contain the standard games Chicken and Stag Hunt.

These three categories are divided into 8 regions each by the conditions $\pm c_1 \lesseqgtr \pm c_2$, determining which common interest arrow is the strongest, $c_1 \lesseqgtr 0$ and $c_2 \lesseqgtr 0$, determining in which direction the c_1 and c_2 arrows point. The remainder of this subsection will be devoted to analysis of the regions and their sub-regions within each of these categories. The focus will be on the regions that we have found interesting.

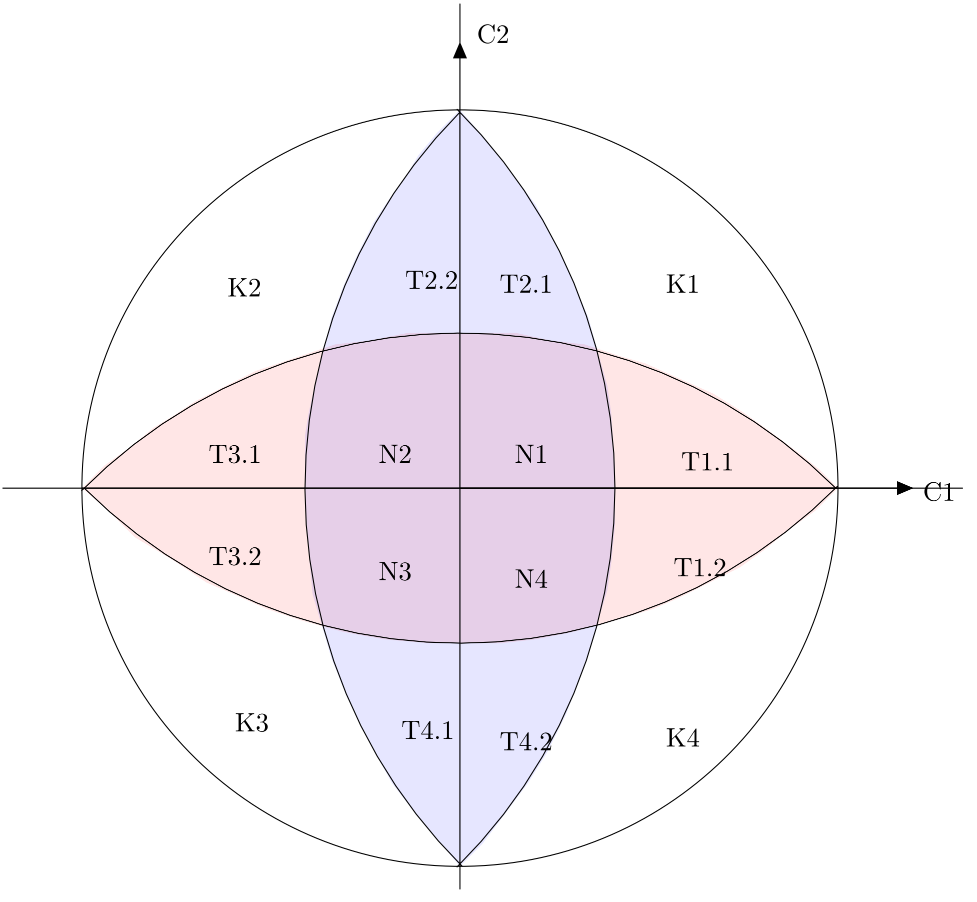


Figure 2.1: Map over Regions in the Projective Plane

Below are descriptions of all the 24 regions of the sphere where $z_1 < 0$, since these are the games inside the unit circle in the plane we project upon. Here are also only games with strict inequalities between the parameters included, i.e. we do not include games that lie on the boundary between regions. Each region is divided into sub-regions. Every game in a region has the same combination of Nash equilibria and Huertas optimality. The sub-regions are divided based on more subtle differences. Included information for each region is:

1. one example of a game representing that region with its conflict and common interest components,

2. the parameter values for the specific region and

3. a table of what outcomes are Nash equilibrium and Huertas optimal. The Nash equilibria are labelled N and Huertas optimalities are labelled H.

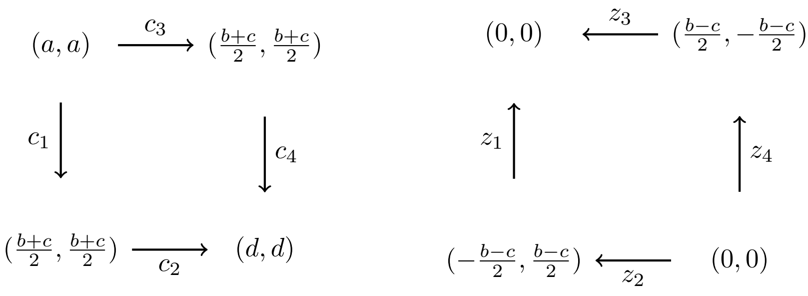
Obs: The a,b,c and d payoffs used in this section are ordered as in Table 2.1.

(a, a)	(b, c)
(c, b)	(d, d)

Table 2.1: Payoff Matrix

When c_1, c_2 and z_1 are all greater than zero, the arrows in the common interest and conflict part will point as shown in the figure below. Switching signs on an arrow means changing its direction. With the arrows defined this way the c_i arrows pull in the same direction as the conflict when they are positive and in the opposite direction when they are negative. Increasing an arrows absolute value means increasing its strength. Note that $c_1 = c_3$, $c_2 = c_4$ and $z_1 = z_2 = z_3 = z_4$. Expressed in payoff values we have that

$$\begin{cases} c_1 = \frac{b+c}{2} - a \\ c_2 = d - \frac{b+c}{2} \\ z_1 = \frac{b-c}{2} \end{cases} .$$



Stronger Zero-Sum

In these regions the conflict is stronger than the common interest. The region where the conflict arrows points in the opposite direction, the N_3 region, contains the Restricted Prisoner's Dilemma. In region N_1 the conflict pulls in the same direction as the common interest, resulting in the Deadlock game. These regions are further divided by the line $c_2 = c_1$, determining whether the c_1 or c_2 arrow is stronger. The remaining regions N_2 and N_4 are divided by the line $c_2 = -c_1$. The sub-regions where $c_2 < -c_1$ are non-restricted versions of the Prisoner's Dilemma and those where $c_2 < -c_1$ contain Deadlock games. In the figure below the decomposition of a Prisoner's Dilemma game from region N_3 is shown and a decomposition of a Deadlock game from region N_1.

All the regions where $c_2 < -c_1$ are versions the Prisoner's Dilemma. They all have the same ordinal payoffs, so in all of these games it is a dominant strategy to defect for both players. If the game is played only once, i.e. as a one-shot game, the games are no different and can be regarded as one single region. This is because the only thing that matters in this case is short term payoff and the best way to achieve this is by always cooperating.

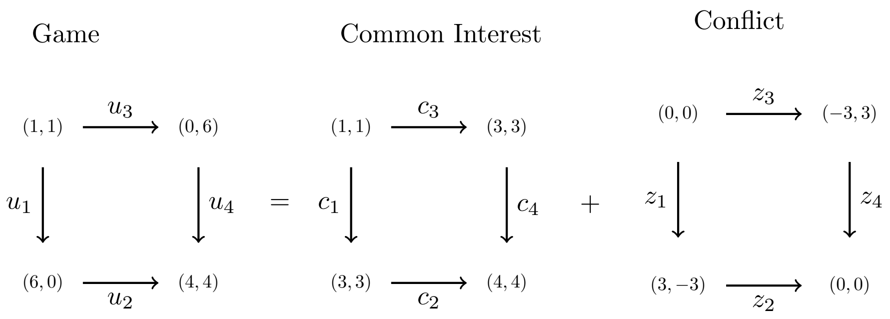
Deadlock is similar to Prisoner's Dilemma, except for that the conflict drags the players in the same direction as the common interest, reinforcing it instead of pulling them away from it. One could therefore argue that, even though they are similar in some respects, they are opposites of each other.

Region N1

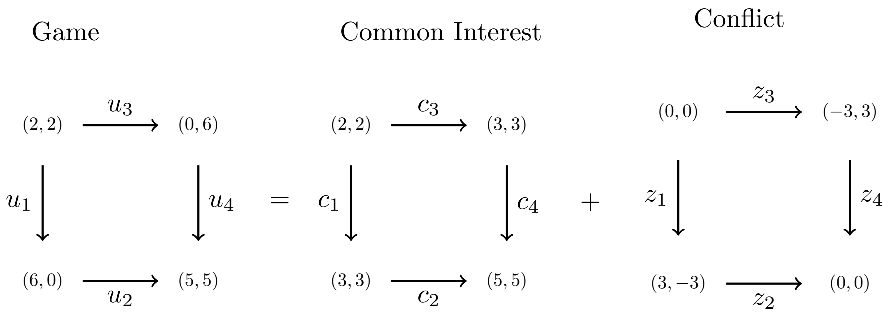
The N_1 region consists of consists of two regions containing Deadlock games. They are divided by whether c_1 or c_2 is greater than the other. Since both c_1 and c_2 are positive the conflict is pulling the players in the same direction as the common interest, so they are not affected by it in the same way as in the Prisoner's Dilemma.

$$\begin{cases} c_1 > 0 \\ c_2 > 0 \\ |z_1| > |c_1| \\ |z_1| > |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$



Sub-region $|c_1| < |c_2|$



H	
	N

Region N3

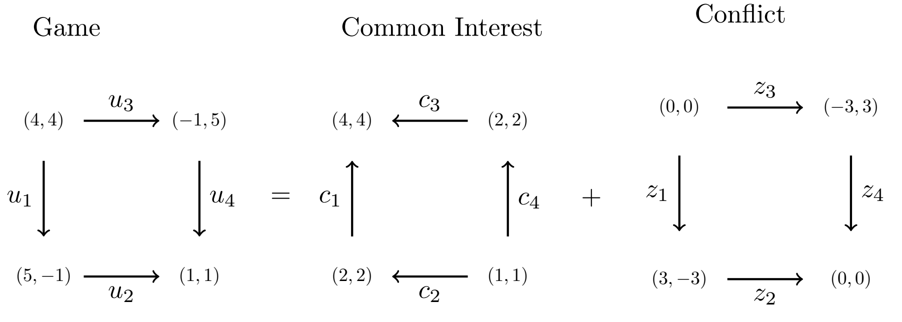
The N_3 region contains the restricted Prisoner's Dilemma, i.e. $2d < b + c < 2a$. In this region both c_1 and c_2 are negative, which means that the conflict pulls the players away from their common interest which results in the Prisoner's Dilemma. This region is divided in two smaller regions in the same way as the previous one, N_1.

$$\begin{cases} c_1 < 0 \\ c_2 < 0 \\ |z_1| > |c_1| \\ |z_1| > |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$

In this part of N_3 the c_1 arrow is stronger than the c_2 arrow. This results in the $d-b > c-a$ inequality which means that it costs more to signal for cooperation than to accept it.

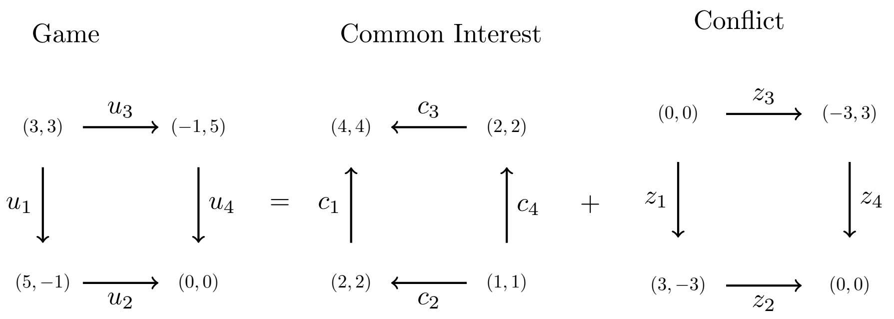
$$\begin{cases} 2d < b+c < 2a \\ c > a > d > b \\ d-b > c-a \end{cases}$$



Sub-region $|c_1| < |c_2|$

In this part of N_3 the c_2 arrow is stronger than the c_1 arrow. This results in the $d-b < c-a$ inequality which means that it costs less to signal for cooperation than to accept it.

$$\begin{cases} 2d < b+c < 2a \\ c > a > d > b \\ d-b < c-a \end{cases}$$



H	
	N

Region N2

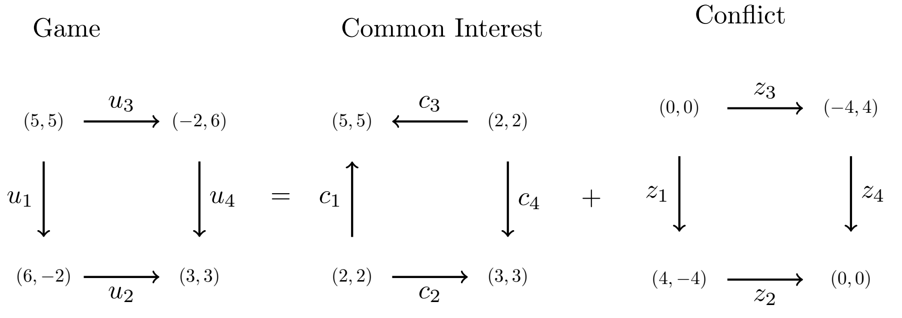
In this region c_2 is positive and c_1 is negative, so the conflict is partly working against the common interest and partly with it. Since c_1 is negative the conflict will work against this common interest, and because it is stronger, change its direction. The common interest represented by c_2 goes in the same direction as the conflict, so this will remain unaffected. Depending on whether $|c_1|$ or $|c_2|$ is greatest, the games in this region will be either Prisoner's Dilemma or Deadlock.

$$\begin{cases} c_1 < 0 \\ c_2 > 0 \\ |z_1| > |c_1| \\ |z_1| > |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$

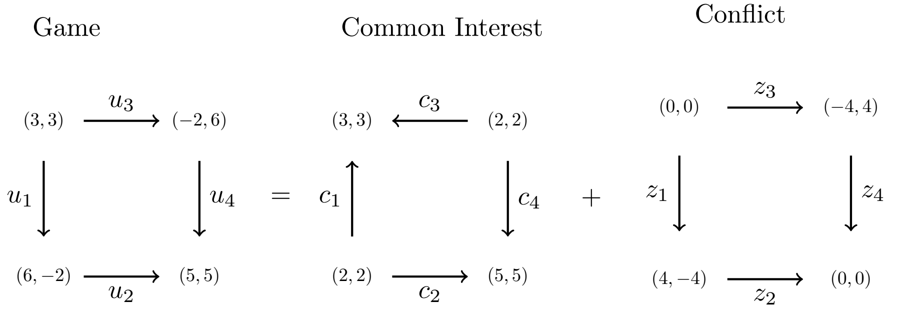
In this part of the region $|c_1| > |c_2|$. This means that the players common interest is stronger towards c_1 than c_2 and thus they jointly prefer to go towards c_1. But since the conflict drags the players in the opposite direction as the more preferred c_1 direction and in the same direction as c_2, they end up in a by both less preferred outcome. This results in a Prisoner's Dilemma.

$$\begin{cases} b + c < 2d \\ c > a > d > b \\ a - d < d - b \end{cases}$$



Sub-region $|c_1| < |c_2|$

In this part of N_2, unlike the previous part, c_1 is less than c_2, which means that the conflict is pulling the players towards their most preferred common interest. The result is a version of Deadlock.



