Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

DIPLOMARBEIT

Best Responses in Repeated Games with Conditional Strategies

ausgeführt am

Institut für
Analysis und Scientific Computing
TU Wien

unter der Anleitung von

Univ.Prof. Dr. Christian Hilbe

und

Ao.Prof.in Dr.in Gabriela Schranz-Kirlinger

und

Dr.in Nikoleta E. Glynatsi

durch

Franziska Lesigang BSc
Matrikelnummer: 12002206

Wien, am 23. Juli 2025

Kurzfassung

Wiederholte Spiele bilden in der Spieltheorie einen grundlegenden Rahmen um zu modellieren, wie sich strategische Interaktionen mit der Zeit entwickeln, wenn Spieler_innen ihre Entscheidungen auf Grundlage vergangener Ereignisse treffen. Zwei besonders interessante Klassen sind reaktive-n-Strategien, bei denen die Wahl der Aktion von den letzten n Aktionen des Gegners abhängt, und selbst-reaktive-n-Strategien, die ausschließlich die eigene Handlungshistorie berücksichtigen. Diese Arbeit behandelt eine zentrale Frage: Welche Informationen benötigt eine Spielerin, Alice, um ihre Auszahlung gegen ihren Gegner, Bob, zu maximieren, wenn dieser eine reaktive-n-Strategie spielt?

Die Studie von [Glynatsi, Akin, Nowak, Hilbe: PNAS, 2024] zeigte, dass Alice in Zwei-Aktionen-Spielen nicht mehr Informationen benötigt als Bob. Zudem kann Alice die für sie maximal mögliche Auszahlung erreichen wenn sie eine pure selbst-reaktive-n-Strategie verwendet. Das erste Ziel dieser Arbeit ist es, dieses Ergebnis auf alle Spiele mit endlich vielen Aktionen zu erweitern. Als unmittelbare Konsequenz ergibt sich, dass das Problem, ob eine reaktive-n-Strategie ein symmetrisches Nash-Gleichgewicht darstellt, entscheidbar ist.

Im zweiten Teil der Arbeit wird gezeigt, dass sich die Interaktion zwischen reaktiven und selbst-reaktiven Strategien als Zyklen auf de-Bruijn-Graphen darstellen lässt. Dieser neue Ansatz reduziert die Komplexität der Entscheidung, ob eine reaktive-n-Strategie ein Nash-Gleichgewicht bildet, erheblich.

Zuletzt betrachten wir eine spezielle Klasse wiederholter Spiele, sogenannte additive Spiele. Unter Verwendung der de-Bruijn-Darstellung zeigen wir, dass Alice in additiven Spielen ihre maximale Auszahlung gegen ihren reaktiven-n-Gegner Bob mit einer puren selbst-reaktiven-$(n-1)$-Strategie erreichen kann. Alice trägt also keinen Schaden, wenn sie sich an ein Ereignis weniger erinnert als Bob. Neben seiner konzeptionellen Bedeutung vereinfacht dieses Ergebnis zusätzlich die Identifikation symmetrischer Nash-Gleichgewichte unter reaktiven-n-Strategien.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Abstract

Repeated games are a foundational framework in game theory, modeling how strategic interactions unfold over time as players make decisions based on prior outcomes. Two classes of interest are reactive-n strategies, which respond to the co-player's last n actions, and self-reactive-n strategies, which consider only the player's own history. This thesis addresses a key question: What information does one player, Alice, need to maximize her payoff against her opponent, Bob, if he plays a reactive-n strategy?

Previous work by [Glynatsi, Akin, Nowak, Hilbe: PNAS, 2024] established for two action games, that Alice requires no more information than is available to Bob. Moreover, Alice can receive the highest possible payoff when she uses a pure self-reactive-n strategy. The first goal of this thesis is to extend this result to all finite action games. As an immediate consequence, the problem of determining whether a reactive-n strategy is a symmetric Nash equilibrium is decidable.

In the second part of the thesis, we observe that the interaction between these two classes of strategies, reactive and self-reactive, can be described as cycles on de Bruijn graphs. This new approach drastically reduces the complexity of deciding if a reactive-n strategy is a Nash equilibrium.

Finally, the thesis focuses on a specific class of repeated games called additive games. Employing the de Bruijn graph representation, we prove that in additive games, Alice can receive her highest possible payoff against her reactive-n opponent Bob by playing a pure self-reactive-$(n-1)$ strategy. Thus, Alice is allowed to remember less than Bob without sacrificing payoff. Apart from being noteworthy on a conceptual level, this result simplifies the task of identifying symmetric Nash equilibria among reactive-n strategies even further.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Danksagung

Diese Arbeit wäre ohne die Unterstützung vieler Menschen nicht möglich gewesen. Mein besonderer Dank gilt an erster Stelle meinen Betreuenden Nikoleta Glynatsi, Christian Hilbe und Gabriela Schranz-Kirlinger. Ich danke ihnen nicht nur für die fachliche Begleitung und das konstruktive Feedback während der Entstehung dieser Arbeit, sondern auch für die wertschätzende und verlässliche Betreuung darüber hinaus.

Ich bedanke mich bei Nikoleta Glynatsi, die sich immer Zeit für meine vielen Fragen genommen und unermüdlich bei der Korrektur meiner Arbeiten geholfen hat. Ihr offenes Ohr, ihre Geduld und ihr Engagement sind für mich eine große Unterstützung und Motivation. Christian Hilbe danke ich besonders dafür, dass er mir die Möglichkeit gegeben hat, mein Praktikum am Max-Planck-Institut in Plön zu absolvieren. Er hat mich nicht nur bei all meinen Arbeiten zuverlässig unterstützt, sondern mir auch von Anfang an großes Vertrauen entgegengebracht. Ich bin beiden dankbar dafür, dass sie mir die Tür zur Welt der Wissenschaft geöffnet und mich ermutigt haben, meinen eigenen Weg darin zu beginnen.

Gabriela Schranz-Kirlinger möchte ich danken – für ihre jahrelange Begleitung während meines Studiums und dafür, dass ich Teil von fem*ma sein durfte. In diesem Zusammenhang gilt mein Dank auch all den inspirierenden Frauen bei fem*ma, die mich in vielerlei Hinsicht geprägt und bestärkt haben.

Ein großer Dank gilt auch all meinen Freundinnen und Freunden, die mich über die Jahre meines Studiums hinweg begleitet und mein Leben weit über das Studium hinaus mitgeprägt haben. Catherine, Steffi, Julian, Christof, Godwin, Flo, Johannes, Maxi, Maxi, Paul, Alex, David, Mario, Henner und viele weitere. An dieser Stelle möchte ich mich auch bei Martin Rubey bedanken, der uns immer unterstützt hat und uns einen Ort an der Universität gegeben hat, an dem wir arbeiten, lernen und zusammenkommen konnten.

Ganz besonders danke ich Chris, für seine stetige Unterstützung und seine Ruhe durch alle Höhen und Tiefen.

Ich möchte Mama, Papa, Klemens und meiner ganzen Familie von Herzen danken – für ihre Unterstützung, ihr Vertrauen und ihren Rückhalt, besonders in einer für uns alle schwierigen Zeit.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Hilfsmittel der künstlichen Intelligenz wurden nur verwendet, um sprachliche Formulierungen und Zeichensetzung Korrektur zu lesen und gegebenenfalls zu verbessern.

Wien, am 23. Juli 2025

Franziska Lesigang

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Contents

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1 Introduction

Decisions are an unavoidable part of life. Naturally, the question arises: Which decision is best? To approach this problem theoretically, consider a *game* defined by *actions*, *players* and *payoffs*. Each player selects an action, and based on these choices, each receives a corresponding scalar-valued payoff. It is assumed that every player strives to maximize their own payoff.

A famous example of such a game is the *prisoner's dilemma*. In the most common version by [AH81] two players are considered, who can both choose to either cooperate (by playing action C) or defect (by playing action D). If both decide to cooperate, the payoff is 3, if both defect, the payoff is 1. However, if one player decides to cooperate and the other to defect, the defective player receives the payoff 5 while the cooperative player leaves empty handed, i.e., with the payoff 0. It is best for a player to choose defection, independently of what the opponent chooses. Interestingly, this dynamic changes, if the game is played iteratively. In simpler words, exploiting an opponent is only worthwhile if you are not going to face them again. The theory of *repeated games* addresses this observation in a general setting, thus becoming a foundational framework in game theory ([MS06]). It allows the study of complex behaviors such as cooperation and direct reciprocity ([Now06], [GV16]).

In repeated games, the underlying game, that is played iteratively, possibly even infinitely often, is referred to as the *stage game*. This thesis focuses on repeated games *without discounting*, meaning the stage game is played infinitely often. Naturally, the concept of *history* arises. It is the union of all stage games that have already been played. It is rationally sound for a player to base their decisions on the history, or at least on parts of it. The subset of the game's history, on which the player's action depends on, is referred to as the player's *memory* at a given stage game. The size of the memory is defined as the distance from the current stage game to the earliest stage game the player recalls. Players can then be classified by their maximum memory size. Particularly cognitively plausible are players with bounded memory size ([Ued21], [MO11]). One important class of bounded memory players, that have been the focus of many evolutionary studies, are *memory-n strategies* ([HMCN17], [RR18]). These are players whose memory at any stage consists for the past $n \in \mathbb{N}$ games. This is equivalent to a player having a memory size of n. A subclass of memory-n strategies are reactive-n strategies, those that solely depend on their opponents last n actions ([GANH24]). In many real-life situations, it is evident that individuals tend to place greater emphasis on how they are treated rather than on how they treat others. This observation is modeled through reactive-n strategies. Throughout this thesis it is assumed that opponents use reactive-n strategies.

After having established the player's abilities, the question is raised: How powerful is memory? Can a player with a larger memory size achieve a higher payoff? In [PD12],

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

it is shown that in the repeated prisoner's dilemma, if the opponent uses a memory-n strategy, a player can always achieve the highest possible payoff using a strategy with the same memory size. This result is further generalized in [LNZ20] to more general strategy spaces and game types. Building on this, [GANH24] improves the result for reactive-n strategies in 2 action games. For any reactive-n opponent, one can construct a self-reactive-n strategy that achieves the maximum payoff against this opponent. These strategies recall only the player's own past n actions and match the opponent's memory size.

A strategy that achieves the highest possible payoff against a fixed opponent, is referred to as a *best response*. Best responses are especially of interest as they are fundamental in the identification of *Nash Equilibria*, a central concept in game theory [Nas50]. A Nash Equilibrium is a state where no player can improve their payoff by unilaterally switching to another strategy. All results in this thesis contribute to reducing the set that contains best responses to any arbitrary reactive-n strategy. As a consequence, this reduction simplifies the task of identifying Nash equilibria.

The first reduction is obtained by extending the result of [GANH24] to all games in which players have finitely many actions to choose from, rather than just two. We constructively prove the existence of a self-reactive-n strategy that is a best response to a fixed but arbitrary reactive-n opponent. This allows us to reduce the set of strategies that contains best responses to all reactive-n strategies from the set of all possible strategies to the smaller set of all self-reactive-n strategies. Secondly, we further reduce this to a *finite* set by proving the existence of a best response among all deterministic (or *pure*) self-reactive-n strategies. Deterministic strategies are those that choose to cooperate or defect with probability 1 or 0, respectively. As before, the proof is constructive and is inspired by [GANH24] and [PD12].

Having reduced the search for a best response to a finite set, it is now possible to computationally solve the problem. However, the number of pure self-reactive-n strategies grows exponentially with the memory size. This motivates the search for additional reductions. To this end, we factorize the set of all pure self-reactive-n strategies by the payoff they receive against an arbitrary opponent in an arbitrary game. More precisely, two strategies are called *equivalent* if and only if they receive the same payoff against every opponent in every game. While this may initially appear to be an insignificant reduction, the thesis goes on to challenge and ultimately contradict this assumption.

The key step is to map the histories of pure self-reactive-n strategies as vertices on a graph, and connecting histories that immediately succeed each other by a directed edge. Based on this mapping, we observe that every pure self-reactive-n strategy's moves eventually reach a cycle on this graph and follow it indefinitely. If two strategies eventually proceed along the same cycle, they receive the same payoff in any game against any reactive opponent. This holds regardless of their (possibly different) moves prior to entering the cycle. Following this observation, we provide a bijection between the equivalence classes of pure self-reactive-n strategies and all possible cycles in the graph. Interestingly, the graph that precisely contains all such cycles is the *de Bruijn graph*, a well-studied structure in graph theory with numerous applications, for example in genetics [CPT11]. This result establishes a

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

novel connection between de Bruijn graph theory and game theory, leading to a noteworthy side result. While a closed-form expression for the exact number of cycles in the de Bruijn graph, and thus the number of equivalence classes of strategies, is not known, numerical results indicate that the reduction is substantial.

Not only do the factorization of pure self-reactive strategies and their bijection to cycles on the de Bruijn graph offer a substantial reduction, they also open the door to applying graph-theoretical tools in the study of repeated games. This perspective allows us to address another natural and intriguing question: Can a player remember less than their opponent without sacrificing payoff? Surprisingly, under two simple conditions, the answer is yes. The two conditions are that the opponent uses a reactive-n strategy, and that the game is *additive*, also known as a game with *equal gains from switching* [NS90]. In that case, a player can use less memory and still achieve the maximum possible payoff. Additive games have played a key role in evolutionary game theory [MFH14; MRH21; CP23]. They are defined by the property that each player's payoff can be decomposed into components depending solely on the player's own action and that of their opponent. In this setting, we prove that for every reactive-n opponent, there exists a best response strategy that is pure self-reactive-$(n-1)$. This result is both conceptually significant and practically useful, as it dramatically reduces the number of strategies that need to be explored to find a best response.

In summary, this thesis presents four significant reductions in the search for best responses when the co-player uses a reactive strategy. Along the way, it introduces new methods that may have broader impact beyond the immediate results. The final main result has already been published [LHG25], while the others are yet to follow.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

2 Game Theory, Repeated Games and Best Responses

In this chapter, we introduce essential concepts that will be the main focus of this thesis. First, we mathematically define games, repeated games, and histories of games, see Section 2.1. In Section 2.2, we explore strategies and analyze how different types of strategies relate to each other. We then define payoffs in repeated games and study their existence in Section 2.3. Finally, in Section 2.4 we characterize best responses and discuss their significance in game theory.