H	
	N

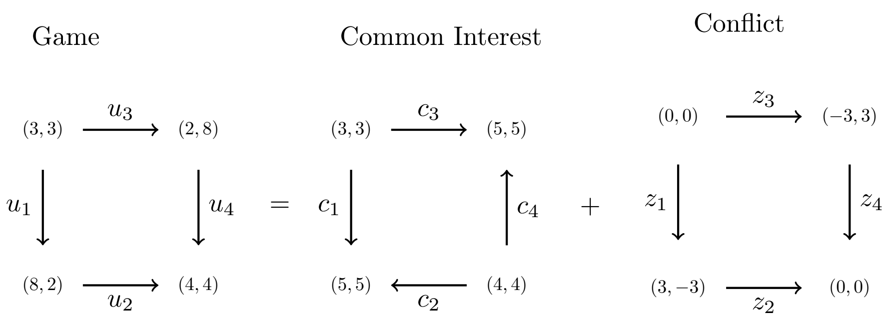
Region N4

This regions is similar to the previous one, N_2, the only difference being that c_1 and c_2 have changed sign. This will also be divided in two depending on the relation between $|c_1|$ and $|c_2|$ and result in Prisoner's Dilemma or Deadlock depending on this.

$$\begin{cases} c_1 > 0 \\ c_2 < 0 \\ |z_1| > |c_1| \\ |z_1| > |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$

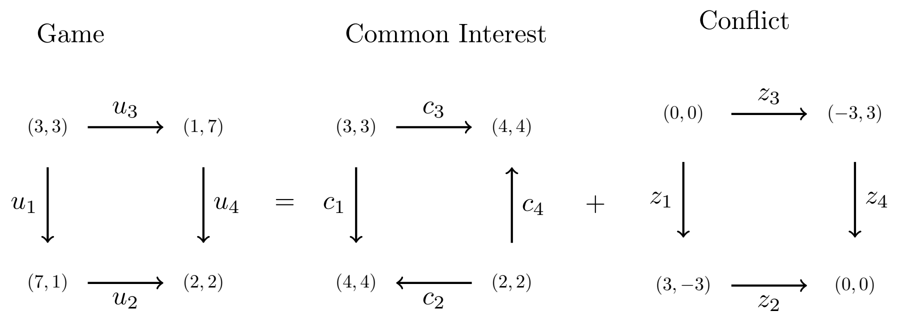
In this part of N_4 the common interest is stronger in the c_1 direction. Since c_1 is positive in this, the conflict does not interfere with the common interest. The result is Deadlock.



Sub-region $|c_1| < |c_2|$

This is the Prisoner's Dilemma part of N_4. The players common interest is towards the negative direction of c_2, but since the conflict is stronger and pulls the players in the opposite direction, the result is the less preferred c_1 direction.

$$\begin{cases} b + c > 2a \\ c > a > d > b \end{cases}$$



H	
	N

Weaker Zero-Sum

The K_k region contain games where the conflict is weaker than the common interest. These games are therefore mostly no-conflict games. The K_4 region is interesting because the common interest lies on the anti-diagonal, which the conflict directly affects. This results in the players getting different payoffs in the common interest outcomes. This results in the Civic Duty Game (Rasmusen, 1989), also called the Restricted Battle of the Sexes (Harris, 1969).

In the Civic Duty Game the conflict is not strong enough to move the players away from the common interest. It does however make it a game of *mixed conflict* in the sense that it is in both players common interest that they coordinate on playing different strategies, but depending on how they coordinate, one player will always be more satisfied than the other. So even though the conflict is not strong enough to break the Nash equilibra in the common interest game, it does add some tension to it. The K_4 region is further divided by the line $c_2 = -c_1$, determining whether the c_1 arrow is stronger than the c_2 arrow. As shown in the Comparative Literature Analysis section, this is an interesting distinction to make.

The other regions of this type are no-conflict regions where the conflict does not affect the common interest enough to make them interesting. The games in the regions K_3 and K_1 are no-conflict games with dominant strategies. This simply does not lead to any interesting analysis since the outcome is too "good". The games in the K_2 region have no dominant strategies, but the players receive the same payoff in the Nash equilibria so there is no real conflict between them. See Appendix A for a more detailed description of these games.

Region K1

In this region c_1 and c_2 are positive, so the common interest and the conflict are pulling in the same direction. Since the common interest is greater than the conflict it means that the players will receive the their highest payoff. Since c_1 and c_2 are pulling in the same direction, it does not matter which one is stronger. The resulting games are therefore games of no-conflict.

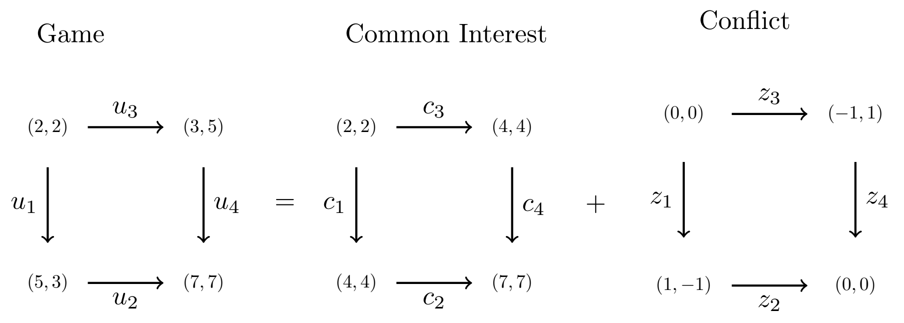
$$\begin{cases} c_1 > 0 \\ c_2 > 0 \\ |z_1| < |c_1| \\ |z_1| < |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$

Game		Common Interest		Conflict

$$\begin{array}{ccc} (1,1) & \xrightarrow{u_3} & (3,5) \\ u_1 \big\downarrow & & \big\downarrow u_4 \\ (5,3) & \xrightarrow[u_2]{} & (6,6) \end{array} \quad = \quad \begin{array}{ccc} (1,1) & \xrightarrow{c_3} & (4,4) \\ c_1 \big\downarrow & & \big\downarrow c_4 \\ (4,4) & \xrightarrow[c_2]{} & (6,6) \end{array} \quad + \quad \begin{array}{ccc} (0,0) & \xrightarrow{z_3} & (-1,1) \\ z_1 \big\downarrow & & \big\downarrow z_4 \\ (1,-1) & \xrightarrow[z_2]{} & (0,0) \end{array}$$

Sub-region $|c_1| < |c_2|$



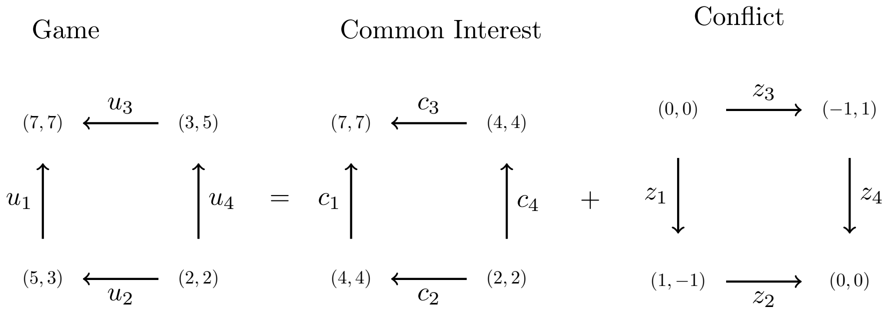
	NH

Region K3

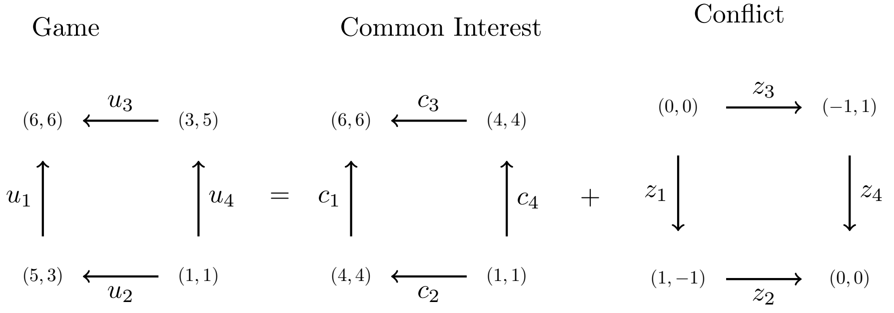
This region contains the same kind of games as those found in K_1, in the sense that the common interest pulls the players to one unique outcome and the conflict is too weak to pull them away from it. In this case the conflict is working against the common interest, but is so weak that it has no interesting difference from the games contained in region K_1.

$$\begin{cases} c_1 < 0 \\ c_2 < 0 \\ |z_1| < |c_1| \\ |z_1| < |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$



Sub-region $|c_1| < |c_2|$



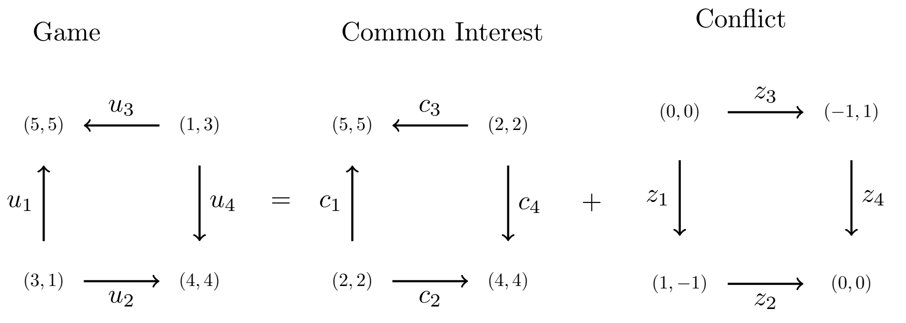
NH	

Region K2

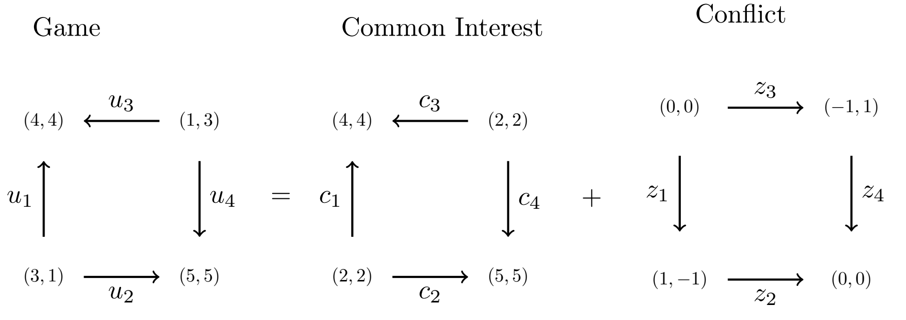
This region differs from K_1 and K_3 in that is has two diagonal Nash equilibria. This is because c_1 and c_2 have different signs. Since both players receive the same payoffs in the diagonal, and the diagonal contains their highest payoffs, this makes these games no-conflict games which we find are of less interest.

$$
\begin{cases}
c_1 < 0 \\
c_2 > 0 \\
|z_1| < |c_1| \\
|z_1| < |c_2|
\end{cases}
$$

Sub-region $|c_1| > |c_2|$



Sub-region $|c_1| < |c_2|$



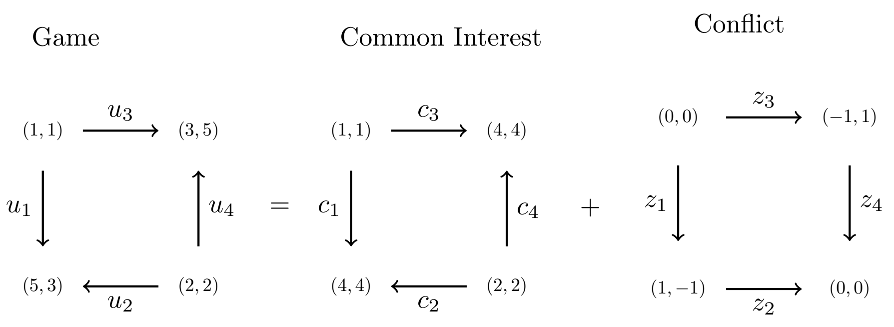
NH	
	NH

Region K4

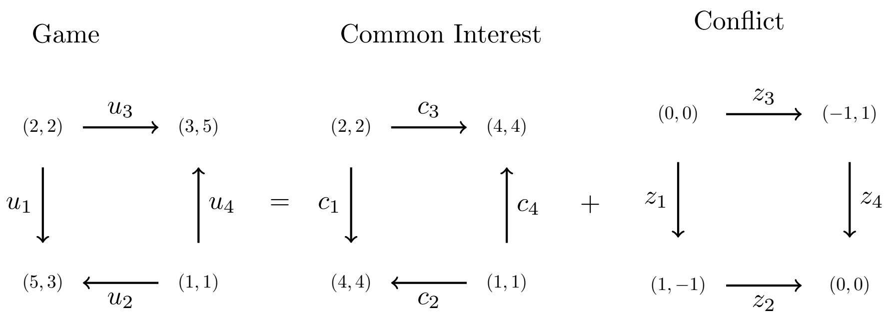
The games contained in this region are similar to the ones in K_2 in the sense that the common interest arrows point the players to two different outcomes, meaning that the resulting game has two Nash equilibria. The difference is that the common interest pulls the players to the anti-diagonal outcomes. Because the game is symmetric, the conflict only affects the payoffs in the anti-diagonal outcomes. Therefore it has an impact on the common interest even though it is too weak to completely change the direction of the arrows. This results in two different versions of the Civic Duty Game (Rasmusen, 1989), also referred to as the symmetric or the Restricted Battle of the Sexes by Harris.

$$\begin{cases} c_1 > 0 \\ c_2 < 0 \\ |z_1| < |c_1| \\ |z_1| < |c_2| \end{cases}$$

Sub-region $|c_1| > |c_2|$



Sub-region $|c_1| < |c_2|$



	NH
NH	

Mixed Strength

The T_k regions are regions of Mixed Conflict, meaning the conflict is stronger than one of the common interest arrows but weaker than the other. The T_3 region contains Stag Hunt, the decomposition is shown in the figure below. The conflict in this regions is weaker than the c_1 arrow but stronger than the c_2 arrow. This region is further divided based on if c_1 is greater or less than zero. The T_4 region contains two versions of Chicken, divided by if c_2 is positive or negative.

The T_1 and T_2 regions appears to be of less interest than the T_3 and T_4 regions. The games in region T_1 are similar to Deadlock. Both players have dominant strategies and the payoffs they receive are the second highest. In the T_2 region the games also have a unique Nash equilibria and this outcome Pareto dominates all other outcomes, which make them less interesting for further analysis.

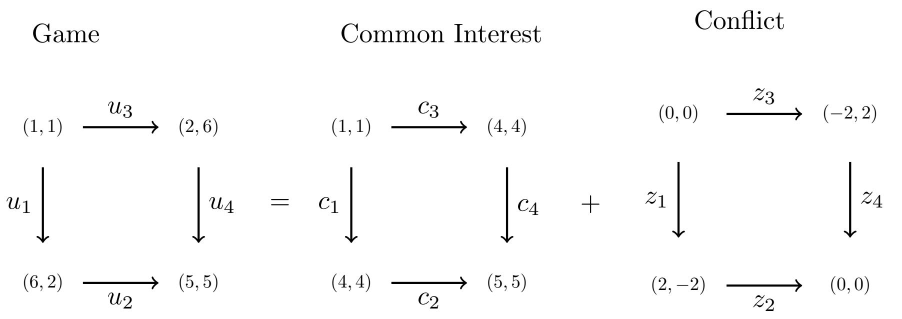
Region T1

In this region c_1 is always pulling in the same direction as the conflict, i.e. z_1. The $|c_2| < |z_1| < |c_1|$ which defines this region means that z_1 is stronger than the c_2 arrow and weaker than the c_1 arrow. This region is split into two parts depending on if c_2 is positive or negative.

$$\begin{cases} c_1 > 0 \\ |z_1| < |c_1| \\ |z_1| > |c_2| \end{cases}$$

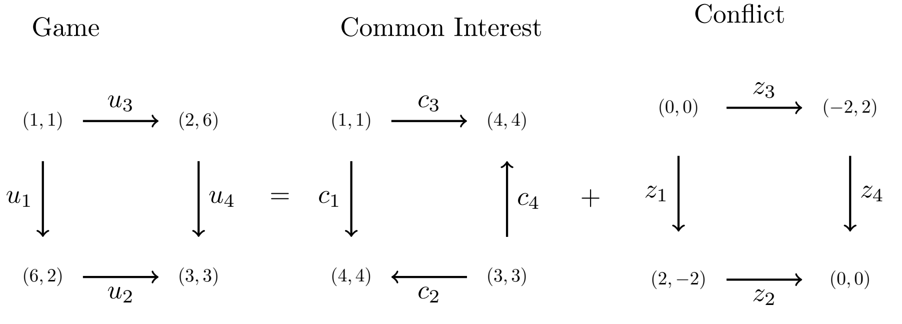
Sub-region $c_2 > 0$

Here $c_2 > 0$, so the c_2 arrow points in the same direction as the conflict. This means that the conflict has no important impact on the common interest.



Sub-region $c_2 < 0$

Here $c_2 < 0$, so the c_2 arrow points in the opposite direction as the conflict. In this case the conflict changes the direction of the c_2 arrow, which means that it has higher impact than in the former game, even though the resulting game is similar. The games in this region has the property that $b + c > 2d$ which is not fulfilled in the part where $c_2 > 0$. The players might, in iterated play, choose to cooperate by alternating between the 01 and 10 outcomes.



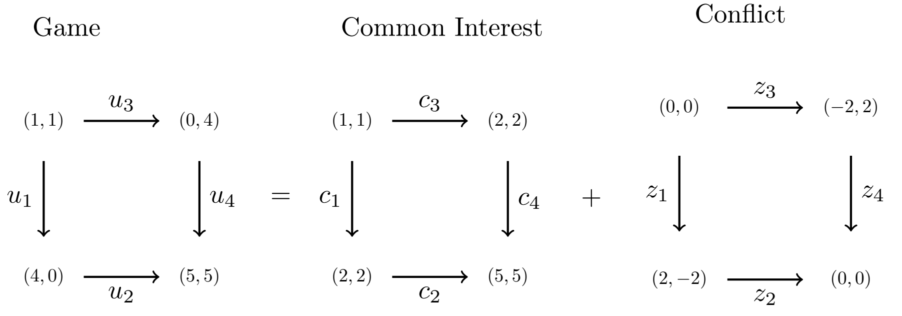
	H
H	N

Region T2

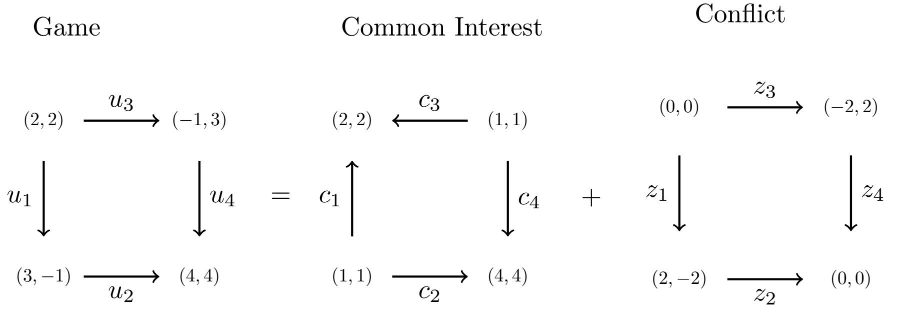
This region is similar to the T_1 region with the difference that c_1 and c_2 parameters are "swapped" in the inequalities.

$$\begin{cases} c_2 > 0 \\ |z_1| > |c_1| \\ |z_1| < |c_2| \end{cases}$$

Sub-region $c_1 > 0$



Sub-region $c_1 < 0$



H	
	NH

Region T3

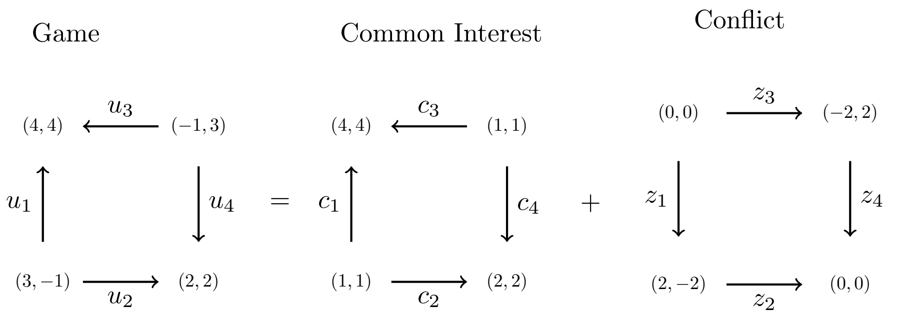
This region contains two versions of Stag Hunt, one where $c_2 > 0$ and one where $c_2 < 0$. The conflict is stronger than the c_2 arrow and weaker than the c_1 arrow.

The conflict works in the opposite direction as c_1 in this region.

$$\begin{cases} c_1 < 0 \\ |z_1| < |c_1| \\ |z_1| > |c_2| \end{cases}$$

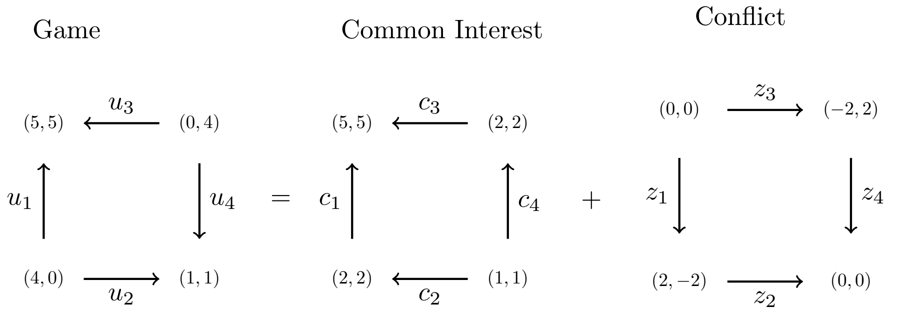
Sub-region $c_2 > 0$

In this sub-region c_2 is positive, which means that the arrow points in the same direction as the conflict arrow. c_1 and c_2 have different signs so they point in different directions. The common interest is both directed towards hunting dear and stag, where stag is more preferred. The conflict, however, ads the risk of pursuing the common interest and the temptation to hunt hare because its a lower risk alternative.



Sub-region with $c_2 < 0$

Here c_2 is negative, so the c_2 arrow points in the same direction as the c_1 arrow and in the opposite direction as the conflict arrow. In this game it is actually never a common interest to hunt hare, but the conflict drags the players toward it. Since hunting hare is a much less attractive strategy in this game the risk of hunting stag is reduced.



NH	
	N

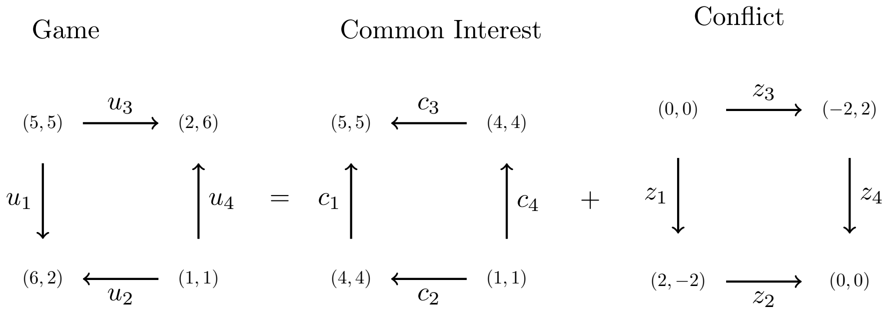
Region T4, Chicken

This region contains the Chicken game. It is similar to T_3, containing Stag Hunt, but here c_1 and c_2 have changed place in the inequalities. The conflict is stronger than the c_1 arrow and weaker than the c_2 arrow.

$$\begin{cases} c_2 < 0 \\ |z_1| > |c_1| \\ |z_1| < |c_2| \end{cases}$$

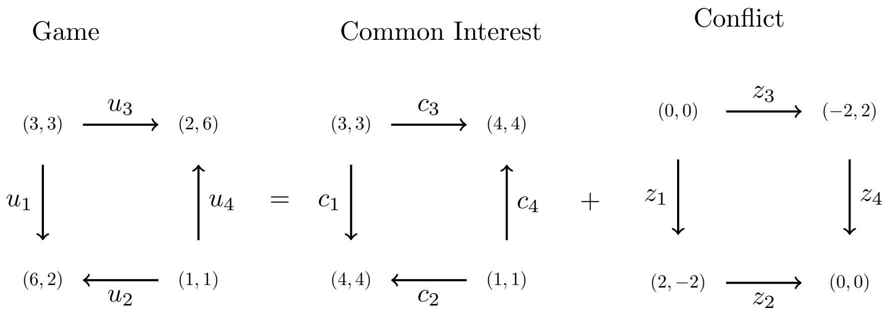
Sub-region $c_1 < 0$

The games in this sub-region are the so called restricted Chicken (Harris, 1969), (Ells and Sermat, 1966) satisfying the inequality $b + c < 2a$. Here the common interest is mainly to avoid the worst outcome, towards which the conflict is pulling the players. Since the conflict is stronger than c_1 the conflict pulls the players away from their most desired outcome. But since the conflict is weaker than c_2 it is not strong enough to pull them to the worst outcome, only closer to it. This also adds the risk of going for the less cooperative outcome, because if both players do this they receive their worst possible payoff.



Sub-region $c_1 > 0$

In this sub-region the common interest arrow c_1 is positive so it pulls in the same direction as the conflict. The common interest therefore pulls the player towards the anti-diagonal. Because of the inequality $b + c > 2a$ cooperation is no longer has the same meaning as in the restricted chicken. In iterated play it is better for the players to alternate between the 01 and 10 outcomes.



H	N
N	

Comparison Between the Regions

All the decomposed games shown above, except for the Civic Duty Game, have a unique common interest. But the conflict works in different ways to make it better for the players to betray one another in order to improve their own payoff.

Prisoner's Dilemma and Stag Hunt have the same kind of common interest component with a unique most favorable outcome. In both cases the conflict works against the common interest. The conflict in Stag Hunt is however weaker than in Prisoner's Dilemma. In Prisoner's Dilemma, the conflict is strong enough to make it a dominant strategy for the players to move away from the common interest. In Stag Hunt, the common interest is too strong for the conflict to move them away from hunting stag, but the worst common interest outcome which is both players hunting hare, becomes a Nash equilibrium. The conflict thus ads the risk of failure for cooperating on the most desirable outcome.

Chicken has simililar common interest as Stag Hunt and Prisoner's Dilemma, but here the players want to avoid the worst outcome at all costs. The conflict does however pull them away from their common interest, breaking to unique common interest equilibrium. The conflict is not strong enough to make the worst outcome a Nash equilibrium, but it pulls them towards it adding the risk if ending up in the worst outcome if both players play too greedily.

Conclusions

The decomposition of the standard game suggests that what makes these games interesting is the interplay between the common interest and conflict components of the game. Because all of the standard games have the common property that the conflict works against the common interest, the hypotheses that how interesting a game is depends on the combination of different types of conflict and common interest. Since all of the standard games are divided into different regions this does indicate that the properties that make a symmetric 2×2 game interesting can be explained by decomposing it and analysing how the conflict affects the common interest.

Nash and Huertas Optimality

A natural question that arises is if the decomposition model captures all interesting combinations of NE and HO. To investigate this question we need to express the NE and HO conditions in the variables c_1, c_2 and z_1. In equation 2.5

a game vector $\vec{v} \in G$ is defined as the vector

$$\vec{v} = \frac{1}{|\vec{v'}|}\vec{v'}, \text{ where } \vec{v'} = A\vec{y} \text{ and } \vec{y} \text{ is some vector} \\ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, a,b,c,d \in \mathbb{R} : a+b+c+d = 0. \tag{2.6}$$

A is an invertible matrix with

$$A^{-1} = \frac{1}{4}\begin{bmatrix} 2 & -3 & -1 & 0 \\ 2 & 1 & -1 & 4 \\ 2 & 1 & -1 & -4 \\ 2 & 1 & 3 & 0 \end{bmatrix} \tag{2.7}$$

and hence we can write

$$\vec{v} = \frac{1}{|\vec{v'}|}\vec{v'} = \frac{1}{|\vec{v'}|}A\vec{y} \Leftrightarrow |\vec{v'}|A^{-1}\vec{v} = \vec{y}. \tag{2.8}$$

In section 2.6 the conditions on the payoffs a, b, c and d for outcome ij, $i, j \in \{0, 1\}$ to be NE are stated. Equation 2.8 makes it possible to express these conditions in the variables c_1, c_2 and z_1 as stated in proposition 2.3.1.

Proposition 2.3.1. Given a game $\vec{v} \in G$ then

(i) 00 is NE iff $z_1 \geq c_1$,

(ii) 01 and 10 are NE iff $z_1 \leq c_1$ and $z_1 \geq c_2$,

(iii) 11 is NE iff $z_1 \leq c_2$.