2.1 Games and Repeated Games

We model the interaction between two people, who we will refer to as players. We assume that players obtain a value from this interaction or *game*, as we continue to call it. Moreover, we postulate that players strive for a high payoff. The payoff is solely dependent on the actions of both players. We assume both players to choose one out of $m \in \mathbb{N}$ actions. The set of actions is denoted by $A := \{A_1, \dots, A_m\}$. We focus on one player, whose payoff we aim to study, and refer to them as the *focal player*. The player opposed to the focal player is referred to as the *opponent player*, or in short *opponent*. The *payoff matrix*

$$G := \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{mm} \end{pmatrix} \in \mathbb{R}^{m \times m} \tag{2.1}$$

defines the game. For example, if the focal player chooses action A_i and their opponent chooses action A_j, then the focal player receives the payoff $g_{ij} \in \mathbb{R}$. We focus on *symmetric games*, where the opponent's payoff matrix is the payoff matrix of the focal player. Thus, the opponent's payoff can be calculated by considering the transpose matrix. In our example, where the focal player and opponent choose action A_i and A_j, respectively, the opponent receives the payoff g_{ji} when the game is symmetric. In an *asymmetric game*, on the other hand, the payoff matrices of each player do not need to bear any relation to one another. The two payoff matrices are often represented as one payoff bimatrix, where each entry is a 2-tuple. The first entry refers to the payoff of the focal player, while the second defines the payoff of the opponent. While this thesis focuses on symmetric games, all results can be generalized to asymmetric games.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Example 2.1. The prisoner's dilemma as introduced in Chapter 1 is a symmetric game. It is defined by the set of actions $A := \{\mathrm{C}, \mathrm{D}\}$ and the payoff matrix

$$\begin{pmatrix} R & S \\ T & P \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} (R,R) & (S,T) \\ (T,S) & (P,P) \end{pmatrix}$$

is the representation of the payoff matrix in a form that is more commonly used in asymmetric games. However, here it allows us to observe the symmetry of the game and provides insight into why the transpose of the payoff matrix is of interest in symmetric games.

The focus of the thesis at hand is the study of *repeated games*. That is, we assume the two players engage in a fixed game iteratively. We refer to the game that is played at every iteration as the *stage game*. *Game* will from now on denote the whole process of the repeated game. We introduce the *game's history* at each step of the iteration. Formally, we define the history of the game after $t \in \mathbb{N}$ steps as an element of

$$H_t := (A \times A)^t$$

which maps $k \in \{0, \dots, t-1\}$ to the actions that both players chose in the k-th game. Furthermore, we denote by H_t^i the set of the focal player's own past t actions and by H_t^{-i} the opponent's past t actions. Mathematically, H_t^i and H_t^{-i} are equivalent to A^t.

2.2 Strategies

In each round of the repeated game players follow a rule to determine their next action. We refer to this rule as the player's *strategy*. This strategy depends on the history of the game so far, and it outputs a distribution over the set of actions.

Definition 2.2. *A* strategy σ *is a sequence of functions, such that for all* $t \in \mathbb{N}$

$$\sigma_t \in \{f \mid f : H_t \to \Delta^{m-1}\}, \tag{2.2}$$

where $\Delta^{m-1} := \left\{\boldsymbol{x} \in \mathbb{R}^m_{\geq 0} \mid \sum_{i=1}^m x_i = 1\right\}$. *That is, at step* $t \in \mathbb{N}$ *the strategy* σ *maps each history* $\boldsymbol{h} \in H_t$ *to a distribution over the action space* A. *The opponent's strategy is defined similarly.*

Remark 2.3. *In this thesis, we identify the set* $\{f \mid f : H_0 \to \Delta^{m-1}\}$ *with the set* Δ^{m-1}.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

We are particularly interested in strategies with bounded recall.

Definition 2.4. *A strategy σ is a* finite memory strategy *if*

$$\begin{aligned} &\exists n \in \mathbb{N} : \forall t, \hat{t} \in \mathbb{N}_{\geq n} \forall \boldsymbol{h} \in H_t \forall \hat{\boldsymbol{h}} \in H_{\hat{t}} : \\ &P_n(\boldsymbol{h}) = P_n(\hat{\boldsymbol{h}}) \Rightarrow \sigma_t(\boldsymbol{h}) = \sigma_{\hat{t}}(\hat{\boldsymbol{h}}), \end{aligned} \tag{2.3}$$

where P_n is the projection onto the last n components. If the input has fewer than n components, then P_n acts as the identity function. We refer to the set of all finite memory strategies as Π.

Recall that for every $t \in \mathbb{N}$, σ_t is a function that takes a history and maps it to a distribution over the action set. Thus, (2.3) refers to a strategy that is only dependent on a fixed number of previous moves and is independent of the time at which the moves were observed.

A subset of finite memory strategies that has been of particular interest in evolutionary and classical game theory are *memory-n strategies*. In contrast to the definition used in other literature, see [GANH24], we include the initial n moves in the definition. While the initial moves will play a crucial role in upcoming chapters, the idea of memory-n strategies is kept in tact. Note that memory-n strategies are not subsets of finite memory strategies by their mathematical definition. However, Lemma 2.6 introduces an embedding from memory-n strategies into finite memory strategies, allowing us to interpret the former as a subset of the latter.

Definition 2.5. *A memory-n strategy $\boldsymbol{\sigma}$ is defined as a 2-tuple*

$$\boldsymbol{\sigma} = (\sigma^{<n}, \sigma) \in A^n \times \{f \mid f : H_n \to \Delta^{m-1}\}. \tag{2.4}$$

The first entry $\sigma^{<n}$ refers to the moves the player takes before a history of size n is observed. After n games, the actions the player takes are dependent on the second entry σ. We denote the set of all memory-n strategies by M_n.

From a cognitive perspective, it is natural to require that any memory-k strategy is also a memory-n strategy for $n \geq k$. This implies that the sets M_n, for $n \in \mathbb{N}$, should be nested under inclusion. This monotonicity, however, is not ensured by Definition 2.5. However, with the following embedding we are able to consider them as a nested sequence of sets. The second embedding ensures, that all memory-n strategies can be considered as finite memory strategies. This is visualized in Figure 2.1.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

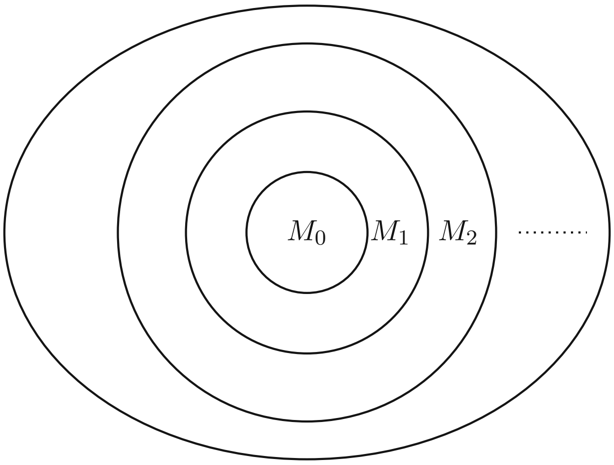


Figure 2.1: The embeddings found in Lemma 2.6 structure the sets of memory-n strategies for increasing $n \in \mathbb{N}$. The figure visualizes that the set of memory-n strategies form a monotonically increasing series of subsets that are all contained in Π.

Lemma 2.6. *For any $n, N \in \mathbb{N}$ with $n \leq N$ there exists an embedding $\iota_{n,N}$ from M_n to M_N and an embedding ι_n from M_n to Π, which satisfy the following properties for all $k, n, N \in \mathbb{N}$ with $k < n < N$:*

1. $\iota_{k,N} = \iota_{n,N} \circ \iota_{k,n}$ *and*
2. $\cup_{n \in \mathbb{N}} \iota_n(M_n) \subseteq \Pi$

Proof. We fix arbitrary $n, N \in \mathbb{N}$ and define $\iota_{n,N}((\sigma^{<n}, \sigma)) := (\tilde{\sigma}^{<N}, \tilde{\sigma})$, where $\tilde{\sigma} := \sigma \circ P_n$ and $(\tilde{\sigma}^{<N})_t$ for $t \in \{0, \ldots, N-1\}$ is iteratively defined by

$$(\tilde{\sigma}^{<N})_t = \begin{cases} A_1 & \text{for } t < N - n \\ (\sigma^{<n})_{t-N+n} & \text{else.} \end{cases}$$

To prove that $\iota_{n,N}$ is injective, we fix two arbitrary memory-n strategies $\boldsymbol{\sigma_1}$ and $\boldsymbol{\sigma_2}$ and assume that $\iota_{n,N}(\boldsymbol{\sigma_1}) = \iota_{n,N}(\boldsymbol{\sigma_2}) = \boldsymbol{\tilde{\sigma}}$. By the definition of $\iota_{n,N}$ we observe that $\sigma_1 \circ P_n = \sigma_2 \circ P_n$. Since $P_n = id$ on the domain H_n, we obtain that $\sigma_1 = \sigma_2$. The observation, that $\sigma_1^{<n} = (\tilde{\sigma}_i^{<N})_{i=N-n}^{N-1} = \sigma_2^{<n}$, concludes $\boldsymbol{\sigma_1} = \boldsymbol{\sigma_2}$.

Consider $k, n, N \in \mathbb{N}$ where $k < n < N$. With $P_k = P_n \circ P_k$ it follows immediately that $\iota_{k,N} = \iota_{n,k} \circ \iota_{k,n}$.

Furthermore, we define $\iota_n(\boldsymbol{\sigma}) := (\tilde{\sigma}_t)_{t \in \mathbb{N}}$, where

$$\tilde{\sigma}_t := \begin{cases} \sigma \circ P_n & \text{for } t \geq n \\ \delta_{(\sigma^{<n})_t} & \text{else.} \end{cases}$$

Note that $(\sigma^{<n})_t$ is an element of the action set A and δ is defined as the Kronecker delta on A. For every $\tilde{n} \in \mathbb{N}$ the function $\iota_{\tilde{n}}$ maps into Π as condition (2.3) holds for

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

$n = \tilde{n}$. The injection of ι_n is shown similarly to the injection of $\iota_{n,N}$. The property $\cup_{n\in\mathbb{N}}\iota_n(M_n) \subseteq \Pi$ follows immediately. □

The property of the embeddings $\iota_{\cdot,\cdot}$ ensure their commutativity. Thus, it is irrelevant if a memory-k strategy is embedded in M_N or first embedded in M_n and afterwards in M_N. This is visualized in Figure 2.2.

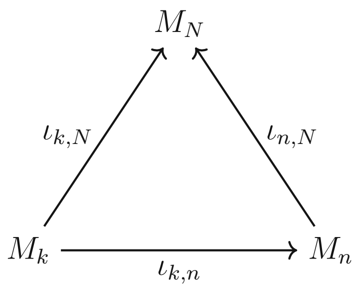


Figure 2.2: The embeddings obtained in Lemma 2.6 commute for $k < n < N \in \mathbb{N}$.

We further define a subset of memory-n strategies that are solely dependent on the opponent's actions.

Definition 2.7. *A reactive-n strategy* $\boldsymbol{p}$ *is defined as a 2-tuple*

$$\boldsymbol{p} = (p^{<n}, p) \in A^n \times \{f \mid f : H_n^{-i} \to \Delta^{m-1}\}. \tag{2.5}$$

We denote the set of all reactive-n strategies by R_n*.*

Equally, we define the subset of memory-n strategies that solely depend on the player's own actions.

Definition 2.8. *A self-reactive-n strategy* $\boldsymbol{q}$ *is defined as a 2-tuple*

$$\boldsymbol{q} = (q^{<n}, q) \in A^n \times \{f \mid f : H_n^{i} \to \Delta^{m-1}\}. \tag{2.6}$$

We denote the set of all self-reactive-n strategies by S_n*.*

Definition 2.9. *Any strategy and any sets of strategies is said to be* pure *if we replace* Δ^{m-1} *by the set of all unit vectors of* $\mathbb{R}^m$*.*

Example 2.10. In the repeated prisoner's dilemma, the famous strategy *Tit-for-Tat*, popularized by [AH81], is an example of a memory-1 strategy. A player using *Tit-for-Tat* initially chooses cooperation C. Then, the player copies the last move of the opponent. Thus, the strategy is even a pure reactive-1 strategy. Formally, *Tit-for-Tat* is defined by $\boldsymbol{p} := ((\mathrm{C}), p)$, where

$$p(\mathrm{C}) = (1,0)^T$$
$$p(\mathrm{D}) = (0,1)^T.$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

2.3 Calculation of Payoffs

Given two arbitrary strategies σ and $\tilde{\sigma}$, let $\pi_{\sigma,\tilde{\sigma}}(t)$ denote the focal player's expected payoff in round t.

Definition 2.11. *We define the* repeated-game payoff *as the limiting average,*

$$\pi(\sigma,\tilde{\sigma}) := \lim_{\tau\to\infty} \frac{1}{\tau}\sum_{t=1}^{\tau} \pi_{\sigma,\tilde{\sigma}}(t). \tag{2.7}$$

We note that the limit (2.7) does not need to exist. For a fixed opponent $\boldsymbol{p}$, we refer to the set of strategies, for which the limit (2.7) does exist, with $\Sigma_{\boldsymbol{p}}$. The following approach shows that if the strategies $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\sigma}}$ are memory-n then the limit converges. By Lemma 2.6, it follows that the limit exists for the interaction between any two finite memory strategies.

To compute the payoff of the focal player in the interaction of two memory-n strategies, we represent the game as a Markov chain. To this end, we choose the states of the Markov chain as the game's n-histories $\boldsymbol{h}\in H_n$.

The probability of transitioning from state $\boldsymbol{h}$ to state $\tilde{\boldsymbol{h}}$ is then given by

$$M_{\boldsymbol{h},\tilde{\boldsymbol{h}}} = m_\sigma \cdot m_{\tilde{\sigma}}. \tag{2.8}$$

Here, m_σ is defined as

$$m_\sigma = \begin{cases} \sigma(\boldsymbol{h})_i & \text{if } \boldsymbol{h}_k = \tilde{\boldsymbol{h}}_{k-1}\ \forall\ k\in\{2,\dots,n\} \text{ and } \tilde{h}_n^1 = A_i, \\ 0 & \text{otherwise,} \end{cases}$$

and $m_{\tilde{\sigma}}$ similarly. For $t\geq n$, let $\boldsymbol{v}(t)\in\Delta^{|H_n|-1}$ denote the probability distribution of observing each history in round t. For a given distribution $\boldsymbol{v}(t)$, the next round's distribution is computed as $\boldsymbol{v}(t+1)=\boldsymbol{v}(t)M$. Since M is a stochastic matrix, the Perron–Frobenius theorem ensures that $\boldsymbol{v}(t)$ converges to a limiting distribution $\boldsymbol{v}=(v_{\boldsymbol{h}})_{\boldsymbol{h}\in H_n}$, which is a left eigenvector of M with eigenvalue 1. If M is primitive, this limiting distribution is unique; otherwise it is uniquely determined by the outcome of the first n rounds.

We can use this insight to compute the players' payoffs, by noting that the focal player's expected payoff in round $t\geq n$ can be written as

$$\pi_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}(t) = \boldsymbol{v}(t)\cdot\boldsymbol{u}.$$

Here, $\boldsymbol{u}=(u_{\boldsymbol{h}})$ is the vector that assigns to every n-history the focal player's latest stage payoff. That is, if the history $\boldsymbol{h}=(\boldsymbol{h}^1,\boldsymbol{h}^2)$ is such that $h_n^1=A_i$ and $h_n^2=A_j$,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

then $u_{\boldsymbol{h}} = g_{ij}$. If $\boldsymbol{v}(t) \to \boldsymbol{v}$ for $t \to \infty$, then so does $\frac{1}{\tau}\sum_{t=n}^{\tau} \boldsymbol{v}(t)$. We obtain

$$\pi(\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}(t) = \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=n}^{\tau} (\boldsymbol{v}(t) \cdot \boldsymbol{u})$$
$$= \Big(\lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=n}^{\tau} \boldsymbol{v}(t) \Big) \cdot \boldsymbol{u} = \boldsymbol{v} \cdot \boldsymbol{u}.$$

2.4 Best Responses

In this thesis, we are interested in determining the minimal memory size of a strategy that serves as the best response against another strategy of a given memory size. We focus specifically on reactive-n strategies. From this point onward, unless stated otherwise, we assume that the opponent employs a reactive-n strategy. Given that reactive-n strategy $\boldsymbol{p}$, our goal is to identify a *best response* within the strategy set $\Sigma_{\boldsymbol{p}}$. Specifically, we seek a strategy $\sigma \in \Sigma_{\boldsymbol{p}}$ such that

$$\pi(\sigma, \boldsymbol{p}) \geq \pi(\tilde{\sigma}, \boldsymbol{p}) \quad \text{for all } \tilde{\sigma} \in \Sigma_{\boldsymbol{p}}. \tag{2.9}$$

Note that this definition does not imply the existence of a best response. In the course of this thesis, we demonstrate that, in our setting, a best response always exists. Best responses are important in the search for *symmetric Nash equilibria*, which are a central concept in evolutionary game theory, see [HR04]. Given a symmetric game, i.e., both players have the same action sets and payoff matrix, we call a strategy σ a *symmetric Nash equilibrium*, if for all $\tilde{\sigma} \in \Sigma_{\sigma}$

$$\pi(\sigma, \sigma) \geq \pi(\tilde{\sigma}, \sigma). \tag{2.10}$$

In other words, a strategy is a symmetric Nash equilibrium if it is a best response to itself. Such strategies provide a sense of stability, as each player has no incentive to deviate when both adopt the same strategy. Every symmetric stage game has a symmetric Nash equilibrium, see [Sig10]. This implies the existence of a symmetric Nash equilibrium in repeated games, as unconditionally playing the stage game's symmetric Nash equilibrium is a symmetric Nash equilibrium in the repeated game. In this thesis, we focus exclusively on symmetric Nash equilibria, disregarding general, asymmetric Nash equilibria. Accordingly, any reference to Nash equilibria throughout the text should be understood as referring specifically to symmetric Nash equilibria.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

3 Pure Self-Reactive Best Responses in m-Action Games

In this chapter, we generalize the work of [GANH24] to all games with finitely many actions. More specifically, we prove that if one player adopts a reactive-n strategy, then the other player can always find a best response within the set of pure self-reactive-n strategies, denoted by S_n^{pure}. While this result was previously established in [LNZ20], we take a different approach. The result is stated as follows.

Theorem 3.1. *Let $\boldsymbol{p} \in R_n$ be a reactive-n strategy. Then, there exists $\tilde{\boldsymbol{p}} \in S_n^{pure}$ with*

$$\pi(\tilde{\boldsymbol{p}}, \boldsymbol{p}) \geq \pi(\sigma, \boldsymbol{p}) \quad \text{for all } \sigma \in \Sigma_{\boldsymbol{p}}. \tag{3.1}$$

To prove this theorem, we proceed in two steps. First, we construct for every strategy a self-reactive-n strategy that receives the same payoff against a fixed but arbitrary reactive-n strategy (Proposition 3.2). Then, we demonstrate that every self-reactive-n strategy is weakly dominated by a pure self-reactive-n strategy (Proposition 3.4). These results and their proofs are presented in the following subsections, namely Section 3.1 and Section 3.2. Combining the results proves not only the existence of a best response, but also Theorem 3.1.

3.1 Self-Reactive Best Responses

Proposition 3.2. *Let $\boldsymbol{p} \in R_n$ and $\sigma \in \Sigma_{\boldsymbol{p}}$ be arbitrary but fixed. There exists $\boldsymbol{q} \in S_n$ such that*

$$\pi(\boldsymbol{q}, \boldsymbol{p}) = \pi(\sigma, \boldsymbol{p}). \tag{3.2}$$

It follows that if a best response against a reactive-n strategy $\boldsymbol{p}$ exists in $\Sigma_{\boldsymbol{p}}$, then a best response also exists in S_n.

Proof. A strategy $\boldsymbol{q} = (q^{<n}, q)$ is a self-reactive-n strategy if and only if q only depends on the player's own last n-moves. Since the histories in the first n games are observed at most once, σ's action choices in these rounds can be interpreted as deterministic. It is thus possible to define $q^{<n}$ as σ's first n moves.

We consider the reactive-n strategy $\boldsymbol{p} \in R_n$ and the strategy $\sigma \in \Sigma_{\boldsymbol{p}}$, where σ is the focal player and $\boldsymbol{p}$ the opponent player. Since $\sigma \in \Sigma_{\boldsymbol{p}}$, the limiting distribution

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

$\boldsymbol{v} = (v_{\boldsymbol{h}})_{\boldsymbol{h}\in H_n}$ is well-defined by

$$v_{\boldsymbol{h}} := \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=1}^{\tau} v_{\boldsymbol{h}}(t). \tag{3.3}$$

For a fixed history $\boldsymbol{h}^i \in H_n^i$ we implicitly define q of the self-reactive-n strategy $\boldsymbol{q} = (q^{<n}, q)$ via

$$\left(\sum_{\boldsymbol{h}^{-i}\in H_n^{-i}} v_{(\boldsymbol{h}^i,\boldsymbol{h}^{-i})}\right) q(\boldsymbol{h}^i) = \sum_{\boldsymbol{h}^{-i}\in H_n^{-i}} \left(\lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=1}^{\tau} v_{(\boldsymbol{h}^i,\boldsymbol{h}^{-i})}(t)\sigma_t((\boldsymbol{h}^i,\boldsymbol{h}^{-i}))\right). \tag{3.4}$$

For all $\boldsymbol{h}^i \in H_n^i$, $\boldsymbol{h}^{-i} \in H_n^{-i}$, and $t \in \mathbb{N}$ we note that $\sigma_t((\boldsymbol{h}^i, \boldsymbol{h}^{-i}))$ is a vector in m-dimensions and thus, so is $q(\boldsymbol{h}^i)$. The limit in the right hand side of (3.4) exists due to (3.3). Additionally, if the sum in the left hand side of (3.4) is zero, then so must be the right hand side. In this case, $q(\boldsymbol{h}^i)$ can be arbitrarily chosen in Δ^{m-1}. We aim to show that the limiting distribution $\boldsymbol{v}$ of the strategy σ playing against the reactive-n strategy $\boldsymbol{p}$ is an eigenvector of the transition matrix M obtained by (2.8) and the self-reactive-n strategy $\boldsymbol{q}$ defined by (3.4) with appropriate initial moves $q^{<n}$. For simplicity, we show $\boldsymbol{v} = \boldsymbol{v}M$ for the history $\boldsymbol{h}_{A_1}$ in which the players both respond with action A_1. We define for $j \in \{i, -i\}$

$$e(\boldsymbol{h}^j, \tilde{\boldsymbol{h}}^j) = \begin{cases} 1 & \text{if } \boldsymbol{h}_k^j = \tilde{\boldsymbol{h}}_{k-1}^j \text{ for all k } \in \{2, \dots, n\}, \\ 0 & \text{else.} \end{cases}$$

For $t \geq n$ we obtain

$$v_{\boldsymbol{h}_{A_1}}(t+1) = \sum_{\boldsymbol{h}^i\in H_n^i} \sum_{\boldsymbol{h}^{-i}\in H_n^{-i}} v_{(\boldsymbol{h}^i,\boldsymbol{h}^{-i})}(t)\ (\sigma_t(\boldsymbol{h}^i, \boldsymbol{h}^{-i}))_1\ p(\boldsymbol{h}^i)_1\ e(\boldsymbol{h}^i, \boldsymbol{h}^i_{A_1})\ e(\boldsymbol{h}^{-i}, \boldsymbol{h}^{-i}_{A_1}).$$

By summing up the equation from $t = 1, \dots, \tau$ and dividing by τ results in

$$\frac{1}{\tau}\sum_{t=1}^{\tau} v_{\boldsymbol{h}_{A_1}}(t+1) = \sum_{\boldsymbol{h}^i\in H_n^i} \sum_{\boldsymbol{h}^{-i}\in H_n^{-i}} \left(\frac{1}{\tau}\sum_{t=1}^{\tau} v_{(\boldsymbol{h}^i,\boldsymbol{h}^{-i})}(t)\ (\sigma_t(\boldsymbol{h}^i, \boldsymbol{h}^{-i}))_1\right)\ p(\boldsymbol{h}^i)_1\ e(\boldsymbol{h}^i, \boldsymbol{h}^i_{A_1})\ e(\boldsymbol{h}^{-i}, \boldsymbol{h}^{-i}_{A_1}).$$

Taking the limit leads to

$$v_{\boldsymbol{h}_{A_1}} = \sum_{\boldsymbol{h}^i\in H_n^i} \sum_{\boldsymbol{h}^{-i}\in H_n^{-i}} v_{(\boldsymbol{h}^i,\boldsymbol{h}^{-i})}\ \left(p(\boldsymbol{h}^i)_1\ e(\boldsymbol{h}^i, \boldsymbol{h}^i_{A_1})\right)\ \left(q(\boldsymbol{h}^{-i})_1\ e(\boldsymbol{h}^{-i}, \boldsymbol{h}^{-i}_{A_1})\right). \tag{3.5}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

With (2.8) this results in

$$v_{\boldsymbol{h}_{A_1}} = \sum_{\boldsymbol{h} \in H_n} v_{\boldsymbol{h}} M_{\boldsymbol{h},\boldsymbol{h}_{A_1}}, \tag{3.6}$$

where M is the transition matrix obtained by $\boldsymbol{q}$ playing against $\boldsymbol{p}$. □

We have now shown that given a reactive-n strategy $\boldsymbol{p}$ it suffices to test $\boldsymbol{p}$ against all self-reactive-n strategies to verify (2.9). This requires calculating the payoff $\pi(\boldsymbol{q}, \boldsymbol{p})$ for self-reactive-n strategies $\boldsymbol{q}$. However, in this setting, both players solely consider the self-reactive-n player's previous actions to make their decision. Thus, the transition matrix M can be reduced to a $m^n \times m^n$ dimensional matrix $\tilde{M}$ defined by

$$\tilde{M}_{\boldsymbol{h}^i,\tilde{\boldsymbol{h}}^i} = \begin{cases} q(\boldsymbol{h}^i)_j & \text{if } h^i_k = \tilde{h}^i_{k-1} \text{ for all } k \in \{2, \dots, n\} \text{ and } \tilde{h}^i_n = A_j, \\ 0 & \text{else.} \end{cases} \tag{3.7}$$

Similarly to Section 2.3, we obtain the limiting distribution $\tilde{\boldsymbol{v}}$ of $\tilde{M}$. Then, the payoff of $\boldsymbol{q}$ playing against $\boldsymbol{p}$ is given by

$$\pi(\boldsymbol{q}, \boldsymbol{p}) = \sum_{\boldsymbol{h}^i \in H^i_n} \tilde{v}_{\boldsymbol{h}^i} \sum_{j,k=1}^{m} g_{jk} q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k. \tag{3.8}$$

3.2 Constraining the Search to Pure Self-Reactive Strategies

Before proving the second proposition, we need an auxiliary result.

Lemma 3.3. *Let $g, h : [0,1]^k \to \mathbb{R}$ be two affine functions and their quotient $f \coloneqq g/h$ be bounded on $[0,1]^k$. Given arbitrary $\boldsymbol{x} = (x_1, \dots, x_k)^T$ and $j \in \{1, \dots, k\}$ define $f_{\boldsymbol{x},j}(t) \coloneqq f(x_1, \dots, x_j + t, \dots, x_k)$. Then $f_{\boldsymbol{x},j}$ is monotonic.*

Proof. Define $x_0 \coloneqq 1$. Since both g and h are affine, there exist $a_0, \dots, a_k \in \mathbb{R}$ and $b_0, \dots, b_k \in \mathbb{R}$ such that

$$g(\boldsymbol{x}) \coloneqq \sum_{j=0}^{k} a_j x_j$$

$$h(\boldsymbol{x}) \coloneqq \sum_{j=0}^{k} b_j x_j$$

Fix an arbitrary $\boldsymbol{x} \in [0,1]^k$ and derive

$$f'_{\boldsymbol{x},j}(t) = \frac{\partial}{\partial t} f(x_1, \dots, x_j + t, \dots, x_k) = \frac{a_j \sum_{i \neq j} b_i x_i - b_j \sum_{i \neq j} a_i x_i}{\left(b_j (x_j + t) \sum_{i \neq j} b_i x_i\right)^2}$$

We observe that the denominator of $f'_{\boldsymbol{x},j}$ is strictly positive, since f is bounded on the domain. The numerator is constant in t. Thus, depending on the sign of

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

the numerator, $f'_{\boldsymbol{x},j}$ is either constant, monotonically increasing or monotonically decreasing. □

Proposition 3.4. *Let* $\boldsymbol{p} \in R_n$ *and* $\boldsymbol{q} \in S_n$ *be arbitrary, but fixed. Then, there exists* $\boldsymbol{q}^* \in S_n^{pure}$ *such that*

$$\pi(\boldsymbol{q}^*, \boldsymbol{p}) \geq \pi(\boldsymbol{q}, \boldsymbol{p}). \tag{3.9}$$

Proof. The proof follows in 3 steps:

1. For $\boldsymbol{h}^i \in H_n^i$ let $M_{\boldsymbol{h}^i}$ be the matrix derived from the transition matrix $\tilde{M}$, as defined in (3.7), by following the two steps:
   - Subtract the I_{m^n} identity matrix from $\tilde{M}$.
   - Set every entry in the last column to zero, except the entry that corresponds to the history $\boldsymbol{h}^i$. This entry is set to one.