Proof. Let $\vec{v}$ be an arbitrary game in G. Equation 2.8 gives us that the payoff vector $\vec{y} = \begin{bmatrix} a & b & c & d \end{bmatrix}^t$ associated with $\vec{v}$ can be written

$$\vec{y} = \alpha A^{-1}\vec{v} \text{ for some } \alpha \in \mathbb{R} : \alpha > 0$$

and hence

$$a = \frac{\alpha}{4}(2x - 3c_1 - c_2), b = \frac{\alpha}{4}(2x + c_1 - c_2 + 4z_1), \\ c = \frac{\alpha}{4}(2x + c_1 - c_2 - 4z_1) \text{ and } d = \frac{\alpha}{4}(2x + c_1 + 3c_2). \tag{2.9}$$

Section 2.6 provides the NE conditions expressed in a, b, c and d and using equation 2.9 we can express these conditions in the variables c_1, c_2 and z_1.

(i) 00 is NE iff $a \geq c \Leftrightarrow \frac{\alpha}{4}(2x - 3c_1 - c_2) \geq \frac{\alpha}{4}(2x + c_1 - c_2 - 4z_1) \Leftrightarrow z_1 \geq c_1$.

(ii) 01 and 10 are NE iff $b \geq d$ and $c \geq a$. $b \geq d \Leftrightarrow \frac{\alpha}{4}(2x + c_1 - c_2 + 4z_1) \geq \frac{\alpha}{4}(2x + c_1 + 3c_2) \Leftrightarrow z_1 \geq c_2$ and $c \geq a \Leftrightarrow z_1 \leq c_1$.

(iii) 11 is NE iff $d \geq b \Leftrightarrow z_1 \leq c_2$.

□

Proposition 2.3.1 implies that the planes $z_1 - c_1 = 0$ and $z_1 - c_2 = 0$ divides the game sphere G into four regions with different types of NE. Since both of these planes are used in our model to describe the relative strengths between $|z_1|$ $|c_1|$ and $|c_2|$, every region in our model has a specific type of NE. This means that our classification captures the concept of NE which supports our theory that all interesting aspects of a game is defined by interactions in its decomposition.

With proposition 2.3.1 we can draw the conclusion that we capture all types of NE that Huertas describes in his model. In fact our model also describes all different types of HO that Huertas uses as a base for his classification. This is stated in proposition 2.3.2.

Proposition 2.3.2. Given a game $\vec{v} \in G$ then

(i) 00 is HO iff $z_1 + c_1 \leq 0$,

(ii) 01 and 10 are HO iff $z_1 + c_1 \geq 0$ and $z_1 + c_2 \leq 0$,

(iii) 11 is HO iff $z_1 + c_2 \geq 0$.

Proof. For an arbitrary game vector $\vec{v} \in G$ we can write the payoffs a , b, c and d associated with $\vec{v}$ as in 2.9. In a 2×2 symmetric game HO is equivalent to NE in the transposed payoff matrix (Huertas-Rosero, 2003). This means that 00 is HO iff $a \geq b$, 01 and 10 are HO iff $b \geq a$ and $c \geq d$ and 11 is HO iff $d \geq c$. By using 2.9 these conditions can be expressed in the variables c_1, c_2 and z_1 as follows.

(i) 00 is HO iff $a \geq b \Leftrightarrow \frac{\alpha}{4}(2x - 3c_1 - c_2) \geq \frac{\alpha}{4}(2x + c_1 - c_2 + 4z_1) \Leftrightarrow z_1 + c_1 \leq 0$.

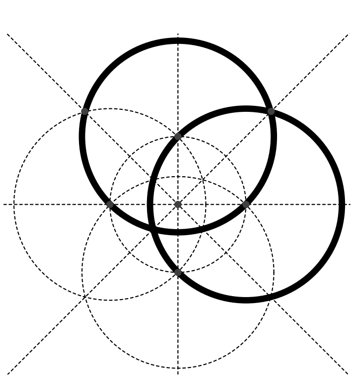
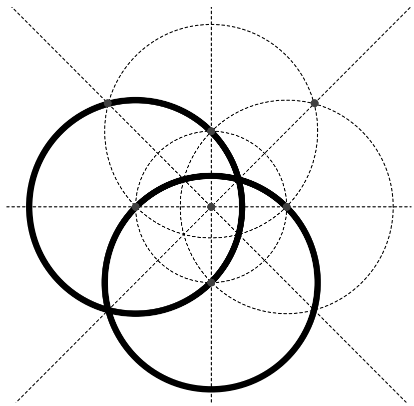
(ii) 01 and 10 is HO iff $b \geq a$ and $c \geq d$. $b \geq a \Leftrightarrow z_1 + c_1 \geq 0$ and $c \geq d \Leftrightarrow \frac{\alpha}{4}(2x + c_1 - c_2 - 4z_1) \geq \frac{\alpha}{4}(2x + c_1 + 3c_2) \Leftrightarrow z_1 + c_2 \leq 0$.

(iii) 11 is HO iff $d \geq c \Leftrightarrow z_1 + c_2 \geq 0$.

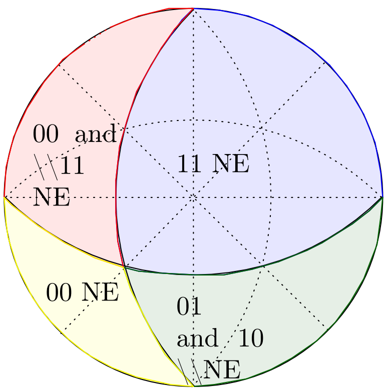
□

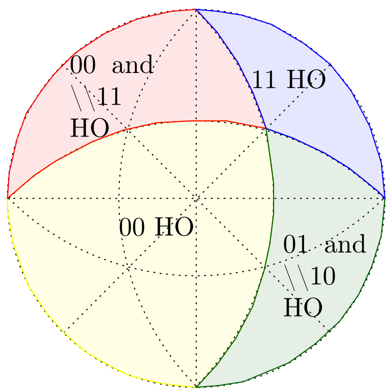
Proposition 2.3.2 states that all four types of HO are described by the planes $z_1 + c_1 = 0$ and $z_1 + c_2 = 0$. Just as the NE planes these planes divides the game sphere into four different pieces with different types of HO and the planes are used in our model for describing the relative relationship between $|z_1|$, $|c_1|$ and $|c_2|$. This means that our classification model captures all varieties of HO that Huertas captures in his model.

The two NE planes $z_1 - c_1 = 0$ and $z_1 - c_2 = 0$ and the two HO planes $z_1 + c_1 = 0$ and $z_1 + c_2 = 0$ are planes through the origin and non of them contains the north pole and hence the stereographic projection maps them onto circles in the plane. The circles obtained from the NE planes are the two bold circles in the left picture below and the two bold circles in the right picture are the circles obtained from the HO planes.

The pictures below illustrates the regions described in proposition 2.3.1 and in proposition 2.3.2.



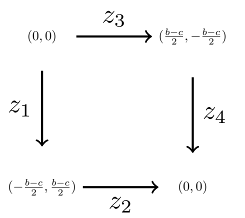


One can clearly see what combinations of NE and HO that are possible in the two illustrations above. These combinations lay the foundation for the classification proposed by Huertas.

The fact that our method of classification, which is intended to divide games into groups depending on their type of decomposition, captures all different types of NE and HO is indeed interesting. It supports the idea that the different properties that makes a game interesting and that separates intuitively different types of games can be found in the decomposition of the games. It seems natural that the type of decomposition of a game determines the NE and HO of the game since the preference relations used to define the variables c_1, c_2 and z_1 are closely related to the concept of NE and HO.

To see the intimate relationship between our variables and for example the type of NE in a region we can consider the following example which illustrates how the sign of c_1, c_2 and z_1 in combination with their size relationship determines the type of NE.

Example 2.3.3. Because of the way we defined the three variables the conflict variable z_1 will always have the following structure within the unit circle due to the fact that $z_1 \leq 0$.

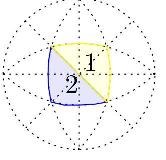


This means that if both c_1 and c_2 are positive the cooperation arrows will point in the same direction as the conflict arrows. Hence the conflict will only strengthen the cooperation arrows that points at 11. This explains why the hole region $c_1, c_2 \geq 0$ has 11 as its only NE.

However there are a few other areas where 11 is NE even though c_1 or c_2 are smaller than zero. This can be explained by the fact that in these regions z_1 is strong enough to shift the effect of the other variables and make 11 NE. Remember that z_1 always pulls towards 11 in the unit circle and if it is strong enough (the value of z_1 is greater than the value of the variables with opposite effect) it will counter effect the other variables and make 11 NE.

On the other hand in the case where $c_1, c_2 \leq 0$ and $|z_1| < |c_1|, |c_2|$ the effect of z_1 will be totally dominated by the other two variables that both pull towards 00 and this makes 00 NE in these regions.

So different combinations of NE and HO can be explained by the sign of the variables and the relative strengths between z_1 and c_1 and between z_1 and c_2. But there are two more lines in the picture that have not yet been explained. The line defined by the plane $c_1 = -c_2$ determines whether $d \geq a$ since $c_1 + c_2 \geq 0 \Leftrightarrow d - a \geq 0$. This is interesting to know since it states if the players prefers a over d or d over a. Consider for example the the two regions 1 and 2 in the figure below.



Both regions have the same NE (11) and the same HO (00) but in region 1 $c_1 \geq -c_2$ and hence $d \geq a$ while $c_1 \leq c_2 \Leftrightarrow d \leq a$ in region 2. This means that in region 1 the players prefers the NE outcome over the HO outcome and in region 2 the players prefer the HO outcome over the NE outcome. In fact, region two is the famous game *Prisoner's dilemma* wile region 1 is the game *Deadlock* and the significant difference between these games is defined in our model by the $c_1 + c_2 = 0$ plane. So this plane can explain some of the ad hoc divisions of classes that Huertas introduces to separate games where the NE outcome is greater than the HO outcome and vice versa.

The essence of the example above is to show that the plane $c_1 + c_2 = 0$ divides the space of games into interesting different classes since it states whether or not the players prefers 00 over 11 and in many situations, for example in the case of prisoner's dilemma and deadlock, this makes the difference between a trivial game (Deadlock) and an interesting game (Prisoner's Dilemma).

The planes in our model gives us a way to determine whether or not the NE outcomes is greater than the HO payoffs or not in any region. If there are two NE outcomes or two HO outcomes then we look at the average payoff. We use the notation $p(NE)$ and $p(HO)$ to denote the average payoff in the NE and HO positions respectively.

Theorem 2.3.4. For every of the 24 classes the in our model the lines and circles in the map determines if $p(NE) \geq p(HO)$.

Proof. There are eight cases to consider.
11 unique NE

If 11 is a unique NE we can have all four types of HO.

Case 1: 11 *is a unique HO.*
This is a trivial case since 11 is the only NE and the only HO and hence $p(NE) \geq p(HO)$.

Case 2: 00 *is a unique HO.*
Of course $p(NE) \geq p(HO) \Leftrightarrow d \geq a \Leftrightarrow c_1 + c_2 \geq 0$.

Case 3: 00 *and* 11 *are HO.*
We know that $p(HO) = \frac{a+d}{2}$ and if $d < a$ then $p(PO) > p(NE)$ and if $d \geq a$ then $p(NE) \geq p(HO)$ and hence $p(NE) \geq p(HO) \Leftrightarrow c_1 + c_2 \geq 0$.

Case 4: 01 *and* 10 *are HO.*
Here we only need to observe that 00 and 11 are HO means that $p(HO) = \frac{b+c}{2}$ and that $c_2 \geq 0 \Leftrightarrow -c - d + 2d \geq 0 \Leftrightarrow d \geq \frac{c+d}{2} \Leftrightarrow p(NE) \geq p(HO)$. That is $p(NE) \geq p(HO)$ if and only if $c_2 \geq 0$.

00 unique NE

As the map of the 24 regions suggest if 00 is a unique NE then 00 is a unique HO and of course this is a trivial case where $p(NE) \geq p(HO)$.

00 and 11 are NE

There are two possible NE-HO combinations in this case.

Case 1: 00 *and* 11 *are HO.*
This is again a trivial case since it is always true that $p(NE) \geq p(HO)$.

Case 2: 00 *is a unique HO.*
Here $p(NE) = \frac{a+d}{2}$ and $p(HO) = a$. If $a \leq d$ then $a = \frac{a+a}{2} \leq \frac{a+d}{2}$ while if $a > d$ then $a = \frac{a+a}{2} > \frac{a+d}{2}$. That is $p(NE) \geq p(HO)$ if and only if $a \leq d \Leftrightarrow c_1+c_2 \geq 0$.

01 and 10 are NE

Again there are two possible combinations of NE and HO that satisfy the condition that both 01 and 10 are NE.

Case 1: 01 *and* 10 *are HO*
In this case it is trival to realize that $p(NE) \geq p(HO)$.

Case 2: 00 *is a unique HO.*
Here $p(NE) = \frac{b+c}{2}$ and $p(HO) = a$. It is easy to see that $c_1 \geq 0 \Leftrightarrow p(NE) \geq p(HO)$ since $c_1 \geq 0 \Leftrightarrow -2a + b + c \geq 0 \Leftrightarrow \frac{b+c}{2} \geq a$.

□

With this proof we have shown that all the aspects of the classification provided by Huertas are found in our classification as well. This means that even though we intended to classify games only based upon their decomposition into cooperative and zero-sum games, we still captured all properties that Huertas-Rosero used to define his classification.

We made an important discovery about the classification provided by Huertas-Rosero, (2004). He claims (Huertas-Rosero, 2004, p.59-60) that he divides the space of games into 12 classes. It turns out however that one of his classes is empty. To be specific the class with two diagonal NE and a unique diagonal HO with a payoff that is less than the other diagonal payoff is empty. To realize this, consider the game with the payoff matrix

$$\begin{bmatrix} (a,a) & (b,c) \\ (c,b) & (d,d) \end{bmatrix}.$$

Suppose that 00 and 11 is NE and suppose without loss of generality that 00 is a unique HO. We know that 00 is NE if and only if $a \geq c$ and 11 is NE if and only if $d \geq b$. Furthermore we know that 00 is HO if and only if $a \geq b$.

Suppose now that the game considered belongs to the class where the HO outcome yields a lower payoff than the other diagonal outcome, i.e. $a \leq d$. This means that $d \geq a \geq b, c \Rightarrow d \geq c$. But $d \geq c$ makes 11 HO which is a contradiction since 00 is a unique HO. This proves that one of the twelve classes that Huertas-Rosero defines is empty and hence does not contain $\frac{1}{24}$ of all games as stated in the article.

Experimental Analysis

In this section we present an experiment designed to find the best strategies in a symmetric 2×2 game, inspired by the computer tournaments held by Axelrod and Hamilton, (1981), followed by a analysis of the results of this experiment. The motivation behind this experiment is to generate empirical data that can be used to analyze the differences between the optimal strategies in different classes in our model. We limit ourselves to deterministic strategies with memory length one (definition 2.4.1) due to practical limitations such as time and computing power.

Experiment design

We will use an evolutionary approach when constructing our experiment. In order to determine the fittest strategy of a given game, we first let every strategy face every other strategy a given number of times. The strategies then get to reproduce depending on how high their total score was after all the fights have been played out. The strategies with higher score will make up a larger fraction of the succeeding generation of strategies. This process is then repeated with the new population made up of the descendants of the former population. Every strategy is represented as a vector of length 5, as in example 2.4.2, determining what action to take depending on what actions were played in the previous iteration and what action to take in the first iteration.We also include the possibility of *mutation*. When the strategies reproduce we add a probability that one of actions in the strategy vector changes. This is to add the possibility of an established population to be invaded by a mutated strategy. A population consisting of only one strategy that initially out-conquered the other strategies might be weak to a certain type of strategy invading it as a result of mutation for example.

We repeat the process described above a given number of times for different games in our classification and analyse what strategies are being played. A formal description of the algorithm used to design the experiment is presented below.

In definition 2.4.1 below a formal definition of what we mean with *deterministic strategy memory one* is presented.

Definition 2.4.1. Let $G = (S, U, P)$ be a symmetric 2×2 game. A *strategy in iteration* n $s_{i,n} \in S_i$ is a strategy of player $i \in P$ in the n:th iteration of the game, $n \in \mathbb{N}$. We say that $s_{i,n}$ is *deterministic with memory length one* if $s_{i,0} = a \in A_i$ and $s_{i,n} = f_i(s_{i,n-1}, s_{-i,n-1})$ for $n > 0$, where $f_i : A_i \times A_{-i} \to A_i$.

Definition 2.4.1 states that a strategy is deterministic with memory length one if it has a deterministic starting action and if the action in the n:th iteration, $n > 0$, of an iterated game is uniquely determined by the outcome of the $n - 1$

iteration of the game. An example of a deterministic strategy of memory length one is presented in example 2.4.2.

Example 2.4.2. In a symmetric 2×2 game an example of a deterministic strategy of length one is presented as in the table below.

00	01	10	11	Start
0	1	0	1	0

This strategy states that the player starts by playing action 0. If the outcome is 00 the action in the first iteration will be 0 again but if the outcome is 01 the action in the first iteration is 1. This strategy is in fact designed to imitate the last move of the opponent in every iteration. This strategy is referred to as *Tit for Tat* (Axelrod and Hamilton, 1981).

Since $|A_i| = 2$ there are $|A_0 \times A_1| = 4$ possible outcomes in a symmetric 2×2 game. Player i has 2 possible actions to take as a response to each outcome and 2 possible starting positions and this means that there are $2^5 = 32$ possible deterministic strategies with memory length one. We denote the set of these strategies with S_D. Each of the two players can choose strategy independently of the other player and this means that there are $32^2 = 1024$ deterministic strategy profiles with memory length one for each symmetric 2×2 game.

The purpose of this experiment is to find the fittest deterministic strategies with memory length one in a specific symmetric 2×2 game and compare the fittest strategies in games belonging to different classes. To define what we mean by *fittest strategies* we need to introduce the concept of *populations of strategies*.

Definition 2.4.3 (Population of strategies)**.** We define a *population of strategies*, denoted $\mathcal{P}$ to be a set of strategies s such that $s \in \mathcal{P} \Rightarrow s \in S_D$. A population of order $n \geq 1$ is a population $\mathcal{P}$ with $|\mathcal{P}| = n$.

A population of strategies is simply a set of deterministic strategies with memory length one. It is important to note that a population of strategies does not have to be a subset of S_D since we allow for the same strategy $s \in S_D$ to occur in a population more than once.

We will now introduce the concept of fights. A fight of length $m \in \mathbb{N}$ is the iterated play of a given game G between two deterministic strategies with memory length one over m iterations. We denote the fight of length m between the strategies s_i and s_j as $F_m(s_i, s_j)$. After a fight $F_m(s_i, s_j)$ we are interested in the payoffs gained by the two strategies and this motivates definition 2.4.4.

Definition 2.4.4 (Total payoff)**.** Given a game G and a specific population of strategies $\mathcal{P}$, let p_{ij} denote the payoff gained by strategy $s_i \in \mathcal{P}$ in a fight $F_m(s_i, s_j)$ against strategy $s_j \in \mathcal{P}$. We define the *total payoff of strategy i* as $p_i = \sum_{s_j \in \mathcal{P}} p_{ij}$.

We say that a strategy s_i is fitter than a strategy s_j, $i \neq j$, in population $\mathcal{P}_0$ if $p_i > p_j$. We define the score of player i as a normalised payoff, so that the scores will not get too large.

Definition 2.4.5 (Score)**.** Given a game G and a specific population of strategies $\mathcal{P}$ we define the *score of player i*, $i \in \mathcal{P}$ as

$$score_i = \begin{cases} \frac{p_i - \min_{s_j \in \mathcal{P}} p_j}{\max_{s_k \in \mathcal{P}} p_k - \min_{s_j \in \mathcal{P}} p_j}, & \text{if } \max_{s_k \in \mathcal{P}} p_k - \min_{s_j \in \mathcal{P}} p_j > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Just knowing if a strategy is fitter than another strategy is not enough for our purposes and hence we introduce a measure of fitness in a population.

Definition 2.4.6 (Fitness)**.** We define the *fitness of strategy i with weight* α as

$$f\alpha_i = \frac{2^{|\mathcal{P}|} \times e^{\alpha \times score_i}}{\sum_{s_j \in \mathcal{P}} e^{\alpha \times score_j}}.$$

The definition of fitness guarantees that for each strategy $s_i \in \mathcal{P}$, $0 \leq f\alpha_i \leq |\mathcal{P}|$ and that $\sum_{s_i \in \mathcal{P}} f\alpha_i = |\mathcal{P}|$. It also guarantees that if s_i is fitter than s_j, i.e. $p_i > p_j$ then $f\alpha_i > f\alpha_j$. We use this definition of fitness to increase the evolutionary pressure, since an increase in score will result in a higher increase in fitness when defined in this way. The parameter α is used to alter the evolutionary pressure, with higher values on α the evolutionary pressure increases. We multiply with the size of the population, $2^{|\mathcal{P}|}$, so that the fitness of strategy i will correspond to the number of offspring strategy i gets in the following generation of strategies.

Now we are ready to construct the algorithm that is our experiment. The algorithm is presented in pseudocode in Algorithm 2.

Given a game G, the idea is to start with the population of strategies $\mathcal{P} = S_D$ and let every pair of strategies fight each other, i.e. play G a given number of iterations. When every fight is done the fitness $f\alpha_i$ is calculated for every strategy $s_i \in \mathcal{P}$ and a new population $\mathcal{P}_{new}$ is generated such that the strategy $s_i \in \mathcal{P}$ occurs in the new population exactly $\lfloor f\alpha_i \rfloor + 1$ times if the decimal part of $f\alpha_i$ is greater than the decimal part of the other strategies fitness and $\lfloor f\alpha_i \rfloor$

otherwise. We denote the size of the *offspring* of strategy $s \in \mathcal{P}$ as $\sigma(s)$.

This means that strategies with high fitness relative to the other strategies get more offspring than the other strategies and weak strategies get less or no offspring. We also introduce the concept of *mutation*, denoted Υ, to the algorithm. When the new population is generated the mutation can randomly cause some of the strategies to mutate, i.e turn into some other strategy in S_D. This is done so that extinct strategies always can challenge the dominating strategies.

This whole cycle is then repeated with $\mathcal{P} = \mathcal{P}_{new}$ as starting population a given number of generations $g \in \mathbb{N}$. The idea is that strategies that are good in the particular game being played will dominate the majority of the populations given that g is large enough.

In our particular case we set $m = 50$, $n = 2\ 000\ 000$ and $s = 0.0001$. To analyse the result we calculate how many percent of the strategies in the last 1 000 000 populations that have a 1 at each of the five positions 00, 01, 10, 11 and *start*. That is we calculate the average strategy over the last million populations. We do this for a large number of randomly generated games and compare how the resulting average strategies differ in different games. To visualise the data from our experiments we plot each game in the projective plane and assign a colour to it depending on the percentage of ones being played in this game. We can then analyse the regions depending on what colour is dominating each region.

```
s = Some strategy in the population 𝒫
p = Probability of mutation
for i = 0 to |s| - 1 do
   r = random[0,1]
   if r ≤ p then
      Mutate action s[i]
   end if
end for
```

Algorithm 1: Mutation

m = Number of rounds
n = Number of generations
$\mathcal{P} = S_D$
for g = 0 to n - 1 **do**
  for i = 0 to $|\mathcal{P}|$ - 1 **do**
    for j = i to $|\mathcal{P}|$ - 1 **do**
      $F_m(\mathcal{P}[i], \mathcal{P}[j])$
    end for
  end for
  $\mathcal{P}_{new} = \{\}$
  for k = 0 to $|\mathcal{P}|$ - 1 **do**
    if $\sigma(\mathcal{P}[k]) > 0$ **then**
      for l = 0 to $\sigma(\mathcal{P}[k])$ **do**
        $\mathcal{P}_{new} = \mathcal{P}_{new} \cup \{\Upsilon(\mathcal{P}[k])\}$
      end for
    end if
  end for
  $\mathcal{P} = \mathcal{P}_{new}$
end for

Algorithm 2: Evolution of Strategies

Result Analysis

In this subsection the results of the experiments presented in the former subsection will be analysed.

Comparative Literature Analysis

In this section a further analysis of the regions of the classification based on decomposition will be held by comparing with relevant literature. The comparisons made are partly with the previously reviewed classifications and partly with different experimental results. From this analysis we found that a lot of our regions have support from other authors as to why they are interesting. A lot of game theorists have shown interest in games that are repeated a finite number of times, where the number of repetitions are unknown by the players (Rapoport, Guyer, and Gordon, 1978), (Harris, 1969), (Lave, 1965), (Ells and Sermat, 1966), (Axelrod and Hamilton, 1981). If you examine the games in this way, finding the Nash equilibria of the one-shot game becomes insufficient to describe the game since the players are able to take future payoff into account.

Subtle changes in the payoff structure can change how the players play the game. In this analysis we show that we capture a lot of conditions that are captured in the classification method proposed in this thesis are proved important in a lot of experimental studies. We also conclude with a comparison of the other classifications reviewed in Chapter 1 and conclude that the proposed classification by decomposition captures all of the aspects found important based on the literature review in Chapter 1.

In this section the actions labelled C stands for cooperation and the actions labelled D stands for defection in the game examples used. They are meant to clarify which strategy is cooperative and which is defective, i.e. which strategy promotes the common good and which is selfish.

Analysis of the Prisoner's Dilemma Regions

In the classification of 2×2 symmetric games proposed by Harris, which is presented in Chapter 1, he distinguishes between three types of Prisoner's Dilemmas by the following conditions on the payoffs:

1. $b + c > 2a$
2. $2d < b + c < 2a$
3. $b + c < 2d$

All of these conditions are capture into different regions in our proposed classification method. To further motivate this partition of the Prisoner's Dilemma comparisons to more thorough studies are made. We get four different versions of the Prisoner's Dilemma, distinguished based on the conditions showed below.

1. $b + c > 2a$
2. $2d < b + c < 2a$ and $a - c > b - d$
3. $2d < b + c < 2a$ and $a - c < b - d$
4. $b + c < 2d$

For the partition based on the inequality $a - c \lessgtr b - d$ we found no evidence from reviewing other literature. We did however find a intuitive meaning, that it affects the cost of signalling for cooperation and accepting it and we propose the conjecture that this will have an impact on how the game is played.

The game in Table 2.2, where $b+c > 2a$, is in the part of N_4 where $|c_2| < |c_1|$. The game in Table 2.3, where $2d < b+c < 2a$, is found in both regions of N_3. This game in Table 2.3 is with $c_1 = c_2$, which corresponds to $a-c = b-d$. The game in Table 2.4, where $b+c < 2d$, is found in the part of N_2 where $|c_1| > |c_2|$.

	C	D
C	$(2,2)$	$(0,6)$
D	$(6,0)$	$(1,1)$

Table 2.2: PD1

	C	D
C	$(3,3)$	$(0,4)$
D	$(4,0)$	$(1,1)$

Table 2.3: PD2

	C	D
C	$(7,7)$	$(0,8)$
D	$(8,0)$	$(5,5)$

Table 2.4: PD3

These three Prisoner's Dilemmas distinguished by Harris, (1969) are not included in the classifications by Huertas-Rosero, (2003), Rapoport, Guyer, and Gordon, (1978) or Robinson and Goforth, (2003).

The variations of the Prisoner's Dilemma first become interesting in iterated play (Lave, 1965). If the game is played only once, the players will surely take action C given that the satisfy the game theoretical axioms of rationality, because it will always guarantee higher payoff in the short term. In iterated play, however, they might decide on cooperating on action C in order to receive a jointly higher payoff in the long run. As Lave argues, if the players play the game many times without knowing how long they are going to keep playing it, they might realize that it is better to deviate from the Nash equilibrium. The players will sacrifice short term payoff, but in doing so they will potentially receive higher total payoff in future iterations (Axelrod and Hamilton, 1981).

In the version of Prisoner's Dilemma in Table 2.2 where $b+c > 2a$ the players may start to alternate between the strategy profiles {C,D} and {D,C} because it gives both higher payoff when the game is repeated (Harris, 1969). Some authors argue that this makes the distinction between cooperation and defection too vague (Harris, 1969), (Axelrod and Hamilton, 1981).

The third version of Prisoner's Dilemma, in Table 2.4, where $b+c < 2d$. The inequality $d-b < a-d$ discussed by Lave, (1965) does not hold. In this region $a < c$, so

$$b+a < b+c < 2d \Rightarrow a-d < d-b.$$

This means that the loss for a failed attempt of signalling for cooperation is higher than the gain of a successfull one (Harris, 1969), (Lave, 1965).

As in the example in Table 2.4, the cost for signalling, by switching from action C to action D, and telling the opposing player that you wish to cooperate is $d-b=5-0=5$ and the gain is only $a-d=7-5=2$ given that the opposing player responds to the signal and switches to action C to D. This means that in the short term, the players loses more if the cooperation attempt fails than they gain if it succeeds.

Many authors advocates the *restricted* Prisoner's Dilemma shown in Table 2.3, where the inequality $2d < b+c < 2a$ is satisfied (Harris, 1969), (Axelrod and Hamilton, 1981). The restricted Prisoner's Dilemma, which is found in region N_3, is however divided into two different games, depending on if $|c_1| > |c_2|$ as in Table 2.5 or if $|c_1| < |c_2|$ as in Table 2.6.

	C	D
C	$(4,4)$	$(-1,5)$
D	$(5,-1)$	$(1,1)$

Table 2.5: PD3.1

	C	D
C	$(3,3)$	$(-1,5)$
D	$(5,-1)$	$(0,0)$

Table 2.6: PD3.2

In the case of $-c_1 > -c_2$ we get the additional restriction that $a-c > b-d$. This means that each player loses more from signaling than player two loses from giving up his short term higher payoff to respond to the signal. In the example shown in Table 2.5 $c-a=1$ and $d-b=2$. Player 1 pays a cost of 2 for signalling to player 2 that he would like to cooperate. Player two only pays a short term cost of 1 for cooperating.

In the other case when $-c_1 < -c_2$ the inequality gets reversed so that $a-c < b-d$. So in this case it is cheap to signal but expensive to accept. An example of this game is shown in Table 2.6.

A conjecture of what could separate these games is that the stability of the games varies between these cases.

The inequality $d-b < a-d$ is satisfied iff $-c_1-2c_2 > -z_1$. In our classification this inequality divides every Prisoner's Dilemma region, except for the one with games as in Table 2.4, into two parts. In these regions the inequality fails to hold when $|z_1|$ is too high, i.e. when close to the origin in our map. This might show a weakness in our classification since Lave found it to have an impact on how the game was played in his experiments.

The higher the conflict the lower chance of cooperation?

Analysis of the Stag Hunt Regions

The T_3 region contains Stag Hunt. The version in table 2.7 satisfies the inequality $b + c < 2d \Leftrightarrow c_2 > 0$ whereas the game in table 2.8 satisfies $b + c > 2d \Leftrightarrow c_2 < 0$. This game was overlooked by Harris, (1969) and just called a no-conflict game. This game has, however, received a lot of attention from other authors (Skyrms, 2004), (Dubois, Willinger, and Nguyen, 2012).

Many authors finds it interesting to analyse Stag Hunt to see whether players will choose the payoff-dominant or the low-risk equilibrium point. A Nash equilibrium outcome is said to be payoff-dominant if it is not strictly Pareto dominated by any other outcome (Van Huyck, Battalio, and Beil, 1991). A low-risk equilibrium can be interpreted as the intersection of the maximin strategies introduced by Rapoport, Guyer, and Gordon, (1978).

Interesting aspects of Stag Hunt are with what frequency the players play the *risk-dominant* or the *payoff-dominant* strategy profile (Van Huyck, Battalio, and Beil, 1991). the A conjecture is that the inequality $b + c > 2d$ makes the *payoff-dominant equilibrium* more frequent and the $b + c < 2d$ makes the *risk-dominant equilibrium* played more often. This makes intuitive sense since the payoff for defecting is smaller in comparison to the other payoffs when $b+c < 2d$.

In the experiments conducted by Dubois, Willinger, and Nguyen, (2012) the result was that cooperation was a lot more frequent in the version where the inequality $b + c > 2d$ is satisfied. They did not however explain it with this inequality, but their experiments are consistent with this conjecture.

	C	D
C	$(4, 4)$	$(-1, 3)$
D	$(3, -1)$	$(2, 2)$

Table 2.7: Stag Hunt 1

	C	D
C	$(5, 5)$	$(0, 4)$
D	$(4, 0)$	$(1, 1)$

Table 2.8: Stag Hunt 2

Rapoport, Guyer, and Gordon, (1978) discusses the difference between Stag Hunt and the game in table 2.9 which is found in region K_3. Although similar, the difference that $b > d$ in the latter game makes cooperation a dominant strategy. The results by Rapoport, Guyer, and Gordon showed that cooperation was a lot more frequent in the game where cooperation is a dominant strategy than in Stag Hunt which was to be expected.

	C	D
C	$(6, 6)$	$(3, 5)$
D	$(5, 3)$	$(2, 2)$

Table 2.9: Cooperation Dominant Game

Analysis of the Chicken Regions

The T_4 region contain the Chicken game. They are separated by if c_1 is positive or negative. If c_1 is positive then $b + c > 2a$ and if c_1 is negative this implies that $b + c < 2a$.

	C	D
C	$(5, 5)$	$(2, 6)$
D	$(6, 2)$	$(1, 1)$

Table 2.10: Restricted Chicken

	C	D
C	$(3, 3)$	$(2, 6)$
D	$(6, 2)$	$(1, 1)$

Table 2.11: Non-restricted Chicken

Ells and Sermat, (1966) argues for the distinction in equation 2.10

$$2d < b + c < 2a. \tag{2.10}$$

The left restriction in equation 2.10 is always satisfied because $d < b < a < c$ in these regions. Ells and Sermat, (1966) showed by conducting experiments of players playing different versions of chicken, that cooperation decreased when $c - b$ increased, i.e. when $|z_1|$ is great. Their experiments also showed that increasing a decreases cooperation more than decreasing b. In our classification a is always equal to $-b$ but this could be interpreted as the players being less likely to cooperate if $b + c > 2a$ than if $b + c < 2a$. This does make intuitive sense, since it is more profitable for players to alternate between the non-cooperative outcomes when $b + c > 2a$ in iterative play. Rapoport, Guyer, and Gordon, (1978) also shows in their experiments that players relatively quickly finds out how to take advantage of the fact that $b + c > 2a$.

The game in Table 2.11 is called *restricted chicken* (Harris, 1969).

Analysis of the Civic Duty Game Regions

For most of the games in these regions we have not found any interesting properties, but the games in the K_4 that are discussed by several game theorists (Rapoport, Guyer, and Gordon, 1978), (Rasmusen, 1989), (Harris, 1969).

The games in table 2.12 and table 2.13 are found in region K_4 and are both Civic Duty Games (Rasmusen, 1989). The difference between these two games is that in the game in table 2.12 the inequality $|c_1| > |c_2|$ is satisfied whereas the game in table 2.13 satisfies the inequality $|c_1| < |c_2|$.

	C	D
C	$(1, 1)$	$(3, 5)$
D	$(5, 3)$	$(2, 2)$

Table 2.12: Leader

	C	D
C	$(2, 2)$	$(3, 5)$
D	$(5, 3)$	$(1, 1)$

Table 2.13: Hero

The game in table 2.12 is called Leader and the game in table 2.13 is called Hero (Rapoport, Guyer, and Gordon, 1978). The cooperative strategy for both players in iterated play is to alternate between the $\{C, D\}$ and $\{D, C\}$ outcomes. Both players maximin strategies are D in the Leader game and C in the Hero game. This means that the natural outcome is $\{D, D\}$ in Leader and $\{C, C\}$ in Hero. The difference between these games, as Rapoport, Guyer, and Gordon, (1978) describes, is that in the Leader game the player who shifts from the natural outcome gets a higher payoff than the one who waits. In Hero it is the other way around. Both players are interested in one of them making an unilateral shift, but would prefer the other to shift first. Therefore the player who sacrifices his largest payoff for the greater good is called the hero.

Rapoport, Guyer, and Gordon also calls the Hero game a ”procrastination game” and the Leader game a ”preemption game”. In the experiments they conducted on these games the natural outcome was a lot more frequent in the Hero game than in the Leader game. This result is natural since the players would rather procrastinate and wait for the other player to make a move in the Hero game.

The Leader and Hero distinction is also included in Harris, (1969) classification, but not in any other classifications reviewed in this thesis. Harris names the game in table 2.13 Restricted Battle of Sexes and the game in table 2.12 Restricted Apology.

Comparison with Reviewed Classifications

From this analysis we can conclude that the proposed classification captures all properties captured in the classification by Huertas-Rosero. By the further analysis of the regions in our classification we conclude that even more interesting properties are included. For example, all the regions in the classification by Harris are also included in our classification but not in the one proposed by Huertas-Rosero. Harris introduces more conditions ad hoc to classify the Restricted Chicken and Restricted Prisoner's Dilemma, whereas we get these classified directly from our parameter conditions. Huertas-Rosero also add ad hoc conditions to divide classes based on if the payoff is large in the HO outcome than in the NE outcome. This is also captured directly from our parameter conditions. Our classification includes all the 12 strategically non-equivalent symmetric 2×2 games that Rapoport, Guyer, and Gordon classifies. We divide these games further in our proposed classification. For example we get several interesting versions of the Prisoner's Dilemma, Chicken and Stag Hunt which are not included in the classification by Rapoport, Guyer, and Gordon. Robinson and Goforth also get 12 different classes of symmetric 2×2 games, the same as Rapoport, Guyer, and Gordon. They therefore do not capture anything in their classification that we do not capture in ours.

From this analysis we therefore conclude that our classification captures all the properties we found interesting that the other classifications reviewed in this thesis capture. This is not done by any other classification of 2×2 symmetric games that we have found and reviewed in this thesis. To summarise this we will add our classification to Table 1.22 presented in Chapter 1.

Classification Author	Mathematical Structure	Interesting Conditions	Mathematical Sophistication	All Standard Games
Rapoport, Guyer, and Gordon	No	Yes	No	No
Robinson and Goforth	Yes	No	Yes	No
Harris	Yes	Yes	No	No
Borm and Du	Yes	No	Yes	No
Huertas-Rosero	Yes	Yes	No	No
Boors, Wangberg	Yes	Yes	Yes	Yes

Table 2.14: Summary

Interesting Properties

An interesting question in game theory is what makes a game interesting or not, how the properties of an interesting game can be formalised and what the correct mathematical conditions for describing such games are. When defining a classification of games the goal is to divide games with different kinds of interesting properties into different classes. Therefore, before one defines a classification, the natural question to answer is *what properties are interesting?* and perhaps even more important, *why do we find these properties interesting?* If a classification of games is to be constructed, this is the question to answer first. After reviewing several different classifications with different foundations, we realise that the answers to these questions are far from trivial and that different authors have come up with diverse ideas about what properties makes a game interesting. In this section we will try to answer these questions formalise the answers in a more philosophical manner.

An interesting objection to the NE condition is discussed by Rasmusen, (1989). Consider the Ranked Coordination game, shown in Table 1.12. Consider also the game of Dangerous Coordination shown in Table 2.15 below.

		Player 2: A	Player 2: B
Player 1	A	$(3, 3)$	$(-1000, 0)$
	B	$(0, 0)$	$(1, 1)$

Table 2.15: Dangerous Coordination

According to NE and HO conditions, the game in Table 2.15 is exactly the same game as the one in Table 1.12 even though the outcome is likely to be different. In the case of Dangerous Coordination, information about the opposing players type makes a big difference, compared to Ranked Coordination. This is an example of why an ordinal scale might lose important information about games and that NE sometimes fails to capture the necessary aspects to fully describe a game.

Nash and Huertas conditions also make no difference between the standard games Battle of the Sexes, Table 1.11 and Ranked Coordination, Table 1.12. Since these games have completely different stories and payoff profiles in the Nash equilibria, it seems like the NE and HO fails to capture some vital information about these games, in this case it fails to capture the conflict of interest between the players.

To conclude this section we conjecture that a way of formalising what makes a game interesting is by decomposing it into a common interest and a conflict part, as done in this section. A formal criterion for what makes a game interesting is that the conflict works against the common interest. The properties that we identified in this thesis to make a game interesting are

1. the conflict pulls the players in the opposite direction as the common interest,

2. the conflict is stronger than the common interest and

3. the conflict directly affects the common interest outcome.

Number of Games

In this section we discuss how the size of the space of strategically non-equivalent $m \times n$ games increase as m and n grows larger. There are two standard definitions of *strategically non-equivalent* games. Robinson and Goforth, (2003) consider two games to be strategically equivalent if one game can be obtained by permuting the rows or columns of the other game's payoff matrix. He does however differ between the two player. A more common definition is that two games are strategically equivalent if this holds or if one game can be obtained by switching the players in the other game.

In this thesis we use the latter definition, but we will nevertheless provide some result that applies when using the first definition. The reason for this is that the step from the first to the second definition is small and the results we present is similar in the two cases. We will refer to the first definition as *strategically non-equivalent with distinct players.*

Distinct players

Here we consider the case when switching the players in a game does not result in a strategically equivalent game in the general case.

Theorem 2.7.1 (Number of $m \times n$ games with distinct players)**.** The size of the space of strategically non-equivalent games with distinct players and strictly ordinal payoffs of size $m \times n$ is

$$\frac{((mn)!)^2}{m!n!}.$$

Proof. The number of ways that the payoffs of the first player can be placed in a $m \times n$ matrix is $(mn)!$. The payoffs of the second player can be placed in $(mn)!$ ways. Since we consider arbitrary $m \times n$ games the total amount of payoff bi-matrices are $(mn!)^2$ according to the multiplication principle.

We can permute the rows of the matrix in $m!$ ways and the columns in $n!$ ways. So each game can be represented by $m!n!$ matrices.

Therefore the total amount of strategically non-equivalent ordinal $m \times n$ game matrices is $\frac{((mn)!)^2}{m!n!}$. □

Corollary 2.7.2 (Number of symmetric $n \times n$ games with distinct players)**.** The size of the space of strategically non-equivalent symmetric games with distinct players and strictly ordinal payoffs of size $n \times n$ is

$$\frac{n^2!}{n!}.$$

Theorem 2.7.1 not only gives us a way to find out exactly how many $m \times n$ games there are for some fixed m and n but it also provides us with information about how fast the size of the space of strategically non-equivalent strictly ordinal $m \times n$ games grows as m and n grows large. The numerator in the fraction will dominate the denominator for large values of m and n and the size of the game space will grow **exponentially**. In Table 2.16 the number of games for some values of m and n are presented.

$m \backslash n$	2	3	4
2	144	43200	33868800
3	43200	3657830400	1.5×10^{15}
4	33868800	1.5×10^{15}	7.6×10^{23}

Table 2.16: Number of $m \times n$ games, distinct players

Table 2.16 gives an idea of how fast the space of $m \times n$ games grow as m and n grows large. This is interesting to know if one would for example generalize the topology classification made by (Topology) to higher dimensions. If one were to generalize the topology classification to 2×3 strictly ordinal games, one would have to classify 43200 games. 43200 is a restively large number in comparison with the 144 strictly ordinal 2×2 games and one might wonder how many 2×2 and 2×3 games there are if one allow non-strictly ordinal games. In the 2×2 case, the answer is 1413, including the 144 strictly ordinal games (Robinson, Goforth, and Cargill, 2007). Games with ties are interesting to study since real-world situations often can be modeled by such games. There is a big difference in the number of 2×2 strictly ordinal games and the number of non-strictly ordinal 2×2 games and that can motivate why Robinson and Goforth only chose to classify the strictly ordinal games. To give an idea about how the number of strictly ordinal $m \times n$ games relate to the number of of $m \times n$ games where ties are allowed, we will start by considering the 2×3 case and then give an lower bound of the size relationship between the two cases.

In a 2×3 game there are $2 * 3 = 6$ payoffs for each player. We start by considering the payoffs for one of the players and we call these payoffs $a_1, a_2, ..., a_6$. Without loss of generality we can suppose that the payoffs are ranged in the order $a_1 < a_2 < ... < a_6$ and the order of the payoffs is therefore given by 5 inequalities. The first step is to find out in how many ways we can change the strict inequalities to equalities. This is a combinatorial problem and to answer it we need to calculate the number of possible combinations of the 5 inequalities.

For example, there is $\binom{5}{0} = 1$ combination of size 0 of the inequalities, meaning that there is exactly one way in which we can keep the strict inequalities between the payoffs. In the same fashion there are $\binom{5}{1} = 5$ combinations of size 1 of the inequalities, which can be interpreted as that there are 5 ways exactly one pair of the payoffs can be equal. With an analogue reasoning we realise that there are

$$\sum_{n=0}^{5} \binom{5}{n} = 2^5 = 32 \tag{2.11}$$

different types of orders of the payoffs and these are listed below. To simplify the representation we denote for example $a_1 < a_2 < ... < a_6$ as 123456 and $a_1 = a_2 < a_3 = a_4 < a_5 < a_6$ as 112234 and so on.

1. 111111	5. 112222	9. 112223	13. 122233
2. 111112	6. 122222	10. 122223	14. 112333
3. 111122	7. 111123	11. 111233	15. 122333
4. 111222	8. 111223	12. 112233	16. 123333

17. 111234	21. 122334	25. 123344	29. 123345
18. 112234	22. 123334	26. 123444	30. 123445
19. 122234	23. 112344	27. 112345	31. 123455
20. 112334	24. 122344	28. 122345	32. 123456

So for each player there are 32 types of ties, where 0 ties is counted as a tie case. That means that there are $32 * 32 = 1024$ different *payoff type combinations*. Each of these combinations contains a set of plausible non-strategically equivalent games.

From now on we will denote the payoff type combination i, j as $P[i, j]$ where $i, j \in \{1, 2, ..., 2^{mn-1}\}$ from the list above is the payoff type of player one and two respectively.

Example 2.7.3. Suppose player one's payoffs $a_1, a_2, ..., a_6$ is of payoff type 4, that is, $a_1 = a_2 = a_3 < a_4 = a_5 = a_6$ and player two's payoffs $b_1, b_2, ..., b_6$ is of payoff type 30 so that $b_1 < b_2 < b_3 < b_4 = b_5 < b_6$. Then the players have payoff type combination $P[4, 30]$ and the payoff types can be expressed as 111222 and 123445 respectively.

We also denote the number of equivalent permutations of payoff type $i \in \{1, 2, ..., p\}$, where

$$p = \sum_{k=0}^{mn-1} \binom{mn-1}{k} = 2^{mn-1} \tag{2.12}$$

in a mxn game, with $|P[i, -]|$ and $|P[-, i]|$ for player one and player two respectively.
In the list below we present $|P[i, -]| = |P[-, i]|$ for every i in the 2×3 case.

1. 720	7. 24	13. 12	19. 6
2. 120	8. 12	14. 12	20. 4
3. 48	9. 12	15. 12	21. 4
4. 36	10. 24	16. 24	22. 6
5. 48	11. 12	17. 6	23. 4
6. 120	12. 8	18. 4	24. 4

25. 4	27. 2	29. 2	31. 2
26. 6	28. 2	30. 2	32. 1

Calculating the number of non-strictly ordered 2×3 games of payoff type combination $P[i, j]$ is a bit more complicated in the general case than calculating the number of games with payoff type combination $P[32, 32]$. This is due to the fact that the number of permutations of the payoff matrix now depend on the structure of the matrix. In the strictly ordinal case there are 12 strategically equivalent permutations of the matrix sice we can permute the rows in $3! = 6$ ways and the columns in $2! = 2$ ways. We present the twelve strategically equivalent matrices below. If we let R_{ij} denote the permutation of row i and j, and C the permutation of the two columns, the matrices below is generated by the permutation sequence $CR_{12}CR_{23}CR_{12}CR_{23}CR_{12}C$.

$$\begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix} \quad \begin{bmatrix}2&1\\4&3\\6&5\end{bmatrix} \quad \begin{bmatrix}4&3\\2&1\\6&5\end{bmatrix} \quad \begin{bmatrix}3&4\\1&2\\5&6\end{bmatrix} \quad \begin{bmatrix}3&4\\5&6\\1&2\end{bmatrix} \quad \begin{bmatrix}4&3\\6&5\\2&1\end{bmatrix}$$

$$\begin{bmatrix}6&5\\4&3\\2&1\end{bmatrix} \quad \begin{bmatrix}5&6\\3&4\\1&2\end{bmatrix} \quad \begin{bmatrix}5&6\\1&2\\3&4\end{bmatrix} \quad \begin{bmatrix}6&5\\2&1\\4&3\end{bmatrix} \quad \begin{bmatrix}2&1\\6&5\\4&3\end{bmatrix} \quad \begin{bmatrix}1&2\\5&6\\3&4\end{bmatrix}$$

Since we are only allowed to permute the rows and the columns, if two arbitrary elements belong to the same row, they will belong to the same row in every possible permutation. In the same way if two elements belong to the same column thew will always belong to the same column. This comes from the fact that the row-permutation operation never can affect what column an element belongs to and the column-permutation operation never can affect what row an element belongs to. This means that if one of the operations of the permutation sequence above would result in an identical matrix due to the structure of the matrix, the matrix has exactly one of the structures listed below.

1. $\begin{bmatrix}1&2\\1&2\\3&4\end{bmatrix}$ Two equal rows. 1 and 2 arbitrary numbers with $1 = 2$ allowed. 3 and 4 are arbitrary numbers with the restriction that $3 \neq 1$ or $4 \neq 2$, or if $1 = 2$ then $3 \neq 4$. $3 \times 2! = 6$ possible permutations.

2. $\begin{bmatrix}1&2\\1&2\\1&2\end{bmatrix}$ Three equal rows. 1 and 2 arbitrary numbers with the restriction $1 \neq 2$. $1 \times 2! = 2$ possible permutations.

3. $\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}$ Two equal columns with distinct rows. $1 \neq 2$, $1 \neq 3$ and $2 \neq 3$.
$3! \times 1 = 6$ possible permutations.

4. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}$ Two equal columns and two equal rows. $1 \neq 2$.
$3 \times 1 = 3$ possible permutations.

5. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$ Two equal columns and three equal rows.
$1 \times 1 = 1$ possible permutation.

6. $\begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{bmatrix}$ Two equal diagonals and one row with the same element in both columns. $1 \neq 2$.
6 possible permutations.

If we introduce a seventh matrix structure that describes all matrices which do not belong to any of the six categories above, the number of possible 2×3 games of payoff type combination $P[i, j]$ is given by

$$\sum_{k=1}^{7} \frac{6!6!Q_k}{|P[i,-]||P[-,j]|q_k} \tag{2.13}$$

$k \in \{1, 2, ..., 7\}$, where Q_k is the fraction of all $P[i, j]$ matrices that has structure k and q_k is the number of permutation associated with the $k : th$ matrix structure. This means that the total number of 2×3 games is given by the triple sum

$$(6!)^2 \times \sum_{i=1}^{32} \sum_{j=1}^{32} \sum_{k=1}^{7} \frac{Q_k}{|P[i,-]||P[-,j]|q_k}. \tag{2.14}$$

The tricky part of finding the number of 2×3 games is therefore to calculating Q_k. It is however easy to get a lower bound of the sum in Equation 2.14 by using the fact that $\sum_{k=1}^{7} Q_k = 1$ and that $q_k \leq 12$ for every k. So we have that

$$\begin{aligned} &(6!)^2 \times \sum_{i=1}^{32} \sum_{j=1}^{32} \sum_{k=1}^{7} \frac{Q_k}{|P[i,-]||P[-,j]|q_k} \geq \\ &(6!)^2 \times \sum_{i=1}^{32} \sum_{j=1}^{32} \frac{1}{|P[i,-]||P[-,j]| \times 12} = 2119320.75. \end{aligned} \tag{2.15}$$

We can also get an upper bound for the sum in Equation 2.14 by using the fact that $q_k \geq 1$. This together with the result in Equation 2.15, gives us that the number of non-strategically equivalent non-strictly ordinal 2×3 games, denoted $\nu_{2\times 3}$, satisfy

$$2119320.75 \leq \nu_{2\times 3} \leq 25431849. \tag{2.16}$$

The gap between the lower bound and the upper bound in Equation 2.16 is to big to say anything useful about the exact value of $\nu_{2\times 3}$ due to the fact that the upper bound is highly pessimistic. However we are not interested in the exact value of $\nu_{2\times 3}$, but rather the lower bound of it. The reason for why the lower bound is interesting is because the reasoning for finding it can be generalized and provide the necessary tools for calculating how fast the number of non-strategically equivalent $m \times n$ games, denoted $\nu_{m\times n}$ grows relative to how fast the number of non-strategically equivalent strictly ordinal $m \times n$ games grows. That is we are interested in the growth rate of $\eta_{m\times n}$ where

$$\eta_{m\times n} \triangleq \frac{\nu_{m\times n}}{\frac{((mn)!)^2}{m!n!}}. \tag{2.17}$$

We can get a lower bound for $\eta_{m\times n}$, denoted $l(\eta_{m\times n})$ by using the lower bound for $\nu_{m\times n}$, denoted $l(\eta_{m\times n})$, in Equation . That is

$$l(\eta_{m\times n}) = \frac{l(\nu_{m\times n})}{\frac{((mn)!)^2}{m!n!}}. \tag{2.18}$$

In the 2×2 case we have that $l(\eta_{2\times 2}) = 9.77$ (since $l(\nu_{2\times 2}) = 1406.25$) and in the 2×3 case we have that $l(\eta_{2\times 3}) = 49.06$. This indicates that $\eta_{m\times n}$ grows quickly as m and n grows larger.

Non-distinct players

Here we answer the same questions as in Section 2.7, but with the addition that two games are considered strategically equivalent if the payoff matrix of one of the games can be obtained by permuting the rows, the columns or the players in the other game.

Lemma 2.7.4 (Number of symmetric $n \times n$ games)**.** The size of the space of strategically non-equivalent symmetric games with strictly ordinal payoffs of size $n \times n$ is

$$\frac{n^2!}{n!}.$$

Proof. Corollary 2.7.2 states that the number of non strategically equivalent symmetric games with distinct players and strictly ordinal payoffs is

$$\frac{n^2!}{n!}.$$

Permuting the players results in the same game in the symmetrical case and therefore there are only one permutation of the players and hence there are

$$\frac{n^2!}{1 \times n!} = \frac{n^2!}{n!}$$

strategically non-equivalent symmetric games with strictly ordinal payoffs of size $n \times n$.

□

Theorem 2.7.5 (Number of $m \times n$ games)**.** The size of the space of strategically non-equivalent games with strictly ordinal payoffs of size $m \times n$ is

$$\begin{cases} \frac{((mn)!)^2}{2m!n!}, & \text{if } m \neq n \\ \frac{(n^2)!}{2n!}\left(\frac{(n^2)!}{n!} + 1\right), & \text{otherwise.} \end{cases}$$

Proof. Let G be a $n \times n$ game with strictly ordinal payoffs. The payoff matrix of G has the form

		Player A 0	Player A 1
Player B	0	(a, e)	(b, f)
	1	(c, g)	(d, h)

and permuting the players result in the new payoff matrix

		Player B 0	Player B 1
Player A	0	(e, a)	(g, c)
	1	(f, b)	(h, d)

.

The two matrices are equal iff $a = e$, $b = g$, $c = f$ and $d = h$, i.e. if G is symmetrical. This means that there is exactly one permutation of the players if G is a symmetric game and two if G is non-symmetric.

Theorem 2.7.1 states that there are

$$\frac{((n^2)!)^2}{(n!)^2}$$

games of size $n \times n$ if we consider the players to be distinct. If we do not wish to distinguish between the two players we get that there are

$$\frac{\frac{((n^2)!)^2}{(n!)^2} - \frac{n^2!}{n!}}{2}$$

non-strategically equivalent, non-symmetric $n \times n$ games with strictly ordinal payoffs and non-distinct players. Lemma 2.7.4 states that there are

$$\frac{n^2!}{n!}$$

such symmetric games and since a game is either symmetric or non-symmetric there are exactly

$$\frac{\frac{((n^2)!)^2}{(n!)^2} - \frac{n^2!}{n!}}{2} + \frac{n^2!}{n!} = \frac{(n^2)!}{2n!}(\frac{(n^2)!}{n!} + 1)$$

non-strategically equivalent $n \times n$ games.
IN the case where $m \neq n$ there are always two permutations of the players and hence we have that there are

$$\frac{((mn)!)^2}{2m!n!}$$

non-strategically equivalent $m \times n$ games if $m \neq n$.

□

Theorem 2.7.5 tells us that there is no major difference between the size of the space of strategically non-equivalent $m \times n$ games with distinct players and strictly ordinal payoffs and the size of such games with non-distinct players. In fact in the case where $m \neq n$ the number of games only differ with a factor of 2 between the two cases. In the Table 2.17 we present the number of $m \times n$ games with non-distinct players nor small values of m and n.

$m \backslash n$	2	3	4
2	78	21600	16934400
3	21600	1828945400	0.75×10^{15}
4	16934400	0.75×10^{15}	3.8×10^{23}

Table 2.17: Number of $m \times n$ games, non-distinct players

The conclusion is that the size of the space of non-strategically equivalent $m \times n$ games grows **exponentially** as m and n grows large.

Number of Nash equilibria

Is the concept of Nash equilibria really interesting and if so, do all different types of distribution of Nash equilibria divide the set of games into interestingly different classes? Initially it might seem obvious that Nash equilibria should, in some way, be a part of the properties used to define interesting classes of games since the Nash equilibria of a game, or lack thereof, often will define the outcome of the game. *Prisoner's dilemma* is a good example of this since the unique NE determines the outcome of the game, given that the players are rational and selfish. However the NE of a game does not always define the outcome of the game and a good example of this is the classical game *Stag hunt* in which there are two NE but the outcome of the game is not deterministic and could very well be a state that is not one of the two NE. This shows that the type of NE a game possesses does not always determine the outcome of the game.

In this section we prove that the classes in our classification can be separated into two categories; classes that contain deterministic games and classes that contain non-deterministic games.
Given that a 2×2 symmetric game has a unique NE and that both players are rational and selfish, then the outcome of the game is deterministic and is bound to be the NE outcome if there are no ties between the payoffs.

Proposition 2.8.1 (Unique Nash in symmetric games)**.** Let G be a symmetric 2×2 game with distinct payoffs, i.e. no pair of payoffs are equal. Then the outcome of G is deterministic iff G has a unique NE.

Proof. Suppose that the payoff matrix of a game is the following.

		Player A 0	Player A 1
Player B	0	(a, a)	(b, c)
	1	(c, b)	(d, d)

Suppose further that a, b, c and d are distinct so that no pair of payoffs are equal. Then the following is true.

1. 00 is NE iff $a > c$.

2. 01 is NE iff $b > d$ and $c > a$.

3. 10 is NE iff $c > a$ and $b > d$.

4. 11 is NE iff $d > b$.

Of course, 2 is true if and only if 3 is true. So neither 01 nor 10 can be a unique NE. That leaves us with two possible unique NE.

In the reasoning belong both players are supposed to be rational in the sense that they always plays to maximize their utility function.

Suppose 00 is a unique NE so that 1 is true but 2,3 and 4 are false. Then we know that $a > c$ and $d < b$ and hence if Player A plays 0, Player B must play 0 and if Player A plays 1 player B must play 0. Player B will therefore always, without the risk of getting a lower payoff, choose to play 0. The exact same reasoning will force Player A to play 0 without risking getting a lower payoff. In this case 0 is a dominant strategy for both players.

This proves that if both players are rational and selfish and the game has a unique NE in 00 the outcome of the game is sure to be 00.

Now suppose instead that 11 is the unique NE in the game so that 1, 2 and 3 are false but 4 is true. Then $a < c$ and $d > b$. If Player A plays 0 Player B must play 1 to maximize his payoff and if Player A plays 1 player B must also play 1 if. No matter what strategy Player A chooses Player B will never risk anything by playing 1 but he will risk getting a lower payoff if he plays 0 and since he is rational he will choose to play 1. The same reasoning will force Player A to play 1 and this means that if 11 is a unique NE then the outcome is sure to be 11.

There are no other possible states that can be unique NE and hence, if a 2×2 symmetric game has a unique NE then that outcome is bound to happen if both players are rational and selfish.

Suppose now that the game has two NE. We can, without loss of generality assume that the two NE outcomes are 01 and 10.
If 01 and 10 are the only two NE in a game then we know that $b > d$ and $c > a$. Given that Player A plays 0 Player B must play 1. If Player A plays 1 Player B must play 0. The exact same reasoning applies if we switch players in the reasoning and since neither of the players can deduce what the other will play the outcome is not deterministic.

The exact same reasoning will show that the game is not deterministic if 00 and 11 are NE. There are no other possible combinations of NE and hence the proof is complete.

□

In the reasoning in the proof to Proposition 2.8.1 it is causal that there are no ties between the payoffs. Consider for example the games below. In the left game Player A will always play 1 and Player B will also always play 1 and hence 11 is a deterministic outcome of the game even though 11 is not a unique NE (00 is also NE). In the right game however the outcome of the game is not deterministic.

		Player A 0	Player A 1
Player B	0	$(1,1)$	$(2,1)$
	1	$(1,2)$	$(3,3)$

		Player A 0	Player A 1
Player B	0	$(1,1)$	$(2,3)$
	1	$(3,2)$	$(0,0)$

This shows that if we allow ties there is no equivalence relationship between a game having a deterministic outcome and it having a unique NE. We can however prove that if such a game has a unique NE, then the game is deterministic.

Proposition 2.8.2 (Unique NE implies outcome)**.** Let G be a symmetric 2×2 game. If G has a unique NE in ii then G is deterministic and ii is the deterministic outcome.

Proof. Since we allow ties among the payoffs we have that

1. 00 is NE iff $a \geq c$.
2. 01 is NE iff $b \geq d$ and $c \geq a$.
3. 10 is NE iff $c \geq a$ and $b \geq d$.
4. 11 is NE iff $d \geq b$.

Neither 01 nor 10 can be a unique NE and that leaves us with to possible unique NE outcomes. Proposition 2.8.1 deals with the case where the inequalities are strict and therefore we need to show that the implication holds in two cases. We use the best response correspondences B_A and B_B to simplify the reasoning.

00 **is unique NE**. We have that $a = c$ and $b > d$. $B_A(0) = \{0, 1\}$, $B_A(1) = \{0\}$ and hence Player A will always play 0. $B_B(0) = \{0, 1\}$, $B_B(1) = \{0\}$ so Player B will also always play 0 and that means that 00 is deterministic.

11 **is unique NE**. We have that $a < c$ and $b = d$. $B_A(0) = \{1\}$, $B_A(1) = \{0, 1\}$ and hence Player A will always play 1. $B_B(0) = \{1\}$, $B_B(1) = \{0, 1\}$ so Player

B will also always play 1 and that means that 11 is deterministic.

□

So if a symmetric 2×2 game has a unique NE then the game is deterministic. This is of course an interesting property in on shot play, but the fact that we cannot determine if a general symmetric game with two NE is deterministic or not is limiting. However, our classification provides enough information about games to separate deterministic and non-deterministic games given that the games belong to exactly one class. To prove this we first need to present Lemma 2.8.3.

Lemma 2.8.3 (Ties on borders)**.** Let G be a symmetric 2×2 game. If G belongs to exactly on class in our classification, then the payoffs of G are distinct.

Proof. In Equation 2.9 we show that the following holds for some constant $\alpha > 0$.

$$\begin{aligned} a &= \frac{\alpha}{4}(2x - 3c_1 - c_2), b = \frac{\alpha}{4}(2x + c_1 - c_2 + 4z_1), \\ c &= \frac{\alpha}{4}(2x + c_1 - c_2 - 4z_1) \text{ and } d = \frac{\alpha}{4}(2x + c_1 + 3c_2). \end{aligned} \tag{2.19}$$

This means that $a = b \Rightarrow z_1 = -c_1$ so any games with $a = c$ lies on the border between at least two classes. In the same way we have:

$a = c \Rightarrow z_1 = c_1$, $a = d \Rightarrow c_1 = -c_2$, $b = c \Rightarrow z_1 = 0$, $b = d \Rightarrow z_1 = c_2$ and $c = d \Rightarrow z_1 = c_2$.

This means that if some pair of payoffs are equal then the game lies on the border between at least two classes.

□

Now we present an important theorem that divides the classes in our model into two categories that separates deterministic and non-deterministic games.

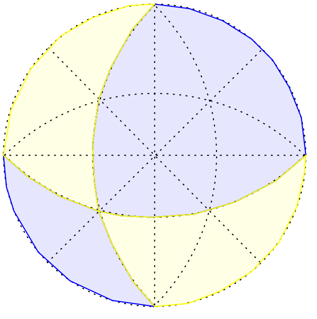
Theorem 2.8.4 (Deterministic games)**.** Let G be a symmetric 2×2 game such that G belong to exactly one class in our classification. Then G is deterministic if and only if G has a unique NE.

Proof. Let G be a game such as described in Theorem 2.8.4. Then Lemma 2.8.3 states that G is a symmetric 2×2 game with distinct payoffs and Proposition 2.8.1 states that G is deterministic iff G has a unique NE.

□

In Figure 2.2 classes with deterministic games are coloured blue and classes with non-deterministic games are coloured yellow. Remember that games on the border between two or more classes are not included in Theorem 2.8.4.

Figure 2.2: Deterministic and non-deterministic classes



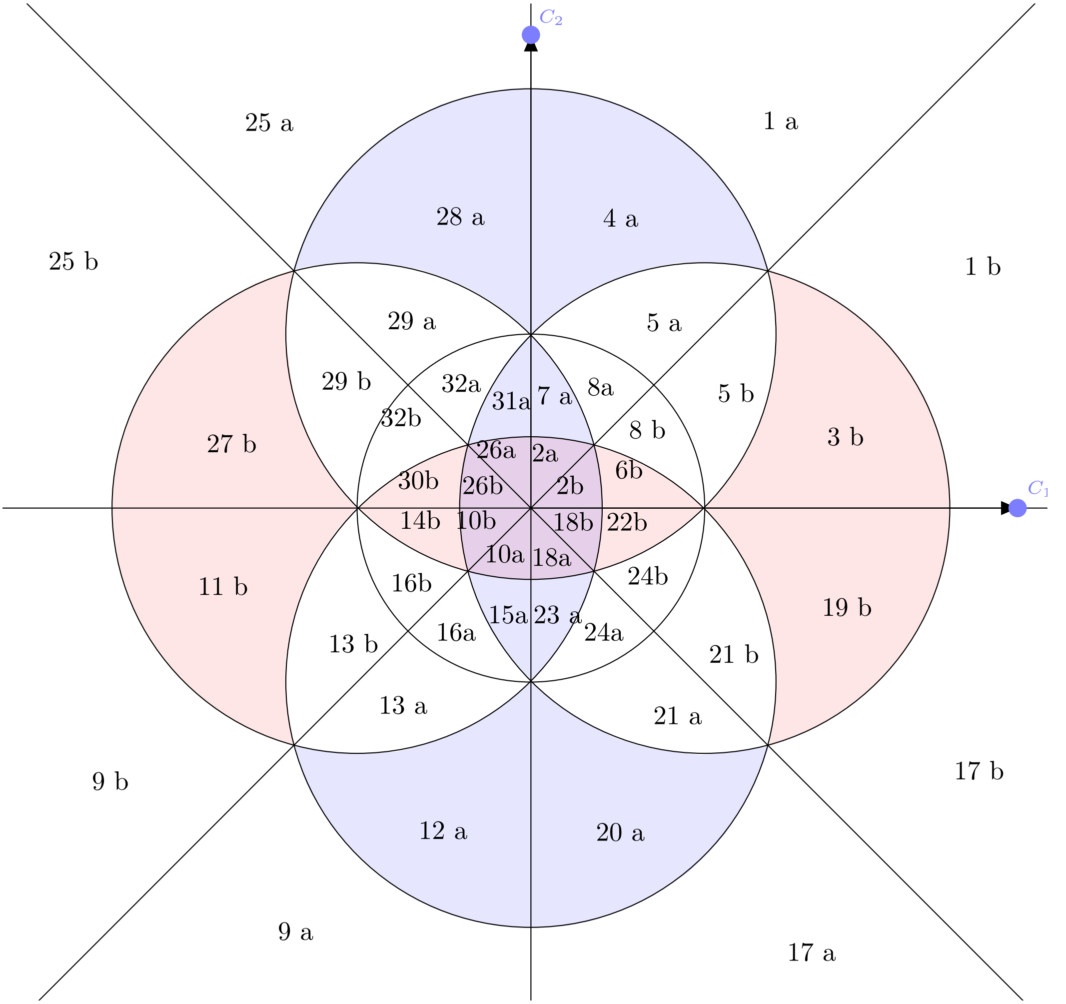
Chapter 3

Discussion

Appendix A

Detailed class description

Here we present at detaild description of the properties of each reagion in the large map received by the stereographic projection.


C_2
C_1
25 a
1 a
28 a
4 a
25 b
1 b
29 a
5 a
29 b
32a
31a
7 a
8a
5 b
32b
8 b
3 b
27 b
26a
2a
6b
30b
26b
2b
14b
10b
18b
22b
10a
18a
24b
11 b
19 b
16b
15a
23 a
24a
13 b
16a
21 b
13 a
21 a
9 b
17 b
12 a
20 a
9 a
17 a

In the illustration of the 48 regions "a" means that $|c_1| \leq |c_2|$ and "b" means that $|c_1| \geq |c_2|$.

1. $c_1 \geq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
2. $c_1 \geq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
3. $c_1 \geq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
4. $c_1 \geq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
5. $c_1 \geq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$
6. $c_1 \geq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
7. $c_1 \geq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
8. $c_1 \geq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$
9. $c_1 \leq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
10. $c_1 \leq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
11. $c_1 \leq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
12. $c_1 \leq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
13. $c_1 \leq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$
14. $c_1 \leq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
15. $c_1 \leq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
16. $c_1 \leq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$
17. $c_1 \geq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
18. $c_1 \geq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
19. $c_1 \geq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
20. $c_1 \geq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
21. $c_1 \geq 0, c_2 \leq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$
22. $c_1 \geq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
23. $c_1 \geq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
24. $c_1 \geq 0, c_2 \leq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$
25. $c_1 \leq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
26. $c_1 \leq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \geq |c_2|$
27. $c_1 \leq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$
28. $c_1 \leq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$
29. $c_1 \leq 0, c_2 \geq 0, z_1 \geq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$

30. $c_1 \leq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \geq |c_2|$

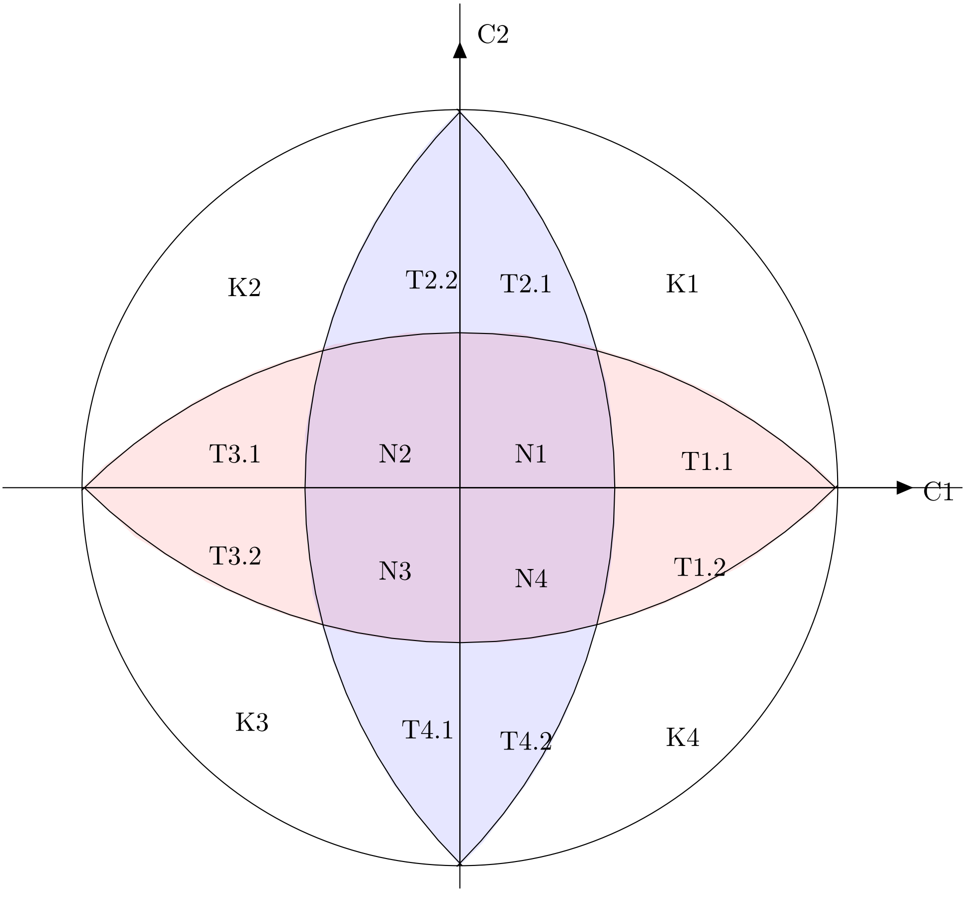
31. $c_1 \leq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \geq |c_1|, |z_1| \leq |c_2|$

32. $c_1 \leq 0, c_2 \geq 0, z_1 \leq 0, |z_1| \leq |c_1|, |z_1| \leq |c_2|$

In the table below we present which pairs of the regions above are equivalent and the name of their combined class. Note that the region on the left hand side and the class of the right hand side of the " $\sim$ " sign have opposite letters a and b.

Class x $\sim$ Class y	Combined class name
$1 \sim 16$	K3
$2 \sim 9$	N1
$3 \sim 15$	T4.1
$4 \sim 14$	T3.2
$5 \sim 10$	N3
$6 \sim 12$	T1.1
$7 \sim 11$	T2.1
$8 \sim 13$	K1
$17 \sim 26$	N2
$18 \sim 25$	N4
$19 \sim 31$	T2.2
$20 \sim 30$	T3.1
$21 \sim 32$	K2
$22 \sim 28$	T1.2
$23 \sim 27$	T4.2
$24 \sim 29$	K4

Table A.1: **Description**


C2
C1
K2
T2.2
T2.1
K1
T3.1
N2
N1
T1.1
T3.2
N3
N4
T1.2
K3
T4.1
T4.2
K4

Linköping University Electronic Press