   We observe, that since the stationary distribution $\tilde{v}_{\boldsymbol{h}^i}$ is an eigenvector of $\tilde{M}$ with eigenvalue 1, $\tilde{v}_{\boldsymbol{h}^i}(\tilde{M} - I_{m^n}) = 0$. Thus, $\det(\tilde{M} - I_{m^n}) = 0$. Following Cramer's rule results in

   $$\mathrm{Adj}(\tilde{M} - I_{m^n})(\tilde{M} - I_{m^n}) = \det(\tilde{M} - I_{m^n})I_{m^n} = 0.$$

   Since the eigenspace to the eigenvalue 1 of the matrix $\tilde{M}$ is one dimensional, every row of the adjugate matrix $\boldsymbol{A} := \mathrm{Adj}(\tilde{M} - I_{m^n})$ is a scalar multiple to the eigenvector $\tilde{v}$. We consider the last row and denote by c the scalar such that

   $$\tilde{v} = c\ (A_{m^n,j})_{j=1}^{m^n}. \tag{3.10}$$

   By definition of $\boldsymbol{A}$, the j-th entry $A_{m^n,j}$ is up to a sign the determinant of the matrix $\tilde{M}$ where the last column and the j-th row have been eliminated. Since the last column was eliminated, these determinants are also the determinants of the Matrix $M_{\boldsymbol{h}^i}$ if the same row and columns are eliminated. Applying Laplace expansion to the last column of $M_{\boldsymbol{h}^i}$ for every $\boldsymbol{h}^i \in H_n^i$, we obtain from (3.10) that

   $$\tilde{v} = c\ (\det(M_{\boldsymbol{h}^i}))_{\boldsymbol{h}^i \in H_n^i}. \tag{3.11}$$

   From $\sum_{\boldsymbol{h}^i \in H_n^i} \tilde{v}_{\boldsymbol{h}^i} = 1$, we infer that

   $$\frac{1}{c} = \sum_{\boldsymbol{h}^i \in H_n^i} \det(M_{\boldsymbol{h}^i}) \tag{3.12}$$

   We return to the payoff $\pi(\boldsymbol{q}, \boldsymbol{p})$ as calculated in (3.8). It follows from (3.11)

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

and (3.12) that

$$\pi(\boldsymbol{q},\boldsymbol{p}) = \frac{\sum_{\boldsymbol{h}^i\in H_n^i} \det(M_{\boldsymbol{h}^i}) \left(\sum_{j,k=1}^m g_{jk} q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k\right)}{\sum_{\boldsymbol{h}^i\in H_n^i} \det(M_{\boldsymbol{h}^i})}. \tag{3.13}$$

2. Let us examine for fixed $\boldsymbol{h}^i \in H_n^i$ the term $\det(M_{\boldsymbol{h}^i})$. Note that the entries of the last column are zero except for the row that corresponds to the history $\boldsymbol{h}^i$. The only entries that contain elements of the vector $q(\boldsymbol{h}^i)$ are in the row that corresponds to the history $\boldsymbol{h}^i$. It follows that the term $\det(M_{\boldsymbol{h}^i})$ is independent of the vector $q(\boldsymbol{h}^i)$. Therefore, both the enumerator and denominator are affine in each vector $q(\tilde{\boldsymbol{h}}^i)$ for $\tilde{\boldsymbol{h}}^i \in H_n^i$. The payoff function is bounded from above and below by the maximum and minimum entry of the payoff matrix G.

3. We consider each history $\boldsymbol{h}^i$ iteratively and show that there exists a best response strategy, that reacts to this history $\boldsymbol{h}^i$ purely. Therefore, let us fix the pure self-reactive-n strategy $\boldsymbol{q}$ and an observed history $\boldsymbol{h}^i$. Since $q(\boldsymbol{h}^i)$ is an element of the space Δ^{m-1} we replace $q(\boldsymbol{h}^i)_m$ by $1-\sum_{j\neq m} q(\boldsymbol{h}^i)_j$. We observe that both the denominator and numerator of π are affine in $q(\boldsymbol{h}^i)_1$ since they are affine in the vector $q(\boldsymbol{h}^i)$ as seen in step 2. From Lemma 3.3, it follows that π is monotonic in $q(\boldsymbol{h}^i)_1$. If it is decreasing, the strategy $\tilde{\boldsymbol{q}}$, defined by $\tilde{q}^{<n} := q^{<n}$ and

$$\tilde{q}(\tilde{\boldsymbol{h}}^i) := \begin{cases} \left(0, q(\tilde{\boldsymbol{h}}^i)_2, \dots, 1-\sum_{j=2}^{m-1} q(\tilde{\boldsymbol{h}}^i)_j\right)^T & \text{for } \tilde{\boldsymbol{h}}^i = \boldsymbol{h}^i \\ q(\tilde{\boldsymbol{h}}^i) & \text{else,} \end{cases}$$

solves the equation $\pi(\tilde{\boldsymbol{q}},\boldsymbol{p}) \geq \pi(\boldsymbol{q},\boldsymbol{p})$. If, on the other hand, $\pi(q(\boldsymbol{h}^i)_1)$ is monotonically increasing, define

$$\tilde{q}(\tilde{\boldsymbol{h}}^i) := \begin{cases} \left(1-\sum_{j=2}^{m-1} q(\tilde{\boldsymbol{h}}^i)_j, q(\tilde{\boldsymbol{h}}^i)_2, \dots, 0\right)^T & \text{for } \tilde{\boldsymbol{h}}^i = \boldsymbol{h}^i \\ q(\tilde{\boldsymbol{h}}^i) & \text{else,} \end{cases}$$

to observe the same result as in the decreasing case. In each instance we can consider the obtained new strategy as as element of Δ^{m-2}. The numerator and denominator of the payoff function π continue to be affine in $\tilde{q}(\boldsymbol{h}^i)$. We can therefore apply the same procedure on $\tilde{\boldsymbol{q}}$. Iteratively we obtain a strategy $\hat{\boldsymbol{q}}$, where $\hat{q}(\boldsymbol{h}^i)$ is an element of the canonical basis of $\mathbb{R}^m$. By applying the same strategy on $\hat{\boldsymbol{q}}$, but for another history $\hat{\boldsymbol{h}}^i$, in which $\hat{q}(\hat{\boldsymbol{h}}^i)$ is probabilistic and iteratively repeating this behavior, we obtain a pure self-reactive-n strategy $\boldsymbol{q}^*$ that fulfills the desired Equation (3.9).

□

We now combine the results to show Theorem 3.1.

Proof. Consider $\boldsymbol{p} \in R_n$. Let $\sigma \in \Sigma_{\boldsymbol{p}}$ be arbitrary but fixed. We obtain from Proposition 3.2 a self-reactive-n strategy $\boldsymbol{q}$ with $\pi(\boldsymbol{q},\boldsymbol{p}) = \pi(\sigma,\boldsymbol{p})$. To the strategy $\boldsymbol{q}$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

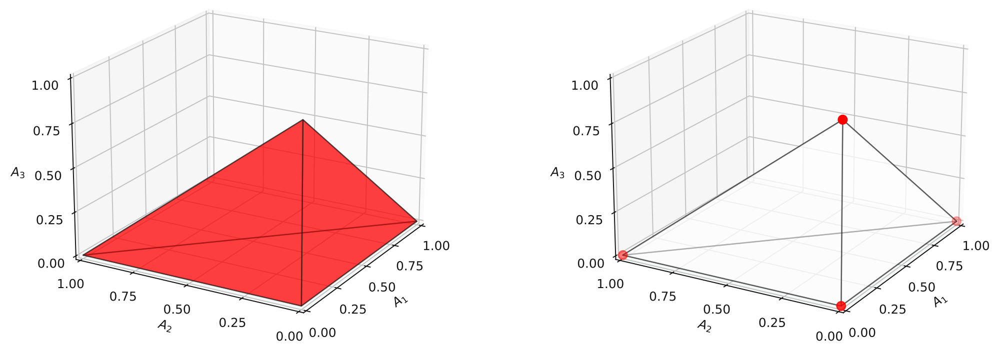
there exists by Proposition 3.4 a pure self-reactive-n strategy $\boldsymbol{q}^*$, such that $\pi(\boldsymbol{q}^*, \boldsymbol{p}) \geq \pi(\boldsymbol{q}, \boldsymbol{p})$. The set of pure self-reactive-n strategies is finite, thus the maximum of $\pi(\tilde{\boldsymbol{q}}, \boldsymbol{p})$ over all $\tilde{\boldsymbol{q}} \in S_n^{pure}$ exists and it holds

$$\pi(\sigma, \boldsymbol{p}) = \pi(\boldsymbol{q}, \boldsymbol{p}) \leq \pi(\boldsymbol{q}^*, \boldsymbol{p}) \leq \max_{\tilde{\boldsymbol{q}} \in S_n^{pure}} \pi(\tilde{\boldsymbol{q}}, \boldsymbol{p}).$$

As σ was chosen arbitrary, (3.1) holds for $\tilde{\boldsymbol{p}} \in \operatorname{argmax}_{\tilde{\boldsymbol{q}} \in S_n^{pure}} \pi(\tilde{\boldsymbol{q}}, \boldsymbol{p})$. □

3.3 Impact

In this section, we have proven that given an arbitrary reactive-n strategy $\boldsymbol{p}$, there always exists a best response within the set of all pure self-reactive-n strategies. This also proved the existence of a best response. The result is remarkable, as it reduces the solution space from an infinite set to a finite subset. An illustration of this result is provided in Figure 3.1.



(a) (b)

Figure 3.1: The figure explores the case of 4 actions. Arbitrary strategies map to the interior, edges, and vertices (a), while pure strategies map only to the vertices (b). The probability of the fourth action being chosen corresponds to 1 minus the probability of all other actions being chosen. Thus, the fourth action does not need to be represented directly in the figures.

As already stated, best responses are essential for the characterization of Nash equilibria. Our results allow us to explicitly compute a number of sufficient conditions for verifying whether a reactive-n strategy is a Nash equilibrium. We can thereby quantify the complexity of the problem in a concrete way. We formalize this idea in the following corollary.

Corollary 3.5. *Consider a symmetric stage game with m actions. Then,to decide if a reactive-n strategy $\boldsymbol{p}$ is a symmetric Nash equilibrium, it takes at most $m^n \cdot m^{(m^n)} + 1$ calculations of payoffs.*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Proof. By Definition 2.10 and Theorem 3.1 $\boldsymbol{p}$ is a symmetric Nash equilibrium if and only if

$$\pi(\boldsymbol{p},\boldsymbol{p}) = \max_{\boldsymbol{q}\in S_n^{\text{pure}}} \pi(\boldsymbol{q},\boldsymbol{p}). \tag{3.14}$$

We thus need to compute the size of the set $S_n^{\text{pure}} = A^n \times \{f \mid f : H_n^i \to \{e_1,\dots,e_m\}\}$. Since A consists of m actions, A^n has m^n elements. The set of histories H_n^i has m^n elements since it is equal to A^n. Thus, there are $m^{(m^n)}$ functions that map from H_n^i into the canonical basis of $\mathbb{R}^m$. Lastly, we need to compute the payoff of $\boldsymbol{p}$ playing against itself. This results in $m^n \cdot m^{(m^n)} + 1$ calculations of payoffs. □

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

4 Factorized Pure Self-Reactive Strategies

We have seen in Chapter 3 that searching for a best response against an arbitrary reactive-n strategy $\boldsymbol{p}$ within the set of all pure self-reactive-n strategies is sufficient. This raises the question: is it actually necessary to test all of them? Numerical observations, such as those in [GANH24], suggest that the answer is negative and that the set of strategies we need to consider can be further reduced. This section aims to mathematically formalize that observation and to identify a subset of pure self-reactive-n strategies that is sufficient for determining a best response. To this end, we study how the behavior of a pure self-reactive strategy maps to a cycle in a specific graph, the *de Bruijn graph*. Using this insight, we define an appropriate equivalence relation to factorize the set of pure self-reactive-n strategies. Finally, we analyze the extent to which this reduces the number of necessary evaluations.

In Section 4.1, we introduce fundamental graph-theoretical concepts, including cycles and the de Bruijn graph, which provide the framework for representing strategies as paths in graphs. We then show in Section 4.2 how each pure self-reactive strategy can be represented as a subgraph of the de Bruijn graph. In Section 4.3 we observe that every pure self-reactive strategy corresponds to a cycle in the de Bruijn graph, leading to an equivalence relation that groups strategies by their associated cycle. We demonstrate in Section 4.4 that payoffs only depend on the cycle a strategy induces, enabling a substantial reduction in the number of strategies that must be evaluated to find a best response.

4.1 Graph Theoretical Preliminaries

In this section, we provide a brief introduction to graph theory by defining all concepts relevant to this thesis.

Definition 4.1. *A directed graph G is a tuple (V, E), where V is a finite set and $E \subseteq \{(u, v) : u, v \in V\}$.*

The elements of the set V are referred to as *vertices* or *nodes*, while the elements of set E are called *edges*. Given the edge $e = (u, v)$, we refer to u as the *tail* of the edge e and v as the *head*. We emphasize that we allow for *loops*, i.e., edges of the form (u, u).

Definition 4.2. *A graph $\tilde{G} = (\tilde{V}, \tilde{E})$ is called a subgraph of the graph $G = (V, E)$ if $\tilde{V} \subseteq V$ and $\tilde{E} \subseteq E$.*

Definition 4.3. *The set of edges $Z := \{e_1, \ldots, e_k\} \subseteq E$ is a* cycle *of the graph G if the following three conditions hold*

1. *There exist k distinct vertices $\{v_1, \dots, v_k\}$, such that for every $i \in \{1, \dots, k\}$ there exist $j_1, j_2 \in \{1, \dots, k\}$ with $e_i = (v_{j_1}, v_{j_2})$.*
2. *No distinct edges have the same head and no distinct edges have the same tail.*
3. *There exists no subset $\tilde{Z}$ of Z, for which 1 and 2 hold.*

A cycle $\{e_1, \dots, e_k\}$ is said to have *length* k. We say a vertex $v \in V$ is on the cycle Z, denoted by $v \in Z$ if it is the tail and head of an edge of the cycle.

A cycle can be viewed as a walk over edges in the graph, following specific rules. Starting from any (initial) vertex in the cycle, one follows edges such that no vertex is met twice, except for the initial vertex, where the walk terminates. An example of a walk that does not form a cycle is presented in Figure 4.1.

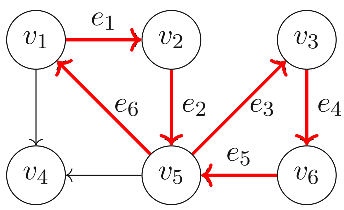


Figure 4.1: The set $\{e_1, e_2, e_3, e_4, e_5, e_6\}$ is not a cycle.

Remark 4.4. *In the literature a cycle is often defined as a* sequence *of edges rather than a set. Taking the sequence approach results in cycles having fixed initial vertices. Since we do not distinguish between sequences of cycles with the same edges but different starting vertices (see Figure 4.2), we must group such sequences together. To do this, we factorize all sequences such that those with the same set of edges fall into the same equivalence class. Instead of factorizing, we define cycles as sets of edges, achieving the same outcome more directly.*

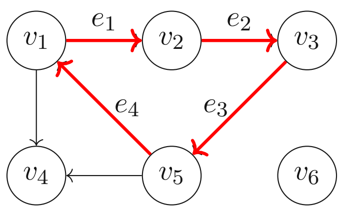


Figure 4.2: The sequences (e_1, e_2, e_3, e_4) and (e_2, e_3, e_4, e_1) describe the same cycle, namely $\{e_1, e_2, e_3, e_4\}$.

Definition 4.5. *A cycle Z of the directed graph $G = (V, E)$ is a* Hamiltonian cycle *if it is of length $|V|$.*

It follows immediately that in a Hamiltonian cycle every vertex on the Graph is reached by the cycle. An Hamiltonian cycle can be interpreted as a walk through the graph, which covers all vertices without reaching any vertex twice and never walking an edge more than once. Note that a Hamiltonian cycle does not need to include all edges.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

4.2 Repeated Games represented on the de Bruijn Graph

In this section, we present how graph-theoretical structures can be used in the study of repeated games. Specifically, we consider the case where nodes of a graph represent observable histories, and edges represent possible transitions between them during iterations of the game.

This is motivated by the fact that Markov chains, as the one described in Section 2.3, can be represented as directed graphs. The vertices are the states of the Markov chain and the edges are weighted by the probability of transitioning from one state to another. Commonly, edges exist if and only if their weight is positive.

In our setting both players rely solely on the previous actions of the self-reactive-n player to make their decisions, as already discussed in Section 3.1. Thus, it is sufficient to only consider the possible histories and transitions of the self-reactive strategy. Similar to the Markov chain approach, we define the vertices of our graph as all observable histories. There exists an edge from one vertex to another if and only if the history of the second vertex can be obtained from the first through a single step of the game's iteration.

Interestingly, the structure of this graph corresponds to the de Bruijn graph, first introduced by [Bru46]. We give the formal definition in Definition 4.6. See a numerical example in Figure 4.3. We then show that each pure self-reactive strategy corresponds to a subgraph capturing all the histories that this strategy allows. In Algorithm A we construct the subgraph and prove its well-definedness.

Definition 4.6. *The n-dimensional* de Bruijn graph $G_n = (V, E)$ *over the alphabet A is defined by*

- $V \coloneqq H_n^i = A^n$ *as the set of vertices,*
- $E \coloneqq \{(v_1, v_2) \in V \times V : (v_1)_k = (v_2)_{k-1}$ *for all* $k \in \{2, \ldots, n\}\}$ *as the set of edges.*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

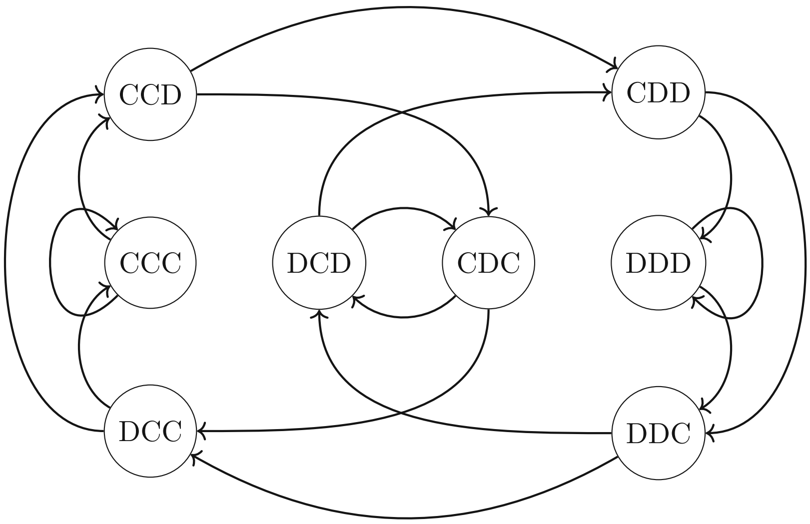


Figure 4.3: The 3-dimensional de Bruijn graph over the alphabet $\{\mathrm{C}, \mathrm{D}\}$. The nodes are all possible histories of a self reactive-3 player, i.e., $V := \{\mathrm{C}, \mathrm{D}\}^3$. In each iteration the history is updated by discarding the earliest rememberable event and appending the most recent move. This is reflected in the graph: two nodes are connected by an edge if and only if the first two entries of the first node match the last two entries of the second node.

Let $n \in \mathbb{N}$ be fixed but arbitrary and denote by $\boldsymbol{q}$ a pure self-reactive-n strategy. We construct the subgraph $G_{\boldsymbol{q}} = (V_{\boldsymbol{q}}, E_{\boldsymbol{q}})$ of the n-dimensional de Bruijn graph over the alphabet A via the following algorithm. For simplicity, we define the function $\mathcal{A} : \{e_1, \ldots, e_m\} \to A$ that maps for every $i \in \{1, \ldots, m\}$ the canonical basis vector e_i of $\mathbb{R}^m$ to the action A_i. Recall that P_{n-1} is the projection of a tuple onto its last $n-1$ components.

Algorithm A (Strategies to Subgraphs)**.** *Given a pure self-reactive-n strategy $\boldsymbol{q} = (q^{<n}, q)$ and the local variables v_{loc} and u_{loc} take the following steps.*

(I) $u_{loc} = q^{<n}$

(II) $V_{\boldsymbol{q}} = \{u_{loc}\}$

(III) $v_{loc} = (P_{n-1}(u_{loc}), \mathcal{A}(q(u_{loc})))$

(IV) $E_{\boldsymbol{q}} = \{(u_{loc}, v_{loc})\}$

(V) ***while*** $v_{loc} \notin V_{\boldsymbol{q}}$*:*

- *a)* $V_{\boldsymbol{q}} = V_{\boldsymbol{q}} \cup \{v_{loc}\}$
- *b)* $u_{loc} = v_{loc}$
- *c)* $v_{loc} = (P_{n-1}(u_{loc}), \mathcal{A}(q(u_{loc})))$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

d) $E_{\boldsymbol{q}} = E_{\boldsymbol{q}} \cup \{(u_{loc}, v_{loc})\}$

(VI) $E_{\boldsymbol{q}} = E_{\boldsymbol{q}} \cup \{(u_{loc}, v_{loc})\}$

(VII) *Return* $(V_{\boldsymbol{q}}, E_{\boldsymbol{q}})$.

Lemma 4.7. *Given a pure self-reactive-n strategy* $\boldsymbol{q}$*, Algorithm A terminates and the returned graph* $G_{\boldsymbol{q}} = (V_{\boldsymbol{q}}, E_{\boldsymbol{q}})$ *is a subgraph of the* n*-dimensional de Bruijn graph* $G_n = (V, E)$ *over the alphabet* A*.*

Proof. We observe that $V_{\boldsymbol{q}}$ is built over the local variables u_{loc} and v_{loc}. In the first initialization u_{loc} is set to $q^{<n}$ which is an element of A^n and thus an element of V. If u_{loc} is in V, then so is $v_{\text{loc}} = (P_{n-1}(u_{\text{loc}}), \mathcal{A}(q(u_{\text{loc}})))$ as it is an element of $A^{n-1} \times A = A^n = V$. Therefore, $V_{\boldsymbol{q}} \subseteq V$. For every edge e in $E_{\boldsymbol{q}}$ there exists $u \in V_{\boldsymbol{q}}$ and $A_j \in A$ such that $e = (u, \underbrace{(P_{n-1}(u), A_j)}_{:=v})$. Since $P_{n-1}(u)$ is the projection onto u's last n-1 components, $u_k = v_{k-1}$ for all $k \in 2, \ldots, n$. By definition 4.6, $E_{\boldsymbol{q}} \subseteq E$. We have thus shown that $G_{\boldsymbol{q}}$ is a subgraph of G_n. It remains to discuss if the algorithm terminates at every input. Suppose there exists an input such that the algorithm does not terminate. If follows, that v_{loc} has never been observed at any interaction and is added to $V_{\boldsymbol{q}}$. However, $V_{\boldsymbol{q}}$ being a subset of V, which is by definition a finite set, leads to a contradiction. □

Note, that the graph $G_{\boldsymbol{q}}$ is independent of $\boldsymbol{p}$. The algorithm starts by initializing $V_{\boldsymbol{q}}$ with the first n-history, that the pure self-reactive-n strategy $\boldsymbol{q}$ observes, i.e., $q^{<n}$. By calculating, which action $\boldsymbol{q}$ chooses to play next, we build our graph, adding all observed histories in the order they appear. Since $\boldsymbol{q}$ is deterministic, i.e., $\boldsymbol{q}$ never chooses differently after the same history, the algorithm terminates upon returning to a history, that $\boldsymbol{q}$ has already observed before. See Example 4.8 for an illustration of the procedure.

Example 4.8. Consider the action set $A = \{\text{C}, \text{D}\}$ and pure self-reactive-3 strategy $\boldsymbol{q} = (q^{<3}, q)$ defined by

$$q^{<3} = (\text{CCD})$$

$$q(v) = \begin{cases} (1,0)^T & \text{if } v \in \{(\text{CCD}), (\text{CDD})\} \\ (0,1)^T & \text{else.} \end{cases}$$

The graph constructed in Algorithm A is shown in Figure 4.4. The Algorithm starts by adding (CCD) to set of vertices. Then, other vertices are added in the order in which they are observed. Edges connecting the vertices are added accordingly.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

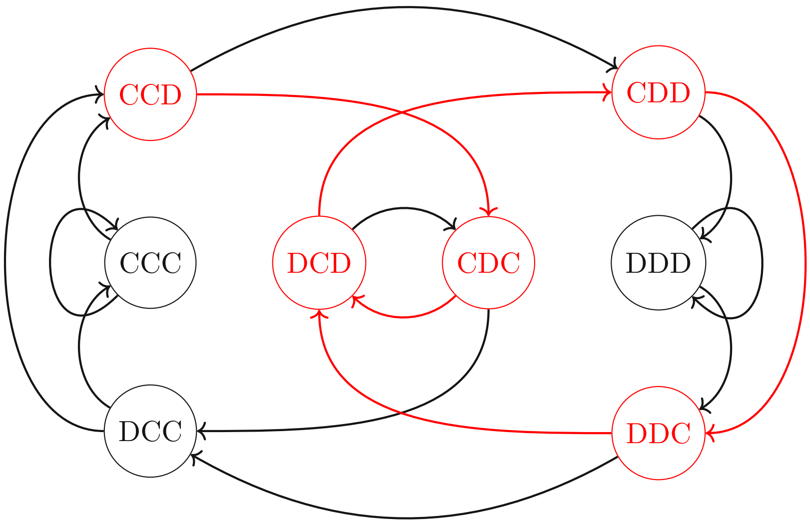


Figure 4.4: In the figure the subset $G_{\boldsymbol{q}} = (V_{\boldsymbol{q}}, E_{\boldsymbol{q}})$ of the n-dimensional de Bruijn graph over the alphabet $\{\mathrm{C}, \mathrm{D}\}$ is marked in red.

Considering the example, note that every history that is observed after and including the history (DCD) is observed infinitely often as they form a cycle. All other histories are observed at most once.

4.3 Classes of Pure Self-Reactive Strategies

By the construction of the graph $G_{\boldsymbol{q}}$ by means of Algorithm A, $G_{\boldsymbol{q}}$ has a unique cycle. We refer to this cycle as the cycle of the pure self-reactive strategy $\boldsymbol{q}$ denoted by $Z_{\boldsymbol{q}}$.

Remark 4.9. *To calculate the cycle, we adapt Algorithm A by taking the following steps.*

1. *Every observed vertex is labeled by the iteration at which it was observed. The history of the initial strategy is labeled 0. A counter is introduced that increases at every step of the while loop and labels every v_{loc} upon addition to $V_{\boldsymbol{q}}$.*

2. *Suppose that the while-loop has terminated after observing v for the second time and that v is labeled k. Delete all vertices labeled smaller. Delete all edges whose head or tail have been removed.*

Then, E_q^{del}, i.e., $E_{\boldsymbol{q}}$ with deletion, defines the cycle $Z_{\boldsymbol{q}}$. The labels can be removed.

We show that this cycle is the decisive ingredient for the calculation of payoffs. To this end, we introduce the following equivalence relation on the set of all pure self-reactive-n strategies. Two strategies $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ are equivalent, in short $\boldsymbol{q_1} \sim \boldsymbol{q_2}$, if their cycles are equal, i.e., $Z_{\boldsymbol{q_1}} = Z_{\boldsymbol{q_2}}$. Given a strategy $\boldsymbol{q} \in S_n^{\text{pure}}$, we denote by $[\boldsymbol{q}]_\sim$ its equivalence class.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Lemma 4.10. *Consider two pure self-reactive-n strategies* $\boldsymbol{q}$ *and* $\tilde{\boldsymbol{q}}$*. Then*

$$\boldsymbol{q} \sim \tilde{\boldsymbol{q}} \Rightarrow \pi(\boldsymbol{q}, \boldsymbol{p}) = \pi(\tilde{\boldsymbol{q}}, \boldsymbol{p}) \quad \text{for all } \boldsymbol{p} \in R_n. \tag{4.1}$$

Proof. Let $\boldsymbol{p} \in R_n$ be arbitrary but fixed. Since the two pure self-reactive-n strategies $\boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ are equivalent, their cycles $Z_{\boldsymbol{q}}$ and $Z_{\tilde{\boldsymbol{q}}}$ are equal. By the construction of cycles, this implies the existence of $N \in \mathbb{N}$ and $k \in \mathbb{N}$ such that $\pi_{\boldsymbol{q},\boldsymbol{p}}(t) = \pi_{\tilde{\boldsymbol{q}},\boldsymbol{p}}(t+k)$ for all $t \geq N$, where $\pi_{\boldsymbol{q},\boldsymbol{p}}(t)$ is the expected payoff of $\boldsymbol{q}$ in round t. The value N can be chosen as the iteration step at which $\boldsymbol{q}$ first observes a node on the cycle. Then, k can be selected as the number of steps until $\tilde{\boldsymbol{q}}$ observes the same node after step N. We observe

$$\begin{aligned}
\pi(\boldsymbol{q}, \boldsymbol{p}) &= \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=1}^{\tau} \pi_{\boldsymbol{q},\boldsymbol{p}}(t) = \lim_{\tau\to\infty} \frac{1}{\tau} \left(\sum_{t=1}^{N} \pi_{\boldsymbol{q},\boldsymbol{p}}(t) + \sum_{t=N+1}^{\tau} \pi_{\boldsymbol{q},\boldsymbol{p}}(t) \right) = \\
&= \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=N+1}^{\tau} \pi_{\tilde{\boldsymbol{q}},\boldsymbol{p}}(t+k) = \lim_{\tau\to\infty} \frac{1}{\tau} \sum_{t=N+1+k}^{\tau} \pi_{\tilde{\boldsymbol{q}},\boldsymbol{p}}(t) = \\
&= \lim_{\tau\to\infty} \frac{1}{\tau} \left(\sum_{t=1}^{N+k} \pi_{\tilde{\boldsymbol{q}},\boldsymbol{p}}(t) + \sum_{t=N+1+k}^{\tau} \pi_{\tilde{\boldsymbol{q}},\boldsymbol{p}}(t) \right) = \pi(\tilde{\boldsymbol{q}}, \boldsymbol{p}).
\end{aligned}$$

□

We emphasize, that the factorization of S_n^{pure} is independent of $\boldsymbol{p}$. There exists a bijective mapping ϕ from $S_n^{\text{pure}}/\sim$ onto the cycles of the n-dimensional de Bruijn graph, denoted by $Cyc(n)$.

From Lemma 4.10 it follows that the definition of the payoff function π on the domain $Cyc(n) \times R_n$ is well-defined by

$$\pi(Z, .) \coloneqq \pi(\boldsymbol{q}, .), \tag{4.2}$$

where $\boldsymbol{q}$ is an arbitrary element of $\phi^{-1}(Z)$.

4.4 Impact

The observations made in Section 4.3 lead to the following improvement of Theorem 3.1.

Theorem 4.11. *Let* $\boldsymbol{p} \in R_n$ *be a reactive-n strategy. Then, there exists* $Z \in Cyc(n)$ *with*

$$\pi(Z, \boldsymbol{p}) \geq \pi(\sigma, \boldsymbol{p}) \quad \text{for all } \sigma \in \Sigma_{\boldsymbol{p}}. \tag{4.3}$$

Considering our goal of efficiently computing a best response against a fixed but arbitrary reactive-n strategy $\boldsymbol{p}$, it suffices to test one representative of each equivalence class or, by (4.2), to test every cycle on the n-dimensional de Bruijn graph. In this section we first focus on how the calculation of payoffs are simplified

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

using cycles. Then, we explore the number of cycles on the de Bruijn graph to obtain an approximation for the number of conditions to calculate symmetric Nash equilibria.

Proposition 4.12. *Consider the reactive-n strategy $\boldsymbol{p}$ and pure self-reactive-n strategy $\boldsymbol{q}$. Let $Z = \phi([\boldsymbol{q}]_\sim)$. Then,*

$$\pi(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{|Z|} \sum_{\boldsymbol{h}^i \in Z} \sum_{j,k=1}^{m} g_{jk} q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k. \tag{4.4}$$

Note, that $q(\boldsymbol{h}^i)$ is a canonical basis vector of $\mathbb{R}^m$. Thus the sum over k is equal to zero for $m-1$ entries of j.

Proof. Considering the payoff calculation of $\boldsymbol{q}$ playing against $\boldsymbol{p}$ as derived in (3.8), we derive a term for the limiting distribution $(\tilde{v}_{\boldsymbol{h}^1})_{\boldsymbol{h}^1 \in H_n^i}$ in dependence of the cycle Z. To this end, observe that every vertex, that is not on the cycle is visited at most 1 time. All vertices on the cycle are visited infinitely often, with visits uniformly distributed. It follows that

$$\tilde{v}_{\boldsymbol{h}^1} = \begin{cases} \frac{1}{|Z|} & \text{if } \boldsymbol{h}^1 \in Z, \\ 0 & \text{else.} \end{cases} \tag{4.5}$$

Combining these observations results in

$$\begin{aligned}
\pi(\boldsymbol{q}, \boldsymbol{p}) &\overset{(3.8)}{=} \sum_{\boldsymbol{h}^1 \in H_n^i} \tilde{v}_{\boldsymbol{h}^1} \sum_{i,j=1}^{m} G_{ij} q(\boldsymbol{h}^1)_i p(\boldsymbol{h}^1)_j \\
&= \sum_{\boldsymbol{h}^1 \in Z} \tilde{v}_{\boldsymbol{h}^1} \sum_{i,j=1}^{m} G_{ij} q(\boldsymbol{h}^1)_i p(\boldsymbol{h}^1)_j + \sum_{\boldsymbol{h}^1 \notin Z} \tilde{v}_{\boldsymbol{h}^1} \sum_{i,j=1}^{m} G_{ij} q(\boldsymbol{h}^1)_i p(\boldsymbol{h}^1)_j \\
&= \sum_{\boldsymbol{h}^1 \in Z} \frac{1}{|Z|} \sum_{i,j=1}^{m} G_{ij} q(\boldsymbol{h}^1)_i p(\boldsymbol{h}^1)_j + \sum_{\boldsymbol{h}^1 \notin Z} 0 \cdot \sum_{i,j=1}^{m} G_{ij} q(\boldsymbol{h}^1)_i p(\boldsymbol{h}^1)_j \\
&= \frac{1}{|Z|} \sum_{\boldsymbol{h}^1 \in Z} \sum_{i,j=1}^{m} G_{ij} q(\boldsymbol{h}^1)_i p(\boldsymbol{h}^1)_j.
\end{aligned}$$

□

Previously, the limiting distribution was obtained by calculating the eigenvector to the eigenvalue 1 of the $m^n \times m^n$ transition matrix defined in (3.7). This step is not necessary by Proposition 4.12. It remains to calculate all cycles on the n-dimensional de Bruijn graph over the action space as the alphabet. While a closed form for the number of cycles of arbitrary dimension and arbitrary action space is not known, we can observe significant reductions already for low dimensions and small action spaces. To explore this, we calculate the number of pure self-reactive-n strategies as we did in Corollary 3.5 and compare it to the number of cycles of the corresponding de Bruijn graph. Table 4.1 illustrates this by considering two actions and calculating

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

both numbers for $n = 1, \ldots, 5$. Additionally, Table 4.2 compares the numbers for three actions and $n = 1, \ldots, 3$.

n-history	**1**	**2**	**3**	**4**	**5**
S_n^{pure}	8	64	2 048	1 048 576	137 438 953 472
$Cyc(n)$	3	6	19	179	30 176

Table 4.1: Comparison of the number of pure self-reactive-n strategies and their equivalence classes for $n \in \{1, \ldots, 5\}$ and $m = 2$.

n-history	**1**	**2**	**3**
S_n^{pure}	81	177 147	205 891 132 094 649
$Cyc(n)$	8	148	3 382 522

Table 4.2: Comparison of the number of pure self-reactive-n strategies and their equivalence classes for $n \in \{1, 2, 3\}$ and $m = 3$.

To obtain a lower bound for the number of cycles, we make use of the exact number of Hamiltonian cycles. The work of [AB87] proves that this number is $(m!)^{m^{n-1}} m^{-n}$. As calculated in Corollary 3.5 there are $m^n \cdot m^{(m^n)}$ pure self-reactive-n strategies, which provides an upper bound.

Not only does the cyclic approach simplify the search for a best response, it also provides a whole set of strategies that are best responses. This is formalized in the following corollary, which immediately follows from Lemma 4.10 and Theorem 4.11.

Corollary 4.13. *Let $\mathbf{p} \in R_n$ be a reactive-n strategy. Then there exists a $Z \in Cyc(n)$ such that every strategy in $\phi^{-1}(Z) \subseteq S_n^{pure}$ is a best response.*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

5 Best Responses in Additive Games

In this chapter, we focus on a subclass of repeated games known as additive games. The graph-theoretical embedding introduced in Chapter 4 provides the necessary tools to prove a central result: for every reactive-n strategy, there exists a best response of strictly smaller memory. We formalize this result in the following theorem.

Theorem 5.1. *Let $\boldsymbol{p} \in R_n$ be a reactive-n strategy and the game be additive. Then, there exists $\tilde{\boldsymbol{p}} \in S_{n-1}^{pure}$ with*

$$\pi(\tilde{\boldsymbol{p}}, \boldsymbol{p}) \geq \pi(\sigma, \boldsymbol{p}) \quad \text{for all } \sigma \in \Sigma_{\boldsymbol{p}}. \tag{5.1}$$

This is a remarkable finding, as it shows that a player can remember less than their opponent and still achieve the optimal payoff. The main result of this chapter has been published in [LHG25]. We take a two-step approach to establishing this result: first, in Section 5.1, we focus on a specific additive game, the two-action donation game, and provide examples where the argument holds for memory sizes 1 and 3. Additionally, we observe why the argument fails for non-additive games. Then, in Section 5.2, we prove Theorem 5.1.

5.1 Donation Game

We return to the repeated prisoner's dilemma, whose stage game is defined in Example 2.1. In this section, we consider both the general repeated prisoner's dilemma and a specific prisoner's dilemma known as the *donation game*. In this game, player's are assumed to pay a fixed cost $c > 0$ for cooperating and receive a fixed benefit $b > c$ if their opponent cooperates. This results in the payoff matrix

$$\begin{pmatrix} b-c & -c \\ b & 0 \end{pmatrix}$$

In the following example, we consider the reactive-1 strategy *Tit-for-Tat* and construct a pure self-reactive-0 strategy, that is a best response in the donation game. This proves Theorem 5.1 for a particular reactive-1 strategy. We further explore why this reduction of memory fails in the general repeated prisoner's dilemma.

Example 5.2. Consider the reactive-1 strategy $\boldsymbol{p}$ defined by

$$p^{<1} = (\mathrm{C})$$
$$p(a) = \begin{cases} (1,0)^T & \text{if } a = \mathrm{C}, \\ (0,1)^T & \text{if } a = \mathrm{D}. \end{cases}$$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Recall from Example 2.10 that this is the famous strategy *Tit-for-Tat*, which became prominent with its success in [AH81]. Let the pure self-reactive-1 strategy $\boldsymbol{q}$, which plays against $\boldsymbol{p}$, be defined by

$$q^{<1} = (\mathrm{C}) \tag{5.2}$$

$$q(a) = \begin{cases} (0,1)^T & \text{if } a = \mathrm{C}, \\ (1,0)^T & \text{if } a = \mathrm{D}. \end{cases} \tag{5.3}$$

We observe that $\phi([\boldsymbol{q}]_\sim) = \{(\mathrm{C},\mathrm{D}),(\mathrm{D},\mathrm{C})\}$, i.e., $\boldsymbol{q}$ observes the histories (C) and (D) in their cycle. Using the payoff calculation derived in (4.4) results in

$$\pi(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}(\underbrace{b}_{(\mathrm{C})} + \underbrace{-c}_{(\mathrm{D})}). \tag{5.4}$$

The underbraces show at which history each component of the payoff is obtained. Since pure self-reactive-0 strategies are strategies that constantly play one action, we will refer to these strategies by *AllC* and *AllD*, respectively. We obtain as their payoffs against $\boldsymbol{p}$

$$\pi(AllC,\boldsymbol{p}) = \underbrace{b-c}_{(\mathrm{C})},$$
$$\pi(AllD,\boldsymbol{p}) = \underbrace{0}_{(\mathrm{D})}.$$

We observe that $\pi(\boldsymbol{q},\boldsymbol{p})$ is the mean over $\pi(AllC,\boldsymbol{p})$ and $\pi(AllD,\boldsymbol{p})$. Thus, it is dominated by at least one of the payoffs. In fact, if $b > c$ then *AllC* is a better response against $\boldsymbol{p}$ than $\boldsymbol{q}$ is, else *AllD* is.

On the other hand, we consider the two strategies $\boldsymbol{q}$ and $\boldsymbol{p}$ to be playing in a general repeated prisoner's dilemma. Using the following counterexample we prove that $\boldsymbol{q}$ cannot be dominated by either *AllC* or *AllD* in general. To this end, we consider the payoff matrix

$$\begin{pmatrix} 3 & 1 \\ 6 & 0 \end{pmatrix}.$$

The payoff of $\boldsymbol{q}$ playing against $\boldsymbol{p}$ is $\pi(\boldsymbol{q},\boldsymbol{p}) = 7/2$, whereas $\pi(AllC,\boldsymbol{p}) = 3$ and $\pi(AllD,\boldsymbol{p}) = 0$. This concludes the counterexample.

Example 5.2 serves as a motivation for Theorem 5.1. We present an additional example that provides insight into why considering cycles, as established in Chapter 4, is essential for the proof of Theorem 5.1. To this end, we continue in the setting of the donation game and consider strategies with higher memory than memory-1.

Example 5.3. Let $\boldsymbol{p}$ be a reactive-3 strategy, defined by

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

$$p^{<3} = (\text{DDD}) \tag{5.5}$$

$$p(\boldsymbol{h}^i) = \begin{cases} (0.4, 0.6)^T & \text{for } \boldsymbol{h}^i = \text{DCC}, \\ (0.3, 0.7)^T & \text{for } \boldsymbol{h}^i = \text{CDC}, \\ (0.2, 0.8)^T & \text{for } \boldsymbol{h}^i = \text{CCD}, \\ (0.1, 0.9)^T & \text{for } \boldsymbol{h}^i = \text{CCC}, \\ (0, 1)^T & \text{else.} \end{cases} \tag{5.6}$$

Consider the pure self-reactive-3 strategy $\boldsymbol{q}$,

$$q^{<3} = (\text{DDD}) \tag{5.7}$$

$$q(\boldsymbol{h}^i) = \begin{cases} (1, 0)^T & \text{for } \boldsymbol{h}^i = \text{DCC}, \\ (1, 0)^T & \text{for } \boldsymbol{h}^i = \text{CDC}, \\ (1, 0)^T & \text{for } \boldsymbol{h}^i = \text{CCD}, \\ (0, 1)^T & \text{for } \boldsymbol{h}^i = \text{CCC}, \\ (1, 0)^T & \text{for } \boldsymbol{h}^i = \text{DDD}, \\ (1, 0)^T & \text{for } \boldsymbol{h}^i = \text{DDC} \\ (0, 1)^T & \text{else.} \end{cases} \tag{5.8}$$

To calculate $\boldsymbol{q}$'s payoff, we first analyze $\boldsymbol{q}$'s cycle, which is marked in red in Figure 5.1.

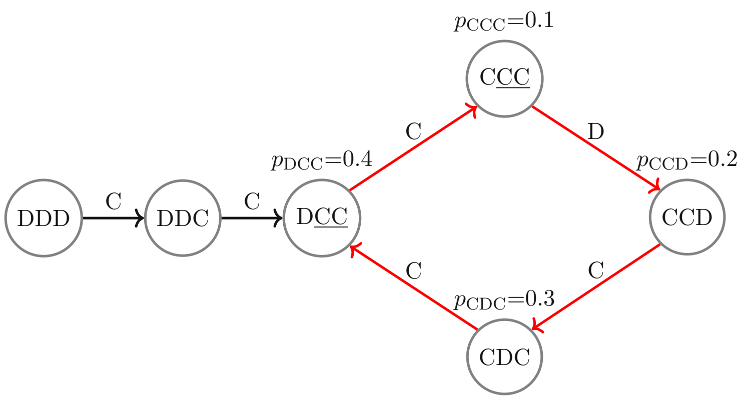


Figure 5.1: The graph $G_{\boldsymbol{q}}$ as derived from Algorithm A.

The probability, that $\boldsymbol{p}$ cooperates in a specific state, is written above the node. We observe that $Z_{\boldsymbol{q}} = \{(\text{DCC}, \text{CCC}), (\text{CCC}, \text{CCD}), (\text{CCD}, \text{CDC}), (\text{CDC}, \text{DCC})\}$.

Calculating $\boldsymbol{q}$'s payoff using (4.4) results in

$$\pi(\boldsymbol{q}, \boldsymbol{p}) = \frac{1}{4}((0.4 + 0.1 + 0.2 + 0.3)b - 3c). \tag{5.9}$$

We observe, that $\boldsymbol{q}$ behaves differently in the states (CCC) and (DCC), where the same 2-history (CC) occurs. Upon observing (CCC), $\boldsymbol{q}$ defects, while $\boldsymbol{q}$ plays

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

cooperation after the state (DCC). Thus, $\boldsymbol{q}$ cannot be mimicked by a self-reactive-2 strategy. We aim to find a pure self-reactive-2 strategy that dominates $\boldsymbol{q}$ against $\boldsymbol{p}$. To this end, we separate the previously discussed histories, that prevent $\boldsymbol{q}$ from being represented by a pure self-reactive-2 strategy. By playing C after (CCC) and D after (DCC), we obtain the graph in Figure 5.2.

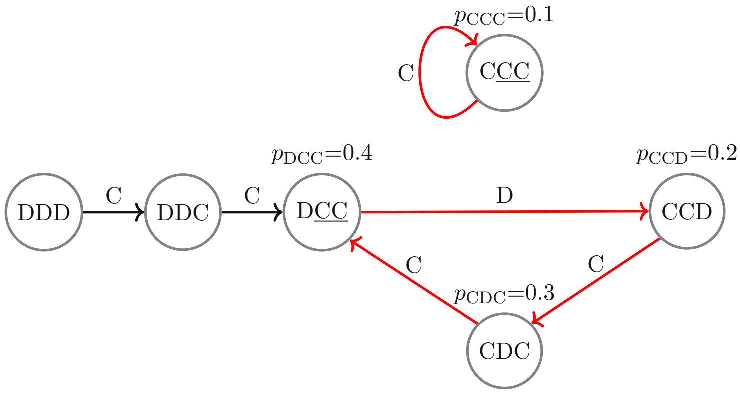


Figure 5.2: The separation of one cycle into two disjoint cycles.

The obtained disjoint cycles

$$Z_1 = \{(\text{CCC}, \text{CCC})\},$$
$$Z_2 = \{(\text{DCC}, \text{CCD}), (\text{CCD}, \text{CDC}), (\text{CDC}, \text{DCC})\}$$

have the payoffs

$$\pi(Z_1, \boldsymbol{p}) = 0.1b - c,$$
$$\pi(Z_2, \boldsymbol{p}) = \frac{1}{3}((0.4 + 0.2 + 0.3)b - 2c).$$

Let $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ be representatives from the equivalence classes $\phi^{-1}(Z_1)$ and $\phi^{-1}(Z_2)$, respectively. We observe that

$$\pi(\boldsymbol{q}, \boldsymbol{p}) \overset{(5.9)}{=} \frac{1}{4}(b - 3c) = \frac{1}{4}(0.1b - c + 3\frac{1}{3}(0.9b - 2c))$$
$$= \frac{1}{4}(\pi(Z_1, \boldsymbol{p}) + 3\pi(Z_2, \boldsymbol{p})) = \frac{1}{4}\pi(\boldsymbol{q_1}, \boldsymbol{p}) + \frac{3}{4}\pi(\boldsymbol{q_2}, \boldsymbol{p}).$$

We note that $\boldsymbol{q}$'s payoff against $\boldsymbol{p}$ is a convex combination of $\boldsymbol{q_1}$'s and $\boldsymbol{q_2}$'s payoff. It is thus dominated by at least one of the strategies. Since both $\boldsymbol{q_1}$ and $\boldsymbol{q_2}$ are pure self-reactive-2 strategies, this concludes the example and proves that there exists a pure self-reactive-2 strategy, that dominates $\boldsymbol{q}$. Which of the two is dominant depends on the values of b and c.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

5.2 Additive games

The following section formalizes the observations previously made and proves them in a general setting. To this end, consider the general repeated game setting as introduced in Chapter 2, where the game is possibly not symmetric and allows for more than two actions. We explore *additivity* of games as defined by [MFH14] and [MRH21]. This property, which holds in the donation game, will be a key characteristic that allows for a memory reduction.

Definition 5.4. *A payoff matrix G is* additive *if there exist vectors $\boldsymbol{a} := (a_1, \ldots, a_m)^T \in \mathbb{R}^m$ and $\boldsymbol{b} := (b_1, \ldots, b_m)^T \in \mathbb{R}^m$ such that*

$$G = \begin{pmatrix} a_1+b_1 & a_1+b_2 & \cdots & a_1+b_m \\ a_2+b_1 & a_2+b_2 & \cdots & a_2+b_m \\ \vdots & \vdots & \ddots & \vdots \\ a_m+b_1 & a_m+b_2 & \cdots & a_m+b_m \end{pmatrix} \in \mathbb{R}^{m\times m}. \tag{5.10}$$

A game is additive *if the focal player's payoff matrix is additive.*

That is, the focal player's payoff can be expressed as the sum of two independent components. Each of the components is solely dependent on one of the player's actions.

Example 5.5. As mentioned before, an important example of an additive game is the donation game. The focal player has to pay a cost $c > 0$ when choosing cooperation C, but gets a benefit $b > c$ if their opponent plays cooperation C. Defecting D pays nothing and gives nothing. This corresponds to an additive game where $\boldsymbol{a} = (-c, 0)$ and $\boldsymbol{b} = (b, 0)$.

The following theorem does not only show a major result, the proof itself is of interest as it provides another insight into why the graph theoretical approach of 4 is relevant.

Theorem 5.6. *Let $\boldsymbol{p} \in R_n$ be a reactive-n strategy and the game be additive. For every cycle $Z \in Cyc(n)$ there exists a cycle $\tilde{Z} \in Cyc(n-1)$ such that*

$$\pi(Z, \boldsymbol{p}) \leq \pi(\tilde{Z}, \boldsymbol{p}).$$

We obtain the following Corollary immediately from Theorem 5.6.

Corollary 5.7. *Let $\boldsymbol{p} \in R_n$ be a reactive-n strategy. Then there exists $Z \in Cyc(n-1)$ with*

$$\pi(Z, \boldsymbol{p}) \geq \pi(\sigma, \boldsymbol{p}) \quad \textit{for all } \sigma \in \Sigma_{\boldsymbol{p}}.$$

Further, the following generalization of Corollary 4.13 is obtained.

Corollary 5.8. *Let $\boldsymbol{p} \in R_n$ be a reactive-n strategy and the game be additive. Then, there exists a $Z \in Cyc(n-1)$ such that every strategy in $\phi^{-1}(Z) \subseteq S_{n-1}^{pure}$ is a best response.*

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

Theorem 5.1 follows directly from Corollary 5.8.

To prove Theorem 5.6, we make observations that will simplify the proof. In Chapter 4 we have established the payoff calculation through (4.4). Because the focal player's payoff matrix is additive, we obtain

$$
\begin{aligned}
\pi(\boldsymbol{q},\boldsymbol{p}) &\overset{(4.4)}{=} \frac{1}{|Z|}\sum_{\boldsymbol{h}^i\in Z}\sum_{j,k=1}^{m} g_{jk}q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k \\
&= \frac{1}{|Z|}\sum_{\boldsymbol{h}^i\in Z}\sum_{j,k=1}^{m}(a_j+b_k)q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k \\
&= \frac{1}{|Z|}\sum_{\boldsymbol{h}^i\in Z}\left(\sum_{j,k=1}^{m} a_j q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k + \sum_{j,k=1}^{m} b_k q(\boldsymbol{h}^i)_j p(\boldsymbol{h}^i)_k\right) \\
&= \frac{1}{|Z|}\sum_{\boldsymbol{h}^i\in Z}\left(\sum_{j=1}^{m} a_j q(\boldsymbol{h}^i)_j \underbrace{\sum_{k=1}^{m} p(\boldsymbol{h}^i)_k}_{=1} + \sum_{k=1}^{m} b_k p(\boldsymbol{h}^i)_k \underbrace{\sum_{j=1}^{m} q(\boldsymbol{h}^i)_j}_{=1}\right) \\
&= \frac{1}{|Z|}\sum_{\boldsymbol{h}^i\in Z}\left(\sum_{j=1}^{m} a_j q(\boldsymbol{h}^i)_j + \sum_{k=1}^{m} b_k p(\boldsymbol{h}^i)_k\right) \\
&= \frac{1}{|Z|}\sum_{\boldsymbol{h}^i\in Z}\left(\boldsymbol{b}\cdot p(\boldsymbol{h}^i) + \boldsymbol{a}\cdot q(\boldsymbol{h}^i)\right),
\end{aligned}
$$

for $\boldsymbol{q} \in \phi^{-1}(Z)$ and $\boldsymbol{p} \in R_n$.

By Remark 2.6, every pure self-reactive-$(n-1)$ strategy can be mapped to a pure self-reactive-n strategy without loss of information. The reverse does not hold in general. We ask ourselves what conditions must hold for a pure self-reactive-n strategy, such that there exists a pure self-reactive-$(n-1)$ strategy, which mimics the relevant behavior. By Chapter 4, this corresponds to the question, when the cycle of the pure self-reactive-n strategy on the n-dimensional de Bruijn graph can be appropriately mapped to the cycle of a pure self-reactive-$(n-1)$ strategy on the $(n-1)$-dimensional de Bruijn graph. To this end, we project all vertices of a cycle in the n-dimensional de Bruijn graph onto their last $n-1$ components. We check if this procedure defines a cycle on the $(n-1)$-dimensional de Bruijn graph. The projection of the vertices corresponds to the strategy forgetting the earliest piece of information they observe.

Lemma 5.9. *Consider the cycle Z on the n-dimensional de Bruijn graph. Sort the vertices of Z in the order of their occurrence as constructed through Algorithm A. Denote by $\boldsymbol{h}^i_k$ the history that ranks k-th in the cycle for $k \in \{1,\dots,|Z|\}$. Then, Z well-defines a cycle on the $(n-1)$-dimensional de Bruijn graph if and only if for all $k, j \in \{1,\dots,|Z|\}$ the implication*

$$P_{n-1}(\boldsymbol{h}^i_k) = P_{n-1}(\boldsymbol{h}^i_j) \Rightarrow P_{n-1}(\boldsymbol{h}^i_{k+1}) = P_{n-1}(\boldsymbol{h}^i_{j+1}) \tag{5.11}$$

holds true. Note that we define $\boldsymbol{h}^i_{|Z|+1} := \boldsymbol{h}^i_1$.

Proof. Suppose for a cycle Z there exist indices j and k such that $P_{n-1}(\boldsymbol{h}^i_k) =$

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

$P_{n-1}(\boldsymbol{h}_j^i)$ but $P_{n-1}(\boldsymbol{h}_{k+1}^i) \neq P_{n-1}(\boldsymbol{h}_{j+1}^i)$. Then, upon projecting every node on the cycle Z onto its last $n-1$ components, will result in two edges having the same tail $P_{n-1}(\boldsymbol{h}_j^i)$ but different heads $P_{n-1}(\boldsymbol{h}_{k+1}^i)$ and $P_{n-1}(\boldsymbol{h}_{j+1}^i)$. The reverse implication is proven similarly. □

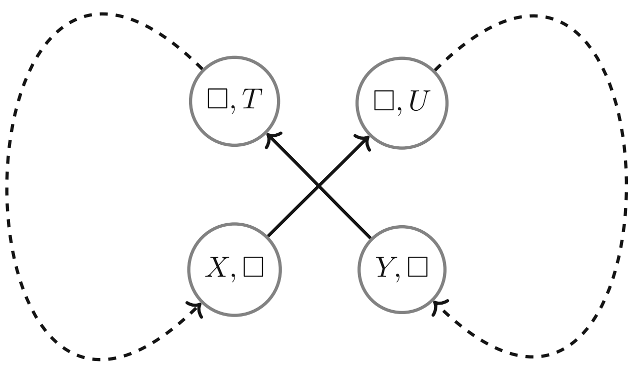
Proof of Theorem 5.6. We sort the vertices of Z as in Lemma 5.9. Let this result in $\{\boldsymbol{h}_1^i, \ldots, \boldsymbol{h}_N^i\}$, where N denotes $|Z|$.

We distinguish two cases.

Case 1. Suppose for all $j, k \in \{1, \ldots, N\}$ the implication (5.11) is true. Then, Z well-defines a cycle $\tilde{Z}$ on the $(n-1)$-dimensional de Bruijn graph with

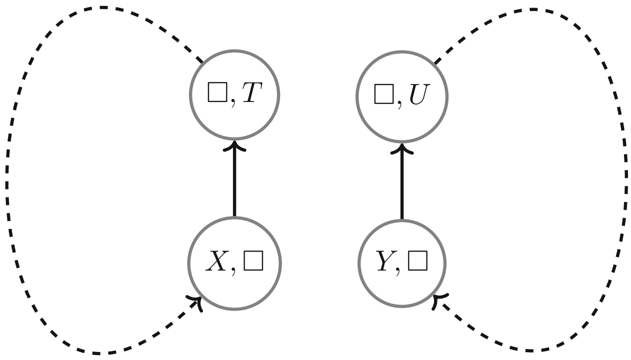
$$\pi(Z, \boldsymbol{p}) = \pi(\tilde{Z}, \boldsymbol{p}).$$

Case 2. Otherwise, there exist $j, k \in \{1, \ldots, N\}$ for which the implication in (5.11) is false. We define $\square := P_{n-1}(\boldsymbol{h}_k^i) = P_{n-1}(\boldsymbol{h}_j^i)$ and the distinct actions $X, Y \in A$ such that $(X, \square) = \boldsymbol{h}_k^i$ and $(Y, \square) = \boldsymbol{h}_j^i$. Further, we denote $(\square, U) = \boldsymbol{h}_{k+1}^i$ and $(\square, T) = \boldsymbol{h}_{j+1}^i$ where $T \neq U$ and $T, U \in A$. Then, the cycle takes the form



where the dashed arrows represent paths, which are possibly empty. We observe the two disjoint cycles

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



and denote them by Z_1 and Z_2. For any representatives $\boldsymbol{q}$, $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ of the classes $\phi^{-1}(Z)$, $\phi^{-1}(Z_1)$ and $\phi^{-1}(Z_2)$, respectively, we observe that

$$q_1(\boldsymbol{h}^i) = \begin{cases} q(\boldsymbol{h}^i) & \text{for } \boldsymbol{h}^i \in Z_1 \setminus \{(X,\square)\} \\ q((Y,\square)) & \text{for } \boldsymbol{h}^i = (X,\square) \end{cases}$$

and

$$q_2(\boldsymbol{h}^i) = \begin{cases} q(\boldsymbol{h}^i) & \text{for } \boldsymbol{h}^i \in Z_2 \setminus \{(Y,\square)\} \\ q((X,\square)) & \text{for } \boldsymbol{h}^i = (Y,\square). \end{cases}$$

Note that $Z_1 \cup Z_2 = Z$ and $Z_1 \cap Z_2 = \emptyset$. Therefore,

$$\begin{aligned} \pi(Z,\boldsymbol{p}) &= \frac{1}{|Z|} \sum_{\boldsymbol{h}^i \in Z} \Big(\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q(\boldsymbol{h}^i)\Big) \\ &= \frac{1}{|Z|}\left(\sum_{\boldsymbol{h}^i \in Z_1} (\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q(\boldsymbol{h}^i)) + \sum_{\boldsymbol{h}^i \in Z_2} (\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q(\boldsymbol{h}^i)) \right) \\ &= \frac{1}{|Z|}\left(\sum_{\boldsymbol{h}^i \in Z_1} (\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q_1(\boldsymbol{h}^i)) + \sum_{\boldsymbol{h}^i \in Z_2} (\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q_2(\boldsymbol{h}^i)) \right) \\ &= \frac{1}{|Z|}\left(\frac{|Z_1|}{|Z_1|} \sum_{\boldsymbol{h}^i \in I_1} (\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q_1(\boldsymbol{h}^i)) + \frac{|Z_2|}{|Z_2|} \sum_{\boldsymbol{h}^i \in Z_2} (\boldsymbol{b}p(\boldsymbol{h}^i) + \boldsymbol{a}q_2(\boldsymbol{h}^i)) \right) \\ &= \frac{|Z_1|}{|Z|}\pi(Z_1,\boldsymbol{p}) + \frac{|Z_2|}{|Z|}\pi(Z_2,\boldsymbol{p}). \end{aligned}$$

That is, the payoff $\pi(Z,\boldsymbol{p})$ is a convex combination of $\pi(Z_1,\boldsymbol{p})$ and $\pi(Z_2,\boldsymbol{p})$. Therefore, it is dominated by at least one of the terms, either $\pi(Z_1,\boldsymbol{p}) \geq \pi(Z,\boldsymbol{p})$ or $\pi(Z_2,\boldsymbol{p}) \geq \pi(Z,\boldsymbol{p})$. We now check the two cases with the dominating strategy. After finitely

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

many steps, we end up in Case 1. □

5.3 Impact

The result has two major consequences. On the one hand, it is of conceptional significance as we provide sufficient conditions for when a player is allowed to remember less than their opponent without loss of payoff. First, the opponent needs to play a reactive-n strategy. Secondly, we require that the stage game is additive. Note that while additivity of the game is restrictive, it has played a crucial role in applied research, especially in evolutionary game theory (see [MRH21], [MFH14], [CP23]).

On the other hand, our result further reduce the number of conditions we need to check to determine whether a reactive strategy is a best response. The reduction is significant as the following Table 5.1 shows. It compares number of conditions needed as calculated in Chapter 3 to the reduced number as obtained in this chapter.

n-history	**1**	**2**	**3**	**4**	**5**
S_n^{pure}	8	64	2 048	1 048 576	137 438 953 472
$Cyc(n-1)$	2	3	6	19	179

Table 5.1: The table compares the number of pure self-reactive-n strategies to the number of cycles in the $(n-1)$-dimensional de Bruijn graph in the donation game.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

6 Discussion

Nash equilibria are a central concept in game theory. A symmetric Nash equilibrium is defined as a strategy that is a best response against itself. Best responses are strategies that achieve the maximum possible payoff against a fixed opponent. Thus, when playing against a symmetric Nash equilibrium, adopting the equilibrium's strategy results in obtaining the maximum possible payoff. If both players use the same symmetric Nash equilibrium, neither can improve their payoff by switching to a different strategy unilaterally. This makes them of great interest, see [Nas50] and [HR04]. Suppose a best response against a strategy $\boldsymbol{p}$ is known. In order to check whether or not $\boldsymbol{p}$ is a symmetric Nash equilibrium, it suffices to compare the best response's payoff against $\boldsymbol{p}$ with the payoff of $\boldsymbol{p}$ against itself. This is why efficiently computing best responses has been the central focus of this thesis.

In Chapter 2, we introduced major concepts for the study of repeated interactions. To this end, we defined games and repeated games before focusing on strategies. As we are particularly interested in studying strategies with bounded recall, i.e., those who can remember at most finitely many past events, we defined finite memory strategies and memory-n strategies. A subset of memory-n strategies are reactive-n strategies. Players who engage in these strategies only recall their opponents' past moves, which makes them both cognitively plausible and amenable to analysis. We further define self-reactive strategies as those that only recall their own past moves. They play a crucial role in the identification of best responses against reactive strategies.

In Chapter 3, we consider a general repeated game and a fixed but arbitrary reactive-n strategy. The work of [LNZ20] proves that in this setting there exists a best response that is a pure self-reactive-n strategy. In [GANH24], they provide an independent proof in the context of reactive-n strategies in the repeated prisoner's dilemma, a two action game. We extend their proof to games with an arbitrary, finite set of actions. This provides an alternative proof of the result in [LNZ20]. Our result not only establishes the existence of a best response but also reduces the search for a Nash equilibrium from an infinite set to a finite one.

The problem arises that the number of calculations though finite is still large already for intermediate memory sizes. Thus, we aim to reduce this set further. In Chapter 4, we observe that strategies with the same long-term behavior obtain the same payoffs, independent of their reactive opponent. We further analyze that these long-term behaviors correspond to cycles on the de Bruijn graph. To be precise, there exists a bijection between them, after the long term behavior, factorized pure self-reactive strategies and the cycles on the de Bruijn graph. This allows us to define payoffs of cycles. This approach results in a significant reduction of the number of calculations needed in the quest to find a best response. For example, given a

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

reactive-3 strategy and a two action game, we only need 19 calculations of payoff instead of 2048. Table 4.1 and Table 4.2 provide additional evidence of the result's significance.

In Chapter 5, we specifically focus on additive games. The new cyclic approach is the foundation needed to be able to proof the existence of a best response among strategies of lower memory. To this end, we note that using cycles to compute payoffs separates the players' action probabilities into distinct summands, rather than combining them multiplicatively. Further, if a cycle is not well-defined as a cycle on the graph of one dimension lower, we can separate the problematic histories and obtain two disjoint cycles. The payoff of the original cycle is then a convex combination of the two disjoint cycles, thus allowing at least one of them to dominate the original cycle. Continuing this argument on the dominant cycle will after finitely many steps lead to a cycle that is well-defined on the graph of lower dimension and thus proving the assertion. Not only is this result remarkable on a conceptional level, it further reduces the number of conditions to check to obtain a best response. As an example, consider a reactive-4 strategy and the two action donation game. We proved in Chapter 3 that calculating 1 048 576 payoffs is sufficient, in Chapter 4 we reduced the number to 179. However, by Chapter 5 it now suffices to calculate 19 payoffs. Further reductions are illustrated in Table 5.1.

6.1 Open Problems

In our setting of repeated games we assume that at each iteration the next game happens with probability 1. An alternative presumption is embedded in the theory of discounting. Here, for every game the next game occurs with probability $\delta < 1$. While we did not discuss discounting in games, it is worth mentioning that Algorithm A can be adapted to receive similar results as we have in Chapter 4. In discounting games, two parts of the pure self-reactive strategy are important for the calculation of payoffs. On the one hand, the payoff received in the long-term behavior which is calculated as a combination of a shifted geometric series over δ and the payoffs obtained in the cycle. On the other, we need to calculate the payoffs of the extended initial moves, those that happen before the long term behavior is reached. All possible payoffs are obtained by combining all possible paths to a vertex (without reaching a node twice) with all possible long term behaviors that start at that vertex.

Furthermore, what remains to explore is to what extend the factorization reduces the number of conditions. This breaks down to the combinatorial problem of finding a closed form for the number of cycles in the n-dimensional de Bruijn graph over an m-element alphabet.

Additionally, it has yet to be shown that testing against all cycles in the de Bruijn graph is not only sufficient but also necessary. This could be proven by constructing for every cycle a game and a reactive-n opponent in which and against which the cycle is the unique best response.

On a broader scale, the question arises whether similar results can be achieved in multiplayer games.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Bibliography

[AB87] T. van Aardenne-Ehrenfest and N. de Bruijn. *Circuits and Trees in Oriented Linear Graphs*. In *Classic Papers in Combinatorics*. 1987, pages 149–163. DOI: 10.1007/978-0-8176-4842-8_12.

[AH81] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981. DOI: 10.1126/science.7466396.

[Bru46] N. de Bruijn. A combinatorial problem. *Proc. K. Ned. Akad. Wet. Ser. A*, 49(7):758–764, 1946.

[CP23] M. C. Couto and S. Pal. Introspection dynamics in asymmetric multi-player games. *Dyn. Games Appl.*, 13(4):1256–1285, 2023. DOI: 10.1007/s13235-023-00525-8.

[CPT11] P. E. C. Compeau, P. A. Pevzner, and G. Tesler. How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*, 29:987–91, 2011. DOI: 10.1038/nbt.2023.

[GANH24] N. E. Glynatsi, E. Akin, M. A. Nowak, and C. Hilbe. Conditional cooperation with longer memory. *Proceedings of the National Academy of Sciences*, 121(50):e2420125121, 2024. DOI: 10.1073/pnas.2420125121.

[GV16] J. García and M. van Veelen. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *J. Econom. Theory*, 161:161–189, 2016. DOI: 10.1016/j.jet.2015.11.007.

[HMCN17] C. Hilbe, L. A. Martinez-Vaquero, K. Chatterjee, and M. A. Nowak. Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences*, 114(18):4715–4720, 2017. DOI: 10.1073/pnas.1621239114.

[HR04] C. A. Holt and A. E. Roth. The Nash equilibrium: A perspective. *Proceedings of the National Academy of Sciences*, 101(12):3999–4002, 2004. DOI: 10.1073/pnas.0308738101.

[LHG25] F. Lesigang, C. Hilbe, and N. E. Glynatsi. Can I afford to remember less than you? Best responses in repeated additive games. *Econom. Lett.*, 250:Paper No. 112300, 4, 2025. DOI: 10.1016/j.econlet.2025.112300.

[LNZ20] R. Levínský, A. Neyman, and M. Zelený. Should I remember more than you? Best responses to factored strategies. *Internat. J. Game Theory*, 49(4):1105–1124, 2020. DOI: 10.1007/s00182-020-00733-1.

[MFH14] W. Maciejewski, F. Fu, and C. Hauert. Evolutionary Game Dynamics in Populations with Heterogenous Structures. *PLOS Computational Biology*, 10:1–16, 2014. DOI: 10.1371/journal.pcbi.1003567.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

[MO11] G. J. Mailath and W. Olszewski. Folk theorems with bounded recall under (almost) perfect monitoring. *Games Econom. Behav.*, 71(1):174–192, 2011. DOI: 10.1016/j.geb.2010.11.002.

[MRH21] A. McAvoy, A. Rao, and C. Hauert. Intriguing effects of selection intensity on the evolution of prosocial behaviors. *PLOS Computational Biology*, 17:1–21, 2021. DOI: 10.1371/journal.pcbi.1009611.

[MS06] G. J. Mailath and L. Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, 2006. DOI: 10.1093/acprof:oso/9780195300796.001.0001.

[Nas50] J. F. Nash Jr. Equilibrium points in n-person games. *Proc. Nat. Acad. Sci. U.S.A.*, 36:48–49, 1950. DOI: 10.1073/pnas.36.1.48.

[Now06] M. A. Nowak. Five Rules for the Evolution of Cooperation. *Science*, 314(5805):1560–1563, 2006. DOI: 10.1126/science.1133755.

[NS90] M. Nowak and K. Sigmund. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematica*, 20:247–265, 1990. DOI: 10.1007/BF00049570.

[PD12] W. H. Press and F. J. Dyson. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26):10409–10413, 2012. DOI: 10.1073/pnas.1206569109.

[RR18] J. Romero and Y. Rosokha. Constructing strategies in the indefinitely repeated prisoner's dilemma game. *European Economic Review*, 104:185–219, 2018. DOI: https://doi.org/10.1016/j.euroecorev.2018.02.008.

[Sig10] K. Sigmund. *The calculus of selfishness*. Princeton University Press, Princeton, NJ, 2010. DOI: 10.1515/9781400832255. URL: https://doi.org/10.1515/9781400832255.

[Ued21] M. Ueda. Memory-two zero-determinant strategies in repeated games. *Royal Society Open Science*, 8:202186, 2021. DOI: 10.1098/rsos.202186.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub