Equilibria in Finite Games

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Anshul Gupta

Department of Computer Science

November 2015

Supervisory team:

Dr. Sven Schewe (Primary) and Prof. Piotr Krysta (Secondary)

Examiner committee:

Prof. Thomas Brihaye and Prof. Karl Tuyls

Contents

Illustrations

List of Figures

List of Tables

Notations

The following abbreviations and notations are found throughout this thesis:

MPG	Mean-payoff game
2MPG	Two-player mean-payoff games
MMPG	Multi-player mean-payoff games
DSG	Discounted sum game
2DSG	Two-player discounted sum games
MDSG	Multi-player discounted sum game
DBA	Deterministic Büchi automata
SP	Strategy profiles
LSP	Leader strategy profiles
ISP	Incentive strategy profiles
PISP	Perfectly incentivised strategy profiles
SCC	Strongly Connected Component
Nash SP	Nash strategy profiles
NE	Nash equilibria
LE	Leader equilibria
IE	Incentive equilibria

FIE	Friendly incentive equilibria
SIE	Secure incentive equilibria
fpayoff	Follower payoff in a strategy profile
lpayoff	Leader payoff in a strategy profile

Abstract

This thesis studies various equilibrium concepts in the context of finite games of infinite duration and in the context of bi-matrix games. We considered the game settings where a special player – the *leader* – assigns the strategy profile to herself and to every other player in the game alike. The leader is given the leeway to benefit from deviation in a strategy profile whereas no other player is allowed to do so. These leader strategy profiles are asymmetric but stable as the stability of strategy profiles is considered w.r.t. all other players. The leader can further incentivise the strategy choices of other players by transferring a share of her own payoff to them that results in incentive strategy profiles. Among these class of strategy profiles, an 'optimal' leader resp. incentive strategy profile would give maximal reward to the leader and is a leader resp. incentive equilibrium. We note that computing leader and incentive equilibrium is no more expensive than computing Nash equilibrium. For multi-player non-terminating games, their complexity is NP complete in general and equals the complexity of computing two-player games when the number of players is kept fixed. We establish the use of memory and study the effect of increasing the memory size in leader strategy profiles in the context of discounted sum games. We discuss various follower behavioural models in bi-matrix games assuming both friendly follower and an adversarial follower. This leads to friendly incentive equilibrium and secure incentive equilibrium for the resp. follower behaviour. While the construction of friendly incentive equilibrium is tractable and straight forward the secure incentive equilibrium needs a constructive approach to establish their existence and tractability. Our overall observation is that the leader return in an incentive equilibrium is always higher (or equal to) her return in a leader equilibrium that in turn would provide higher or equal leader return than from a Nash equilibrium. Optimal strategy profiles assigned this way therefore prove beneficial for the leader.

Acknowledgements

With all due respect, I thank God for giving me the opportunity to pursue my Ph.D. at the University of Liverpool.

I sincerely and deeply thank my supervisor, Sven Schewe, for his guidance, patience, and motivation throughout my years of study. He has been very supportive in many different ways and has constantly encouraged me during all these years. His immense knowledge has always helped me to learn a lot from him. His friendly behaviour and positive attitude are exemplary and I regard it as my honour to have done my Ph.D. under his excellent supervision.

I would like to thank Alexei Lisitsa, Martin Gairing, and Dominik Wojtczak, for being my academic advisors, and Piotr Krysta, for being my second supervisor. I thank them for all their advice and useful ideas. My thanks to Dominik for taking his time to read my papers. His suggestions and feedback has always been valuable. I want to thank Ashutosh Trivedi for numerous discussions related to mean-payoff games. It was really nice to work on a paper together with him. I would like to thank Thomas Brihaye and Karl Tuyls for kindly agreeing to be my examiners and for their insightful comments on my thesis. It was an enriching discussion session with both of them on my thesis and their feedback has really helped me to improve my thesis.

My thanks goes to the Department of Computer Science for supporting my Ph.D. and for providing a wonderful research environment.

Finally, I thank my family, for all their care and support. My special thanks are for my parents for always inspiring me to aim high and for showering upon me their unconditional love and affection. Last but the most important one, I wish to thank my husband Vivek for always being there through thick and thin and for being the most supportive person. He has always been extremely caring and loving. I cannot recall a single moment when he was not able to co-operate with me. I thank him for doing all those little things like cooking a meal or making a cup of tea and for taking time from his work just to accompany me on my conference trips. All these small gestures are indeed the things that matters most.

Preface

I declare that this thesis is composed by myself and that the work contained herein is my own, except where explicitly stated otherwise, and that this work was undertaken by me during my period of study at the University of Liverpool, United Kingdom. This thesis has not been submitted for any other degree or qualification except as specified here.

Main results of this thesis are contained in chapters 3, 4 and 5 and have already been published. Most parts of Chapter 3 has been published in AAMAS 2015 [GS15]. Chapter 4 is based on two conference publications. Parts of Chapter 4 are published in TIME 2014 [GS14] and some other parts of the chapter are accepted for publication at SEFM 2016 [GST+16]. Chapter 5 has been published in GandALF 2015 [GSW15].

A journal article, titled *'Buying Optimal Payoffs in Bi-Matrix Games'*, based on the conference paper [GS15], is presently under submission.

Chapter 1

Introduction

We study leader and incentive equilibrium in the context of bi-matrix games and multi-player infinite duration games. In the studied game settings a designated player (the *leader*) is in a position to assign the strategy profile to herself and to every other player in the game alike. All other players who merely follow the strategy profile as assigned by the leader are called as *followers*. This is in contrast to the Nash equilibrium that are defined symmetrically as here the symmetric condition in a strategy profile is relaxed for the leader. We say a strategy profile is stable if no one except the leader has an incentive to deviate. Stable strategy profiles assigned this way are called *leader strategy profiles*. A leader strategy profile is considered 'optimal' if it provides maximal reward to the leader. An optimal leader strategy profile is called a *leader equilibrium*.

We further propose a natural generalisation of leader strategy profiles – *incentive strategy profiles* – in multi-player games and in bi-matrix games. In an incentive strategy profile, the leader is further allowed to influence the behaviour of other players in the game. The leader can incentivise her followers to follow an assigned strategy profile. For this, she transfers part of her own payoff to them. We can say that the leader gives these non-negative incentives to others in order to make them comply with the assigned strategy profile. An incentive strategy profile that gives maximal reward to the leader is considered 'optimal' and is called an *incentive equilibrium*. Stability in an incentive equilibrium is considered in the same way as in a leader equilibrium, but now incentives are also taken into account.

As a general convention followed in this thesis, we refer to the leader player as *'she'* and her follower(s) as *'he'*. We also make the common assumption on the behaviour of players that they are pure 'rational' individuals in that they tend to maximise their own payoff. The game settings we study are leader centric and the focus is therefore to maximise the leader's reward. We first discuss the context in which this thesis is studied in the following sections and then briefly discuss the equilibrium concepts outlined above in Section 1.3. We give main contribution of our work in Section 1.4 and discuss the related work in Section 1.5. We give an outline of this thesis in Section 1.6.

1.1 Bi-matrix Games

We study incentive equilibria (and its variants) in bi-matrix games. A bi-matrix game is a finite strategic form (or: normal form) two-player game in which both players have a finite set of available actions. The game is a strategic game, as it involves strategic interaction among the participating entities, known as *players*. The strategic game is a way of describing the game using a matrix and is used to describe the game where players make simultaneous choices. The *payoff* or *utility* given to each player is represented here in a matrix – hence the name – bi-matrix game.

There is one *row player* and one *column player*. The 'row player' actions are identified by the rows and the 'column player' actions are identified by the columns of a bi-matrix. The strategy of the row player is to select a row, while the strategy of the column player is to select a column. If a player selects an action deterministically, then the selected strategy is called *pure*. By playing a pure strategy, players select what action to choose from a finite set of available actions.

A bi-matrix game with m pure strategies of the row player and n pure strategies of the column player can be represented by payoff matrices of size $m \times n$. The payoff matrices of two players can be combined into one payoff bi-matrix of the same dimension. The objective of both players in a bi-matrix game is to maximise their resp. payoff (or: utility) from a selected strategy profile.

An example of a bi-matrix game is shown in Table 1.1. The set of available actions for the row player is (S, T) and the set of available actions for the column player is (P, R). Once pure strategies are selected, each player receives the payoff value determined by the entry in the selected (row and column) box of the bi-matrix. The first value in the selected box refers to the payoff of the row player. The second value of the selected box refers to the payoff of the column player. For example, if the row player selects S, and the column player selects P, then the strategy profile is (S, P). The payoff of the row player and the column player from this strategy profile are 0 and 2 respectively. If each of the pair in the payoff matrix adds up to zero, then the game is a zero-sum game. Otherwise, it is a non-zero sum game.

	P	R
S	$0, 2$	$4, 1$
T	$1, 4$	$1, 1$

TABLE 1.1: A bi-matrix example.

The players are allowed to make a randomised decision by playing a probability distribution over the rows or columns. Such randomised strategies are called *mixed*. A mixed strategy can therefore be viewed as a probability distribution over a player's pure strategy set. The combined strategy selection by both players is termed as a *strategy profile*. For a mixed strategy profile of a player, the sum of all probabilities defined

over pure strategies should be equal to 1. A pure strategy can therefore be viewed as a special case of a mixed strategy profile where one action is played with probability 1.

When using a mixed strategy, the payoff to a player is the expected payoff induced by the payoff matrix and the probability distributions induced by the mixed strategies. If we refer to Table 1.1, a mixed strategy profile is, where player 1 plays S with 75% chance and T with 25% chance. Similarly, player 2 plays P with 75% chance and R with 25% chance. The payoff of the row player and the column player from this mixed strategy profile are 1 and 2.125, respectively.

1.2 Finite games of infinite duration

We study leader and incentive equilibria in multi-player finite games that are of infinite duration. These are turn taking graph based games that are played on a finite directed graph, called the game arena. The game arena on which the game is played is defined as a tuple that consists of a finite set of players, a finite set of vertices and a finite set of directed edges. The vertex set is partitioned into various subsets such that every subset belongs to exactly one of the players. An integer reward function assigns an integer value to every edge.

There is a designated start vertex and a token is placed initially on it. Players take turns in creating a play. At every vertex, the player who owns the vertex will select an outgoing transition to push the token forward along an edge in the game arena. Whenever an edge is visited, every player receives a payoff value given on that edge transition. Players take turns in moving a token along the edges of the graph and successively create an infinite path. The infinite path thus formed is called a *play*. Every player then receives a payoff value based on how the infinite path is evaluated (see below). The objective of the players is to maximise their aggregated reward or aggregated payoff value of the infinite path.

An infinite sequence can be aggregated into a single value that is the value of a play. For our study, we focus on the following two mechanisms for aggregating an infinite sequence. First is to consider the *mean-payoff value* (or: *limit-average value*) and the second is to consider the *discounted-payoff value*. The former way is used in mean-payoff games and the latter is used in discounted-payoff games. Mean-payoff games and discounted-payoff games are examples of quantitative games in that players have *quantitative objectives* in these games. They are similar in the way how they are played, but they differ in how an infinite path is evaluated. For mean-payoff games, we study leader equilibria and extend our study to incentive equilibria as well. For discounted-payoff games, we study leader equilibria in detail. More specifically, we study bounded memory strategy profiles in these games and establish the use of memory. We now introduce each of these game type in detail.

Mean-payoff Games. Mean-payoff games were first introduced by Ehrenfeucht & Mycielski in [EM79]. These are the quantitative games where players have an objective

of maximising their mean-payoff value from an infinite path. Classically, these are two-player zero-sum games. They are zero-sum games in that sum of the rewards on every edge add up to zero. In two-player games, the vertex set is bipartite and partitioned among the two players. The game is played as follows. A token is placed initially on a designated start vertex. The two players – Maximiser (Max) and Minimiser (Min) – take turns in moving the token, depending on whether the vertex is placed at a 'Max' vertex or at a 'Min' vertex. At every vertex, the player who owns the vertex will select an outgoing edge to move the token to the next vertex. This way, the players jointly construct an infinite path called a *play*. The value of a play is evaluated using the limit-average or mean-payoff function. Each player receives a value that is the mean average of the sum of their rewards on the edges visited in the infinite path. For any play, player 'Max' wins a value that is the average of an infinite play. The player 'Min' loses this value. Two-player mean-payoff games were shown to be positionally determined [EM79], i.e., for every vertex v, there is a value $\mathbf{v}(v)$, such that players 'Max' and 'Min' guarantee payoff values $\geq \mathbf{v}(v)$ and $\leq \mathbf{v}(v)$ respectively. The mean-payoff games therefore asserts the existence of optimal positional strategies and given that players follow the positional strategies, the path followed would lead to a simple loop, where it would stay forever. For a mean-payoff game $\mathcal{G}$, if v_0 is an initial vertex and an infinite path $\pi = \langle v_0, v_1, \ldots \rangle$ is generated, then the value of the play π is computed as follows

$$\mathcal{G}(\pi) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} r(e_i).$$

Here, e_i is the i^{th} edge transition and $r(e_i)$ is the reward incurred upon taking this edge transition. In multi-player mean-payoff games, there is a finite set of players and every player follows an objective of maximising their limit-average reward from an infinite play. Every player receives a payoff value that is the average of individual rewards accumulated over an infinite path. Note that the multi-player games and all games with two players in it are not necessarily zero-sum games. Two-player zero-sum games are the games where players have complete antagonistic objectives.

As an example, we now refer to the game graph as shown in the Figure 1.1. It represents a multi-player mean-payoff game with three players – player 1, player 2 and player 3. Vertices 1 and 4 belong to player 1, vertices 2 and 5 belong to player 2 and vertex 3 belongs to player 3. Vertex 1 is taken as an initial vertex and is denoted with an incoming arrow. Edges in the game graph are annotated with reward vector that shows reward of player 1, player 2, and player 3 in the respective order. Note that rewards on the edges that are not part of an infinite path are not shown, as their reward does not hold any significance on player's overall reward from a play.

At vertex 1, if player 1 moves the token to vertex 4, then it would result in an infinite path $\langle 1 \cdot 4^{\omega} \rangle$ with an overall reward of 1 for the player 1 and a reward of 0 for both player 2 and player 3. However, if at vertex 1, player 1 always move the token to vertex 2, and player 2 always move the token to vertex 1, then the resultant infinite

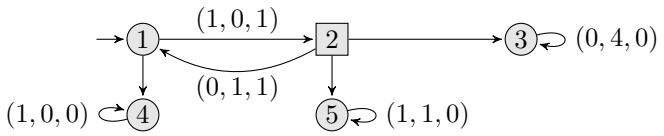

FIGURE 1.1: A multi-player mean-payoff game.

path would be $\langle 1 \cdot 2 \rangle^\omega$ with an overall reward of 0.5 for both player 1 and player 2, and a reward of 1 for the player 3.

Mean-payoff games are interesting to study because of their rare complexity status and for the number of applications that they enjoy. Mean-payoff games have their place in the synthesis and analysis of infinite systems, in logics and games, in the quantitative modelling and verification of reactive systems [Hen13]. Their limit-average criteria made them suitable to model distributed development of systems where several individual components interact among themselves. The objective of the individual components is to maximise their limit-average share of frequency with which they use a shared resource.

Discounted-payoff Games. Discounted-payoff games (or: discounted sum games) form another important class of infinite games with quantitative objectives. Intuitively, they are played in a similar manner to mean-payoff games, but the evaluation function used to evaluate these games is different. The game is played on a finite directed graph where each vertex is owned by exactly one of the players. The game arena consists of a finite set of players, a finite set of vertices and a finite set of directed edges. Initially, a token is placed on a start vertex. Whenever the token is on a vertex, the player who owns this vertex will select an outgoing edge and move the token along this edge. This way, players take turns to jointly construct an infinite play. An integer reward function assigns an integer value to every edge. Every edge in the game graph is annotated with a reward vector. Each value in the reward vector depicts a reward given to a player, whenever that edge transition is taken. In addition, there is a real-valued discount factor λ whose value lies between 0 to 1. The payoff given to players at every edge transition is discounted by λ, where $0 < \lambda < 1$. For a play π, every player receives a reward that is an aggregated sum of their individual discounted rewards over edge transitions taken in π. The objective of each player is therefore to maximise their individual discounted sum of rewards.

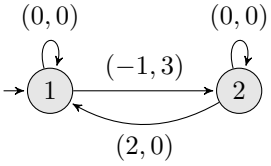

FIGURE 1.2: A discounted-payoff game with discount factor $\frac{1}{2}$.

Thus, for a discounted-payoff game $\mathcal{G}$, we can aggregate the reward over an infinite path $\pi = \langle v_0, v_1, \ldots \rangle$ as

$$\mathcal{G}(\pi) = \sum_{i=0}^{\infty} \lambda^i r(e_i)$$

As an example, we consider the discounted-payoff game from Figure 1.2 with discount factor $\frac{1}{2}$. In this game, the vertex labelled 1 is owned by the player 1 and vertex labelled 2 is owned by the player 2. Assume that, whenever the token is at vertex 1, player 1 moves the token to vertex 2, and whenever the token is at vertex 2, player 2 moves the token to vertex 1. This would result in an infinite path $\langle 1 \cdot 2 \rangle^\omega$. For this path $\langle 1 \cdot 2 \rangle^\omega$, the discounted-payoff of player 1 is computed as follows:

$$-1 \times (0.5)^0 + 2 \times (0.5)^1 - 1 \times (0.5)^2 + 2 \times (0.5)^3 - \cdots = 0$$

Similarly, we compute discounted-payoff of player 2 as follows:

$$3 \times (0.5)^0 + 3 \times (0.5)^2 + 3 \times (0.5)^4 + \cdots = 4$$

Discounted-payoff games were introduced by Shapley in 1953 [Sha53]. He showed that every two-player discounted zero-sum game has a value and that they are positionally determined. It implies that, starting at any vertex v, optimal positional strategies exist for both players. Gimbert and Zielonka [GZ04] considered infinite two-player antagonistic games with popular payoff mechanisms like mean-payoff and discounted-payoff. They gave sufficient conditions to ensure that both players have positional (memoryless) optimal strategies. It is known that mean-payoff games are polynomial time reducible to discounted-payoff games [Con93, ZP96]. Their complexity therefore falls in the same complexity class as mean-payoff games. Discounted-payoff games have applications in temporal logics where important events are discounted in accordance with how late they occur. This aspect of discounted-payoff games has been studied in [dAFH$^+$04] and the importance of discounting has been discussed in detail in [dAHM03].

1.3 Equilibrium concepts in game theory

Game theory [OR94] is all about the strategic interaction of multiple rational players and has found its applications in fields as diverse as economics, political science, biology, and recently, in computer science. It gives a formal approach to model some real time situations as well. It provides a natural framework to study the behaviour of rational players when they interact among themselves strategically. Players are commonly referred to as 'rational' assuming that their moves are motivated by maximising their own payoff. Note that this is a common assumption made in classical game theory.

A central concept in game theory is to find an equilibrium in these games. That is, there should be some mechanism that defines the solution of a strategic game. In any finite strategic game, every player has a finite set of actions to select from. These sets of finite available actions comprise sets of pure strategies. If a player selects an action with

probability 1, then it is termed as a pure strategy. However, players are also allowed to mix their strategy choices. If a player chooses a probability distribution over the set of pure strategies, then the resultant strategy is a mixed strategy. The combined set of strategy selection – one by each player – is termed as a *strategy profile*. An obvious solution approach would ask for – a stable strategy profile – to which every player in the game agrees. This stability of a strategy profile is defined in terms of an equilibrium, i.e., a situation where players are prepared to follow the strategy profile as it is.

The most widespread concept of equilibria was given by John Nash in 1950. A strategy profile is in a Nash equilibrium if no player can do better by changing his or her strategy alone. Nash showed that in any finite strategic game, at least one mixed strategy equilibrium always exists [Nas50]. The term became popular as *Nash equilibrium*.

In a Nash equilibrium, every player agrees to play their intended strategies because they do not have an incentive to deviate from the strategy profile. It depicts rational behaviour of players at its minimal in that players only have to satisfy least criterion of stability that there is no incentive for unilateral deviation. It was shown in [DGP09] that the complexity of computing a mixed strategy Nash equilibrium is PPAD complete. As an example of Nash equilibrium, we now refer to the Table 1.1. A pure Nash equilibrium in this bi-matrix is a strategy profile where row player plays T and column player plays P receiving a payoff of 1 and 4 respectively. Note that no Nash equilibrium in mixed strategies exist in this example.

Von Stackelberg in 1934 gave the concept of *leadership* or *Stackelberg* models [vS34]. They are also known under the term *commitment* models. In a Stackelberg model, one player acts as a leader and the other acts as a follower. The model – unlike a Nash model – is a sequential model. Here, the leader commits to a strategy first and after observing the action chosen by the leader, the follower takes the next turn. The main objective of both players is to maximise their own return. The Stackelberg model makes some basic assumptions, in particular that the leader knows ex-ante that her follower is observing her action. It assumes that both players are rational in that they try to maximise their own return. The solution concept became popular under the term *Stackelberg equilibrium* or *leader equilibrium*. While players select their strategies simultaneously in Nash equilibrium, players move sequentially in a leader equilibrium. Computing them is computationally cheap, as the construction of a leader equilibrium is known to be tractable [CS06].

1.3.1 Solution concepts

We have already introduced the game types that we have studied in this thesis. Although we study a variety of games – bi-matrix games, mean-payoff games and discounted-payoff games – our game settings remain essentially the same.

Our Game Settings. In our game settings, the leader assigns a strategy profile and therefore it is natural to assume that she may not like to stay within Nash restriction.

The leader may therefore select a strategy profile where no follower has an incentive to deviate, while it is explicitly allowed that the leader herself may have an incentive to deviate. The Nash requirement of a stable strategy profile, thus, does not apply to the leader. Given that leader holds this extra power over the game, she can also incentivise various strategy choices of her follower. We therefore allow the leader to transfer part of her own payoff to her followers.

Approach. Our approach focuses on leader centric situations. We intend to compute optimal strategy profiles in these games using different solution approaches, like allowing the leader to benefit from deviation. Our primary objective is to maximise the the leader's payoff. We aim at computing strategy profiles that are optimal w.r.t. the payoff of the leader. For this, we allow the leader to assign strategy profiles in the game. As said above, the leader is allowed to break the Nash symmetry in a strategy profile such that the Nash condition does not hold for her. Thus, the set of 'leader strategy profiles' from a broader class, giving her more leeway in selecting an optimal strategy profile. Although the leader might benefit from deviation, note that no other player is allowed to do so. In an incentive strategy profile, the leader pays a non-negative incentive amount, say $\iota \geq 0$, to each of her followers. This incentive amount is now added to the overall payoff of the resp. follower, and deducted from the payoff of the leader. Thus, the overall payoff of resp. follower is increased by the amount ι, but only if he follows the strategy profile. Otherwise, i.e., if the follower deviates from the assigned strategy profile, his payoff is not affected. Similar to the leader strategy profiles, the leader can only assign strategies such that no follower benefits from deviation. The leader here has more power as compared to the traditional leader equilibrium. In particular, any leader strategy profile, can be viewed as an incentive strategy profile (with zero incentive).

Leader equilibrium. An optimal strategy profile among the class of leader strategy profiles is a leader equilibrium. As an example, we now refer to the multi-player mean-payoff game from Figure 1.1. We assume that player 1 owns vertex 1, leader owns vertex 2, and player 3 owns vertex 3. The reward vectors shown on the edges depict the reward of the players in this order - player 1, leader, and player 3. Initially, at vertex 1, player 1 can either move the token to vertex 4 or to vertex 2. At vertex 2, the leader has an incentive to move the token to vertex 3, because this would give her a maximal payoff of 4. However, player 1 receives a lower payoff at vertex 3 than at vertex 4. Therefore, player 1 would prefer moving the token initially to vertex 4, and not to vertex 2. This would result in the strategy profile $\langle 1 \cdot 4^{\omega} \rangle$. Note that this is also the only Nash equilibrium in this game. It gives the payoff of 1 to the player 1, whereas, both the leader and player 3 receive a payoff of 0.

However, the leader has a better strategy: she can select a strategy profile, where, she might benefit from deviation. In an optimal leader strategy profile, the leader would move the token from vertex 2 to vertex 5. Therefore, a leader equilibrium whose outcome is $\langle 1 \cdot 2 \cdot 5^{\omega} \rangle$ would provide an overall payoff of 1 for both the leader and the player 1.

Note that the leader receives better reward in the leader equilibrium as compared to the only Nash equilibrium.

Incentive equilibrium. An optimal strategy profile among the class of incentive strategy profiles is an incentive equilibrium. As an example, we refer to the bi-matrix game from Table 1.1. The row player is the leader and the column player is her follower. In this example, the only Nash equilibrium is the strategy profile (T, P) with a payoff of 1 for the leader and a payoff of 4 for the follower. Note that, in this example, both Nash and leader equilibrium provide the same leader return. However, the leader can improve her payoff in an incentive equilibrium. If the leader commits to a pure strategy S, then she can incentivise her follower to play the pure strategy R by giving him an incentive of amount $\iota = 1$. The strategy profile (S, R) then becomes an incentive equilibrium. The payoffs of the leader and the follower in this equilibrium are 3 and 2, respectively. Note that the leader pays only the minimal incentive that is needed to make the strategy profile stable. The follower reward in the strategy profile (S, R) therefore equals the follower reward in the strategy profile (S, P) such that the follower has no incentive to deviate from the incentive equilibrium. This also shows that the reward of the leader from an incentive equilibrium is better compared to her reward from a Nash or leader equilibrium. Note that this is a general observation and holds in all cases.

Related equilibria. For bi-matrix games, we consider different assumptions on the behaviour of the follower. One assumption is that the follower is friendly towards his leader in that he chooses, ex-aequo, the strategy assigned by the leader. This behaviour puts an obligation over the leader in that the leader also responds friendly. Therefore, the leader would select a strategy profile that provides ex-aequo follower return. We discuss the resulting equilibria under the term *friendly incentive equilibria*. In any friendly incentive equilibrium, the leader would follow a secondary objective of maximising the follower return. For the strategy profiles that gives an equal return to the leader, she would use the follower return as a tie-breaker.

Another assumption is that the follower acts adversarial in that he may try to harm the leader. Here, the conditions from incentive equilibrium need to be strengthened further. We introduce *secure incentive equilibria* for this case (conditions here are comparable to the secure Nash equilibria). We also introduce *ϵ-optimal incentive equilibrium* for the general case where no secure incentive equilibrium exists. Note that, in a secure strategy profile, the leader can only assign a strategy profile, where every deviation of the follower must either lead to a strict decrease of the follower's return, or not affect the leader return adversely. These are defined on a similar basis as secure Nash equilibria [CHJ06]. We therefore use the term *secure incentive equilibria*.

We note that the construction of all of these incentive equilibria is tractable and the solution concept is therefore computationally cheap.

Reward and Punish strategy profiles. We use *reward and punish strategy profiles* as a means to construct optimal leader strategy profiles and incentive strategy profiles.

We note that the reward and punish strategy profiles can be used as a means to maximise the payoff of the leader from a given strategy profile. They can be used by the leader to dictate the play in a game. In a reward and punish strategy profile [Fri77], the leader promises an optimal reward to every player, but, if a player deviates from the assigned strategy profile, then the leader forms a coalition with all other players to act against the deviating player. That is, the leader initially cooperates with all players to produce a strategy profile. But, if a player deviates, then all other players co-operate to harm this player and neglect their own interests. Thus, while the objective of the deviating player is still to maximise his reward from the strategy profile, the objective of all other players (including the leader) is now changed to minimise the payoff of the deviating player. This would then result in a two-player game where the players have antagonistic objectives. An essential reward criteria on these strategy profiles is as follows. If a player deviates at some vertex, then the overall reward of the player from the resultant two-player game that start at the point of deviation is not higher than the player's reward from the assigned strategy profile. We use reward and punish strategy profiles as a tool to construct strategy profiles that are optimal w.r.t. the leader.

Motivaton. The model we study allows us to consider quantitative specifications where studying a leader of this type has natural justifications. As an example, one could consider a distributed component system where several rational components interact with each other and with a rational controller. The components are considered 'rational' as they try to maximise their individual utilities and a 'rational' controller would try to maximise overall system utility. These properties can be reflected by an automata where individual processes try to maximise the amount of time they spend in an accepting state. Using mean-payoff objectives to represent this, one could encode this as maximising the limit-average time a process is in an accepting state. The rational controller would try to maximise the limit-average time a system's critical resource is being used. As the objectives are not completely antagonistic, the techniques discussed here can be applied to arrive at an optimal solution.

1.4 Contribution

Most of our results are regarding the existence of leader and incentive equilibria in bi-matrix (two players) games and in multi-player turn taking games (with quantitative objectives) and their complexity.

For bi-matrix games, we introduce *incentive equilibria* as a generalisation of *leader equilibria* (Chapter 3 and [GS15]). We show that the leader can improve her reward in incentive equilibria without adding to the computational cost – incentive equilibria are computationally tractable just like leader equilibria. We also contribute conceptually by discussing behavioural assumptions of the follower in incentive equilibria. We discuss the implications of both friendly follower and an adversarial follower. These different implications lead to friendly incentive equilibria and secure incentive equilibria

respectively. We give an algorithm for the computation of friendly incentive equilibria in bi-matrix games. We also report experimental results. We evaluate our solution approach on randomly generated bi-matrix games. For this, we consider 100,000 data-sets each for games with continuous payoff values and games with integer payoff values for the evaluation of friendly incentive equilibria. Our results show that incentive equilibria are superior over leader equilibria and leader equilibria are superior over Nash equilibria.

In multi-player non-terminating games, we contribute by establishing various results. We introduce the concept of *leader strategy profiles* and *leader equilibria* in multi-player mean-payoff games (Chapter 4 and [GS14]). Leader strategy profiles are based on the traditional *reward and punish strategy profiles* [Fri71, BDS13]. In a reward and punish strategy profile, the leader assigns a strategy profile, and the first player who deviates from the strategy profile is punished. We show that solving multi-player mean-payoff games is polynomial time reducible to solving two-player mean-payoff games. We establish the existence of leader equilibrium and show that no Nash equilibrium is superior (Note that this is an obvious implication from the fact that each Nash equilibrium is, in particular, a leader equilibrium). We give a constraint system that can be used to construct an optimal leader strategy profile that provides the maximal leader return.

We establish the NP-completeness of the related decision problem 'is there a leader equilibrium with payoff greater or equal to a threshold' (which equals the bound for Nash equilibria [UW11]). We show that the NP-hardness depends on the number of players: for a bounded number of players, we give a polynomial time reduction to solving two-player mean-payoff games. The complexity of finding leader and Nash equilibria for a bounded number of players therefore directly relates to the complexity of solving two-player games. There are algorithms for solving two-player games in pseudo polynomial time [BCD$^+$11], in smoothed polynomial time [BEF$^+$11] and in PPAD [EY10]. Then, there are fast randomised [BV07] and deterministic [Sch08] strategy improvement algorithms, and the decision problem is in UP$\cap$CoUP [Jur98, ZP96].

We contribute by introducing *incentive equilibria* in multi-player mean-payoff games (Chapter 4 and [GST$^+$16]). One fundamental result is the existence of incentive equilibria in these games. The decision problem related to constructing incentive equilibria is shown to be NP-complete. When the number of players is kept fixed, the complexity of the problem falls in the same class as two-player mean-payoff games. We give results from a tool to evaluate multi-player mean-payoff games. The co-authors Maram Sai Krishna Deepak and Bharath Kumar Padarthi from [GST$^+$16] have worked for the implementation of the tool. However, all the technical details throughout the paper is ours. We implement the strategy improvement algorithm from [Sch08] for finding the mean partitions and extend it to evaluate two-player mean-payoff games. We extend the constraint system from [GS14] by adding incentives to the overall reward of the followers. We construct incentive equilibria by constructing these constraint systems.

Our results show that incentive equilibrium will only provide a better return to the leader than leader equilibrium. We show that the complexity of finding incentive

Nash equilibria	• NP complete for multi-player non-terminating games [UW11] • PPAD complete for bi-matrix games [DGP09]
Leader equilibria	• NP complete for multi-player non-terminating games [GS14, GSW15] • For fixed number of players, complexity equals that of solving two-player games [GS14, GSW15] • Tractable for bi-matrix games [CS06]
Incentive equilibria	• Complexity equals that of computing leader equilibria [GST$^+$16] • Tractable for bi-matrix games [GS15]
Friendly IE	• Tractable for bi-matrix games [GS15]
Secure IE	• Tractable for bi-matrix games (constructive proof is required to establish this) [Chapter 3]

TABLE 1.2: Summary of the complexity results for different equilibria.

equilibria and leader equilibria in multi-player mean-payoff games is same.

We establish the existence of *optimal bounded memory leader strategy profiles* in multi-player discounted-payoff games (Chapter 5 and [GSW15]). Here, we extend the use of leader equilibria to multi-player discounted-payoff games. We discuss the existence of optimal bounded memory leader strategy profiles. We show that in discounted-payoff games the leader can benefit from more memory and that there are cases where infinite memory is also needed. We mainly discuss the construction of strategies that use only bounded memory. We give a simple non-deterministic polynomial time approach for assigning reward and punish strategies that meet or exceed a given payoff bound for the leader and uses memory only within a given bound. We show that the decision problem whether a pure strategy with bounded memory that gives a reward greater than or equal to some threshold value exists is NP-complete.

We summarise the important results in the Table 1.2. In this table, 'Friendly IE' refers to 'friendly incentive equilibria' and 'Secure IE' to 'secure incentive equilibria'.

1.5 Related work

Our results concern the existence of leader equilibria and incentive equilibria. Some important notions of equilibria in game theory [OR94] are Nash equilibria and leader equilibria. John Nash introduced the concept of Nash equilibrium in 1950 [Nas50] and showed that at least one mixed strategy Nash equilibrium always exists in strategic games. It was shown in [DGP09] that computing mixed strategy Nash equilibrium in strategic games is PPAD complete. Von Stackelberg introduced leader equilibria that is also known as Stackelberg equilibria in [vS34]. It has been studied in depth in Oligopoly theory [Fri77]. Conitzer and Sandholm [CS06] studied the computation of Stackelberg strategies in bi-matrix games. Von Stengel and Zamir [vSZ04, vSZ10] studied leadership game with mixed strategies and showed that the possibility to commit to a strategy profile in bi-matrix games is always beneficial for the committing player. An endogenous game model where both players can offer side payment to each other has been studied in [MOJ05]. They have considered the simultaneous determination of contracts and

the game is played in stages. In regard to complexity, leader and incentive equilibria are computationally cheap solutions, as their construction is known to be tractable [CS06],[GS15].

Various work pertaining to the existence of Nash equilibria in infinite duration games is known. Ummels and Wojtczak [UW11] studied the complexity of determining Nash equilibria in limit-average games. They showed that the decision problem of finding a Nash equilibria is NP-complete for pure (not allowing randomisation) strategy profiles, while the problem is undecidable for arbitrary randomised strategies. The undecidability result of [UW11] for Nash equilibria in arbitrary randomised strategies can be easily extended to leader equilibria. Ummels [Umm08] analysed the complexity of Nash equilibria in infinite multi-player games with co-Büchi or parity winning conditions. He established the NP-completeness of Nash equilibrium in infinite games with these objectives. Ummels has studied the concept of subgame perfect equilibrium for the case of infinite games in [Umm06]. Brihaye et al. [BDS13] have studied the existence of simple Nash equilibria in non-terminating games with various mixed reward functions.

The reward and punish strategies that we study in Chapter 4 and 5 are inspired by similar strategies introduced in [Fri71]. We show in Chapter 4 that keeping the number of players fixed, the problem of finding equilibria in multi-player mean-payoff games, is polynomial time reducible to solving two-player mean-payoff games. Although mean-payoff games [ZP96] are positionally determined [EM79], it remains an open question whether there exists a polynomial time algorithm to solve these games. We use the strategy improvement algorithm from [Sch08] as an underlying algorithm to evaluate two-player mean-payoff games.

Shapley showed that every two player discounted zero-sum game has a value and that optimal positional strategies exist for both players [Sha53]. Fink [Fin64] generalised their work by showing that every discounted sum game has a Nash equilibrium. Berg and Kitti [BK13] studied subgame perfect pure strategy equilibria in discounted sum games. They analysed subgame perfect equilibria in games with perfect information.

We study incentive equilibria in bi-matrix games and mean-payoff games. The work related to this is a mechanism designer modelled as a player in the game who has the opportunity to modify the game [ASA10]. Their focus is on pure strategies only. Another related work is [MT04], where an external party, who has no control over the rules of the game, can influence the outcome of the game by committing to non-negative monetary transfers for the different strategy profiles. Stark, in [Sta89], studied how the introduction of altruism into non co-operative game settings can lead to an improved quality for both agents. Stark [Sta85] further discussed special altruism – 'a mutual altruism' and its role in various contexts.

The secure incentive equilibria that we study in Chapter 3 are defined on a similar basis as secure Nash equilibria. Secure Nash equilibria were studied for multi-player games with quantitative objectives in [CHJ06] and [DPFK$^+$14]. These results, however, pertain to secure Nash equilibrium only. Chatterjee et al. [CHJ06] have studied secure

Nash equilibria in two-player non zero-sum games. De Pril et al. [DPFK$^+$14] establish the existence of secure equilibria in multi-player perfect information turn-based games for games with probabilistic transitions and for games with deterministic transitions.

1.6 Outline of this thesis

We give formal definitions of equilibrium concepts introduced here in Chapter 2. We start with the discussion on incentive equilibria in bi-matrix games in Chapter 3 where we consider various follower behavioural models assuming a friendly and a non-friendly follower. They lead to friendly incentive equilibria and secure incentive equilibria respectively. We then discuss the existence of leader and incentive equilibria in multi-player mean-payoff games in Chapter 4. We discuss the existence of optimal bounded memory leader strategy profiles in multi-player discounted sum games in Chapter 5. We conclude and summarise the results of this thesis in Chapter 6.

Chapter 2

Definitions

We define here various equilibrium concepts that we study in this thesis. We give these definitions on general n-player strategic form (or: normal form) game and formally define a leader strategy profile, leader equilibrium, incentive strategy profile, and incentive equilibrium.

Definition 2.1. A game $\mathcal{G}$ in normal form consists of

- a set P of players $\{1, \cdots, n\}$.
- a set S_p of pure actions corresponding to each player $p \in \{1, \cdots, n\}$. A strategy profile σ for the game $\mathcal{G}$ is $\sigma = (s_1, s_2, \cdots, s_n)$ where $s_p \in S_p$ for $p = (1, \cdots, n)$.
- a reward function $u_p(\sigma) : S \to \mathbb{R}$ that gives a reward r_p to player p whenever a strategy profile $\sigma \in S$ is chosen. The set S consists of pure strategy profiles and is defined as a Cartesian product of the individual strategy sets S_p.

The way that the player p chooses an action is defined by a strategy σ_p. A family of strategies $\sigma = \{\sigma_p \mid p \in P\}$ is called a strategy profile. Given a strategy profile σ, we write σ_p for the strategy of player $p \in P$ in σ.

A strategy profile σ is pure if each player $p \in P$ plays an action s_p deterministically and is a mixed strategy profile if a player has chosen probability distribution over the set of the player's available actions. A mixed strategy for player p is defined as a distribution over S_p as $\sigma = (p_1.s_1, \ldots, p_k.s_k)$, i.e., $\sum_{i=1}^{m} p_k = 1$ (the sum of the weights is 1) and $p_i \geq 0$ for $1 \leq i \leq k$, if there are k pure strategies. We define support of a mixed strategy σ as the positions with non-zero probability, and denote it as $\mathsf{support}(\sigma) = \{j \leq k \mid \sigma(j) > 0\}$.

We write $\Sigma_p^{\mathcal{G}}$ for the set of strategies (pure and mixed) of player $p \in P$ and $\Pi^{\mathcal{G}}$ for the set of strategy profiles in a game arena $\mathcal{G}$. There is a distinguished leader player $l \in P$.

For a strategy profile σ, a player $p \in P$, and a strategy σ' of p, we write $\sigma_{p,\sigma'}$ for the strategy profile σ' such that $\sigma'(p) = \sigma'$ and $\sigma'(p') = \sigma(p')$ for all $p' \in P \setminus \{p\}$. We are now in a position to formally define Nash and leader (or: Stackelberg) equilibria.

Definition 2.2. A strategy profile σ is a Nash equilibrium if no player would gain from unilateral deviation, i.e., for all players $p \in P$ we have $r_p(\sigma) \geq r_p(\sigma_{p,\sigma'})$ holds for all $\sigma' \in \Sigma_p^{\mathcal{G}}$.

2.1 Leader strategy profiles

In a leader strategy profile no player but the leader may have an incentive to deviate.

Definition 2.3. A strategy profile σ is a leader strategy profile if no player except the leader would gain from unilateral deviation, i.e., for all $p \in P \setminus \{l\}$ we have $r_p(\sigma) \geq r_p(\sigma_{p,\sigma'})$ for all $\sigma' \in \Sigma_p^{\mathcal{G}}$.

An optimal leader strategy profile w.r.t. the leader return is a leader equilibrium.

Definition 2.4. A leader strategy profile σ is a leader equilibrium if leader reward r_l in σ is maximal among the class of leader strategy profiles. I.e., $r_l(\sigma) \geq r_l(\sigma')$ for any other leader strategy profile σ'.

2.2 Incentive strategy profiles

Before defining incentive strategy profiles and incentive equilibrium, we first introduce concept of incentives, bribery vector, incentive profile, and bribery stable strategy profile.

Definition 2.5. An incentive or bribery amount is defined as a non-negative value $\iota \geq 0$ that the leader pays to her followers to follow a strategy profile.

Definition 2.6. We define bribery vector for a player p as $\beta^p = (\beta_1, \ldots, \beta_n)$ with non-negative incentives $\beta_j \geq 0$ for all $j = 1, \ldots, n$ corresponding to each pure strategy of player $p \in P \setminus \{l\}$, and where the leader pays to player p an incentive $\iota_p = \beta_j \geq 0$ to follow pure strategy j. We denote with β the bribery vector $\forall p \in P \setminus \{l\}$ for the strategy profile σ.

Definition 2.7. A strategy profile (σ, β) with bribery vector β is an incentive strategy profile where we denote the leader payoff by $r_l(\sigma) - \sigma \sum_{p \in P \setminus \{l\}} \beta^p$ and payoff for the player p by $r_p(\sigma) + \sigma\beta^p$ and no player p benefits from deviation.

We call a bribery vector $\beta^p = (\beta_1, \ldots, \beta_k)$ for player p a j-bribery and denote it with β_j^p if it incentivises player p for only playing pure strategy j, that is, if $\beta_i^p = 0$ for all $i \neq j$.

A strategy profile is called bribery stable that is stable under bribery or incentives. That is, for a given bribery vector and a strategy profile, no follower can improve his payoff by changing his strategy for the given leader strategy. A player p has therefore no incentive to deviate from a bribery stable strategy profile.

Definition 2.8. We define a bribery stable strategy profile as an incentive strategy profile (σ, β) that is stable in that no follower benefits from deviation i.e., for all $p \in P \setminus \{l\}$ and for all $\sigma' = (\sigma'_q)_{q \in P \setminus \{l\}}$, with $\sigma_q = \sigma'_q$ for all $q \neq p$, we have $r_p(\sigma, \beta) \geq r_p(\sigma', \beta')$ holds for all other bribery strategy profiles (σ', β').

Definition 2.9. A bribery stable strategy profile (σ, β) is an incentive equilibrium if $r_l(\sigma, \beta) \geq r_l(\sigma', \beta')$ holds for all other bribery strategy profiles (σ', β').

Definition 2.10. We next define an incentive profile as $\bar{\iota} = (\iota_p)_{p \in P \setminus \{l\}}$ where ι_p is the value that the leader pays to player $p \in P$ for following a pure strategy.

We use incentive profile in place of bribery vector in the definition of incentive strategy profile and an incentive equilibrium when players (including the leader) play pure strategies only.

We denote the notion of a strategy profile in the presence of an incentive profile $\bar{\iota}$ as a pair $(\overline{\sigma}, \bar{\iota})$, where $\overline{\sigma}$ is a strategy profile assigned by the leader, in which the leader pays incentives to her followers as is given by $\bar{\iota}$. We write $\bar{\iota}_p$ for the incentive for player $p \in P \setminus \{l\}$ in an incentive profile $\bar{\iota}$. We write $\bar{\iota}_p(\overline{\sigma})$ for the incentive to player p for the strategy profile $\overline{\sigma}$ under incentive profile $\bar{\iota}$.

In any incentive strategy profile $(\overline{\sigma}, \bar{\iota})$, no player but the leader may benefit from deviation.

Definition 2.11. For an incentive profile $\bar{\iota}$, a strategy profile $\overline{\sigma}$ is an incentive strategy profile $(\overline{\sigma}, \bar{\iota})$, if no follower can improve his overall payoff from a unilateral deviation. I.e., for all players $p \in P \setminus \{l\}$ we have that $r_p(\overline{\sigma}) + \bar{\iota}_p(\overline{\sigma}) \geq r_p(\overline{\sigma}_{p,\sigma'}) + \bar{\iota}_p(\overline{\sigma}_{p,\sigma'})$ for all $\sigma' \in \Sigma_p^{\mathcal{G}}$.

An optimal strategy profile among this class that provides maximal payoff to the leader is an *incentive equilibrium*.

Definition 2.12. For an incentive profile $\bar{\iota}$, an incentive strategy profile $(\overline{\sigma}, \bar{\iota})$ is an incentive equilibrium if the leader's total payoff for this profile is maximal among all incentive strategy profiles. I.e., for all $(\overline{\sigma}', \bar{\iota}')$ we have that $r_l(\overline{\sigma}) - \sum_{p \in P \setminus \{l\}} \bar{\iota}_p(\overline{\sigma}) \geq r_l(\overline{\sigma}') - \sum_{p \in P \setminus \{l\}} \bar{\iota}'_p(\overline{\sigma}')$.

For a given $\varepsilon > 0$ we call an incentive strategy profile $(\overline{\sigma}, \bar{\iota})$ an ε-incentive equilibrium if the leader's payoff is at most ε worse than that of any other profile. I.e., for all profiles $(\sigma', \bar{\iota}')$ we have that

$$r_l(\overline{\sigma}) - \sum_{p \in P \setminus \{l\}} \bar{\iota}_p(\overline{\sigma}) \geq r_l(\overline{\sigma}') - \sum_{p \in P \setminus \{l\}} \bar{\iota}'_p(\overline{\sigma}') - \varepsilon.$$

Note that the way incentives are computed depends upon the type of game. As an example, for mean-payoff games the incentives are computed in a similar manner as mean-payoffs are computed, i.e., they are aggregated over an infinite play. Thus, for a

mean-payoff game, incentives are extended to infinite play $\pi = \langle v_0, v_1, \ldots \rangle$ in the usual mean-payoff fashion:

$$\iota_p(\pi) \stackrel{\text{def}}{=} \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \iota_p(v_0 \ldots v_{n-1}).$$

Incentive strategy profiles vs. leader strategy profiles. We call an incentive strategy profile a *leader strategy profile* if all incentives are constant 0 functions, and a *Nash strategy profile* if, in addition, $\overline{\sigma}$ is also a Nash equilibrium. We write SP, ISP, LSP, and Nash SP for the set of strategy profiles, incentive strategy profiles, leader strategy profiles, and Nash strategy profiles. The following is a straightforward observation.

$$\textit{Nash SP} \subseteq \textit{LSP} \subseteq \textit{ISP} \subseteq \textit{SP}. \tag{2.1}$$

Chapter 3

Bi-Matrix Games

This chapter is mainly based on the results from [GS15]. In addition to these results, we discuss here the properties and existence of secure incentive equilibrium. Note that we have already introduced bi-matrix games in general in Chapter 1 (cf. Section 1.1). Here, we start with the discussion of our game settings in particular and then we give our solution approach.

We first give few motivational examples in Section 3.3 that exemplifies the main strength of the techniques we introduce in this chapter. Section 3.4 formally introduces bi-matrix games and also discusses the terminology and notations used throughout this chapter. Section 3.5 discusses incentive equilibria in bi-matrix games and the properties like tractability, purity and friendliness of incentive equilibria. We give technical details and techniques we develop to compute various types of equilibria in Section 3.6. We show that the construction of both friendly incentive equilibria and ε-optimal incentive equilibria is tractable. We present our experience with an implementation of an algorithm for friendly incentive equilibria on the randomly generated bi-matrix games in Section 3.9. We recall that as a general convention, we refer to the leader as *she* and her follower(s) as *he*.

3.1 Abstract

We consider non-zero sum bi-matrix games where one player presumes the role of a leader in the Stackelberg model and the other player is her follower. In a leader (or: Stackelberg) equilibrium, it suffices for the leader to commit to an optimal mixed strategy and to assign a pure strategy to her follower. This approach provides better solutions for the leader when compared to Nash equilibria, where the decisions are made independently and simultaneously. We show that the leader can improve her reward further when she is allowed to incentivise the strategy selection of her follower: we discuss a setting, where the leader can pay some of her own utility to her follower for assigning a particular strategy profile. We call the resulting strategy profile 'stable under bribery conditions' if the follower would not benefit from deviation. Among these strategy profiles, we call those with an optimal return for the leader 'incentive equilibria'. Clearly, leader – and

even Nash – equilibria are stable under bribery condition (for no incentive). Leader and Nash equilibria therefore cannot be superior (w.r.t. leader return) to incentive equilibria.

Like for leader equilibria, this basic model of incentive equilibria makes assumptions on the behaviour of the follower. One assumption that we take for granted is that both players play rational in that their main objective is to maximise their own payoff. The second assumption is that followers are friendly to their leader in that they choose, ex aequo, the strategy suggested by her. This implies that they implicitly take the welfare of the leader into account, at least as long as it does not affect their expected payoff. This second assumption is disputable, and one could as well assume that, ex aequo, the follower acts adversarial towards his leader. We discuss the implications of these different behavioural models. In a nutshell, we argue that the optimistic assumption leads to an obligation for the leader to return this kindness: she should choose, ex aequo, an assignment that provides the highest follower return. We call the resulting equilibria 'friendly incentive equilibria'. This obligation is at least a moral one, but it also has an economical side, as a follower who observes the opposite is not likely to keep making leader friendly choices. The pessimistic assumption leads to a situation, where the 'stable under bribery conditions' needs to be strengthened to a condition comparable to the one known from secure Nash equilibria. In many cases, no optimal incentive equilibrium for this condition exists. We therefore introduce ε-optimal incentive equilibria for this case. We show that the construction of all of these incentive equilibria is tractable.

3.2 Introduction

Stackelberg models [vS34] have been studied in detail in Oligopoly Theory [Fri77]. In a Stackelberg model, one player or firm acts as a market leader and the other is a follower. The model is a sequential move game, where the leader takes the first move and the follower moves afterwards. The main objective of both players is to maximise their own return. The Stackelberg model makes some basic assumptions, in particular that the leader knows ex-ante that the follower observes her action and that both players are rational in that they try to maximise their own return.

We consider a game settings where one player acts as the leader and assigns a strategy to her follower, which is optimal for herself. This is viewed as a Stackelberg or leadership model, where a leader is able to commit to a strategy profile first before her follower moves. We consider Stackelberg models for non-zero sum bi-matrix games that are the normal form games. In our game settings row player is the leader and the column player is her follower. In a stable strategy profile, players select their individual strategies in such a way that no player can do better by changing their strategy alone.

This requirement is justified in Nash's setting as the players have equal power. In a leader equilibrium [vS34], this is no longer the case. As the leader can communicate her move (pure or mixed) up front, it is quite natural to assume that she can also communicate a suggestion for the move of her follower. Yet, the leader cannot freely

assign strategies. Her follower will only follow her suggestion when he is happy with it in the Nash sense of not benefiting from changing his strategy. We refer to strategy profiles with this property as *leader strategy profiles*. A leader equilibrium is simply a leader strategy profile, which is optimal w.r.t. the leader return.

We argue that a leader with the power to communicate can also communicate her strategy and the response she would like to see from her follower and that – also how much – she is willing to pay for compliance. This allows her to incentivise various strategy choices of her follower by paying him a small bribery value. Incentivising the choice of her follower by paying an incentive $\iota \geq 0$ would intuitively change the payoff matrices for both players accordingly: it would decrease the leader's payoff by an amount ι, and increase the follower's payoff by the same amount.

In this setting, the leader has more power compared to Stackelberg's traditional setting: the moves there can be viewed as special cases with an incentive of $\iota = 0$. Similar to the classic case, the leader is restricted to strategies that the follower is prepared to follow, but for determining this, the incentive he would gain (when following) or lose (when deviating) is taken into account. We refer to strategy profiles (including the bribery value) that satisfy this constraint as *incentive strategy profiles*, and to the optimal choices of the leader among them as *incentive equilibria*.

We finally turn to the assumptions that we make on the behaviour of the follower. In all models we study, the main characteristics of the players is that they play *rational* in that their main goal is to maximise their own payoff. This raises the question how they select among strategies that provide the same return for themselves. It is common to assume that the follower follows the leaders suggestion as long as he does not suffer from it. That is, the follower will, ex aequo, do what the leader asks him to do. We argue that this assumption puts an obligation on the leader to be considerate towards the follower return in the same way. Thus, if the leader wants to benefit from a friendly ex aequo choice of her follower, then she should be friendly ex aequo, too. We therefore introduce *friendly incentive equilibria*. An incentive equilibrium is *friendly*, if it provides the highest *follower* return among all incentive equilibria. When always selecting friendly incentive equilibria, the leader returns this kindness of her follower by using his payoff as a tie-breaking criterion ex aequo. In the same way, we define friendly leader equilibria.

However, if we drop the assumption that the follower selects, ex aequo, the strategy suggested by the leader, then we should conservatively assume that the follower has a secondary objective to harm the leader. With this behavioural model of the follower, the leader can only put forward strategy profiles, where every deviation of the follower must either lead to a *strict* decrease of the follower return, or does not decrease the leader return. A similar property has been studied for Nash equilibria as secure Nash equilibria [CHJ05a]; we therefore use the term secure incentive equilibria.

Incentive strategy profiles that meet these requirements are called *secure* incentive strategy profiles. A secure incentive strategy profile with maximal payoff for the leader is called a secure incentive equilibrium. Secure leader strategy profiles and secure leader

equilibria can be defined accordingly. We will see that secure incentive equilibria do not always exist. Moreover, when they exist, then there is at least one among them that has a bribery value 0, and is therefore a secure leader equilibrium, too. To cover the general case where secure incentive equilibria may not exist, we show that incentive equilibria are a very stable concept: we show that ε-optimal secure incentive strategy profiles always exist. In fact, they can be derived from any incentive equilibrium by raising the incentive by an ε amount.

We therefore portray two different approaches of a leader in a Stackelberg equilibrium – in the first approach, the follower acts friendly towards his leader and plays the strategy as assigned to him by the leader. Here, the leader is obliged to select a strategy profile that would, ex aequo, maximise the follower return. The second approach considers an adversarial follower. The leader then has to select strategy profiles that are secure, although they might not be optimal. (However, the relative loss of the leader is arbitrarily small, such that we skip over this detail in the remaining introductory part.)

In a Stackelberg game where one player is in a position to commit before the other player take his move, an extensive form representation of the game is a common assumption. An extensive form of the game is a finite tree representation where each node of the tree is labelled with a player who will choose an action at that node. Each action corresponds to an edge going deep in the tree and payoff vector is given at terminal nodes that depict the payoff for each player. At every node players have perfect information about their earlier moves and this representation therefore makes temporal aspect of the game explicit. For example, an extensive form representation of popular game prisoner's dilemma is shown in the Figure 3.1.

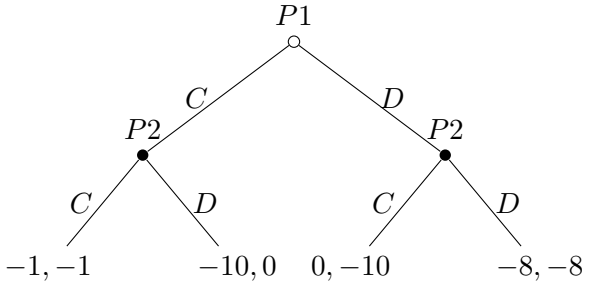

FIGURE 3.1: Prisoner's Dilemma in Extensive-form.

In this figure, Prisoner I at the root node moves first and chooses an action that corresponds to a branch. Prisoner II would then choose his action depending upon the move Prisoner I has taken. Payoff for different strategy profiles are shown at terminal nodes.

If the settings were to be modelled as an extensive-form game, backward induction can be applied on the game tree to compute an equilibrium (subgame perfect). A subgame perfect equilibrium is a refinement of Nash equilibrium with a property that equilibrium computed in every subgame is also an equilibrium. As in an incentive equilibrium the leader needs to communicate her strategy and also the bribery to be

given to her follower, this information needs to be communicated along the tree branch where a decision has to be made.

Assuming that Prisoner I is the leader in the above example (Figure 3.1), she needs to communicate an incentive she would give to Prisoner II (the follower) while taking a move. To compute incentive equilibrium using backward induction, one would start at the bottom of the game tree and compute optimal action for player at each decision node. One would then move a layer up until the root node is reached and an optimal strategy profile is computed. The strategy profile thus obtained is subgame perfect relative to the leader as the leader is allowed to benefit from deviation in our setting.

Although it is natural to assume a tree representation of a leadership game, the size of the tree could become large for an increasing number of strategies. Moreover, if a player is in a position to commit to a mixed strategy (leader in our case) this size could be even larger as every pure strategy in the support of mixed strategy would correspond to a (pure strategy) different branch in the tree. This size would increase further for an increasing number of strategies in the original game. The set of strategy profiles in a Nash equilibrium in an extensive form game also contain unreasonable strategy profiles as it allows for strategies that constitutes non-credible threat that the players would not carry out in real. Therefore to stay simple and convenient, we use simple and original representation for the two-player strategic games i.e., bi-matrix form.

3.3 Motivational examples

As a first motivational example for a friendly incentive equilibrium, we consider a travel agent, who sells flight tickets to a customer. The travel agent receives a concession fee from the companies on every deal. These fees, however, differ for different airlines. Customers who wish to travel may value flights differently. For example, they might have preference over flight timings, connections in between the flights, choice of airport, etc. E.g., they might prefer arriving in Heathrow over arriving in Stansted, or prefer leaving at 10am over leaving at 5am. They value their benefit from buying the various connections on offer differently, e.g., as follows:

Airline	agent-value	customer-value
Air India:	£100	£170
German Wings:	£50	£100
Lufthansa:	£120	£80
Ryan Air:	£5	£200
US Airways:	£100	£150

The agent-value reflects the concession fee the agent would get from the respective flight. The customer-value reflects the value the customer considers an offer to have (the value of the flight to him minus the cost of the ticket).

When the travel agent learns the priorities and valuations of the customer, she can make an offer to suit their needs. The agent cannot increase the offer by the company,

but she can discount them on her own expense. When the agent does not discount any offer, the customer would choose the Ryan Air flight, but the travel agent can simply offer the customer an incentive in form of a £30 discount on the Air India flight, thus increasing her gain from £5 to £70. When the leader announces her action, she effectively simplifies the bi-matrix game to a vector, like the one from this example. Here, by offering a discount of £30, the leader can incur huge profit on her return, as the customer now values both Ryan Airways and Air India equally.

Customer relationship models of this type usually lead to only the customer making choices. However, this still leaves some scope for deviation, if the customer receives an equal value on more than one option. By paying an $\varepsilon > 0$ extra amount, the leader can guarantee that the follower will not deviate from the assigned strategy and the strategy profile then becomes an ε-optimal secure strategy profile. Thus, the travel agent can offer a discount of $£30 + \varepsilon$ (£30.01)to the customer on the Air India flight.

The example can be viewed as a simple bi-vector where one player (leader in our case) has only one choice and therefore she has no direct influence on the strategy selection. As opposed to leader or Nash equilibria, the leader can benefit here from announcing incentives on the various options that the other player has. Effectively, this is the only thing that the leader can do in this case – she can announce incentives in order to encourage the selection of a strategy in her favour by her follower. We come back to this later in this chapter (Section 3.9) and discuss how it gives a better social return and the leader can also improve her own payoff. Interestingly, the payoff for the follower remains unaffected in these cases.

The example therefore translates to simple scenarios where the leader can only select the incentives to be given to her follower, but has only one option. This simplified setting is included for two reasons: it shows that the concept is useful in even simpler scenarios where incentivising is the only way the leader can influence the game and it separates the concerns and shows the influence of incentives in isolation.

Arms race game. A class of games to derive benefit from friendly incentive equilibrium is the class of 'Arms race' games. Consider two countries, say the more powerful country and the less powerful country, who, because of a mutual threat, have to spend a considerable amount of their resources on the production of arms and their military. The countries can save a considerable amount if they both agree on not entering in an arms race. To achieve this, the more powerful country may pledge not to enter an arms race, and at the same time offer the less powerful country (the follower) a trade advantage, or promise support in obtaining public events like Olympic games or any other major event in return for the less powerful country also not to enter into an arms-race. If the pledge is believed and the incentive is sufficiently high such that the follower would have no incentive to enter the arms-race, then the dilemma is overcome. This also exemplifies that the incentive is not the main obstacle, but the trust in the pledge not to enter into the arms-race.

The examples discussed above shows how a friendly incentive equilibrium gives better results than the leader equilibrium or Nash equilibrium. When compared to leader or Nash equilibria, a friendly incentive equilibrium would always give better results for the leader in these games. This also reflects why it is beneficial for the leader to behave friendly to her follower ex aequo when the follower is also friendly.

Another important class of games to derive benefit from incentive equilibria is Prisoner's dilemma [Tuc50]. In games of this class, players are somehow hesitant to co-operate even if it lies in the best of their interests. The leader's ability to incentivise her follower then helps the players to arrive at an optimal solution. We discuss this game in detail in the Section 3.5.

3.3.1 Related Work

Leader equilibria, introduced by von Stackelberg [vS34] and therefore sometimes referred to as Stackelberg equilibria, have been studied in depth in Oligopoly theory [Fri77]. The main contribution of our work is conceptual, and the closest relation of our equilibrium concept is to Stackelberg leader equilibria [CS06, vSZ10]. Ehtamo and Hamalainen in [EH86] considered the construction of optimal incentive strategies in two-player dynamic game problems that are described by integral convex cost criteria. They considered the problem of finding an incentive strategy for the leader by looking at the rational response from the follower, such that, in an optimal strategy, the cost functional of the leader could be minimised. They further considered [EH89] the analytical methods for constructing memory incentive strategies for continuous time decision problem. The strategies they study are time-consistent, which means the continuation of equilibrium solution remains an equilibrium.

Stark, in [Sta89], studied how the introduction of altruism into non co-operative game settings can lead to an improved quality for both the agents. Stark [Sta85] also discussed special altruism – 'a mutual altruism' and their role in various contexts. A realistic altruistic behaviour of players is also considered in fairness equilibrium [Rab93] where a player behaves nice if the other player is also nice with him and a player acts hostile if the other player is also hostile. Fairness equilibria is however not necessarily a superset or subset of Nash equilibria. This is altruism considered at different level: the same people who are altruistic to other altruistic people are motivated to hurt those who hurt them. A player's payoff therefore does not depend only on his choice of actions but also on his beliefs about the other player's actions. The payoff profile therefore depends largely on the player's motives.

A game model where players make binding offers of side payments is studied in [MOJ05]. The game is played in stages where in a first stage players engage in side contracting and payoff functions are altered accordingly. The altered game is then played in the second stage. Eventually, the game is played to be endogenous. The game is also not a sequential game and both players are in a position to offer side payments.

Von Stengel and Zamir [vSZ10] studied a commitment model in a leadership game with mixed extensions of bi-matrix games. They show that the possibility to commit to a strategy profile in bi-matrix games is always beneficial for the committing player. Von Stengel and Zamir gave further results in this regard in [vSZ04]. They show that the set of leader's payoff in a leadership game lies in an interval which is as good as that player's Nash and correlated equilibrium payoff in a leadership game. They discussed the importance of commitment as a means of coordination for considering correlated equilibria.

Conitzer and Sandholm [CS06] gave a first insight into the computation of Stackelberg strategies in normal-form games. [ASA10] considered a mechanism designer modelled as a player in the game who has the opportunity to modify the game. In their setting, the players' utility and social welfare is seen as counter intuitive. E.g., social welfare may arbitrarily come worse and they focus completely on pure strategies, whereas, our solution approach tries to increase the payoff of both players, which often leads to a good social outcome and the equilibrium we study is mixed only for the leader (while the strategies assigned to the follower are pure).

Another related work is [MT04], where an external party, who has no control over the rules of the game, can influence the outcome of the game by committing to non-negative monetary transfers for the different strategy profiles that may be selected by the agents in a multi-agent interaction.

Secure Nash equilibria have been studied for multi-player graph based games with quantitative objectives in [CHJ06] and [DPFK$^+$14]. These results, however, pertain to secure Nash equilibrium only. Chatterjee et al. [CHJ06] have studied secure Nash equilibria in two-player non zero-sum games, where they considered lexicographic objectives of the two players in order to make an equilibrium secure – players first follow their objective of maximising their payoff and then they have a secondary objective of minimising the other player's payoff. If the two players select their strategy profiles in this order, they would form a unique secure Nash equilibrium – i.e., a strategy profile, which is in equilibrium and is also secure. De Pril et al. [DPFK$^+$14] establishes the existence of secure equilibria in multi-player perfect information turn-based games for games with probabilistic transitions and for games with deterministic transitions.

When comparing Nash with leader or incentive equilibria, the latter are computationally cheap solutions as the construction of leader equilibrium is known to be tractable [CS06]. Similarly, we show in this chapter that computing incentive equilibria is also tractable. For Nash equilibria, theoretical computer science has contributed much towards its complexity and how easy it is to find one. Especially, how soon can we expect players to really reach an equilibrium? Roughgarden [Rou10] has considered a concrete 'dynamic' model to learn these behavioural properties, especially, how quickly we can expect players in multi-player games to arrive at an equilibrium. They use a straight forward procedure known as best-response dynamics [Rou10].

In any finite potential game [MS96], starting from an arbitrary initial strategy, best-response dynamics finally converges to a pure Nash equilibrium (PNE), if there exists one. It cycles in game without any PNE. However, in the worst-case, it might take an exponential number of iterations. This is a method of recursively searching for a PNE, and it halts only at a point when it has found one. Best-response dynamics are best applicable in a situation where a large number of agents collectively interact to produce a solution to a problem, with every agent trying to pull the solution in a direction that is favourable to him. Here, each agent would try to optimise their individual objective function.

3.4 Definitions

We define a bi-matrix game as $\mathcal{G}(A, B)$, where A and B are the real valued $m \times n$ payoff matrices for the leader and the follower, respectively. In our settings, leader is the row player and the follower is the column player. We refer to the number of rows by m and to the number of columns by n. We also refer to the entry in row i and column j of A and B by a_{ij} and b_{ij}, respectively. Since we have only two players, we represent leader's strategy with σ and follower's strategy with δ and represent a *strategy profile* as a pair of strategies $\langle \sigma, \delta \rangle$. A (mixed) strategy of the leader is a probability vector $\sigma = (p_1, \ldots, p_m)$, i.e., $\sum_{i=1}^{m} p_i = 1$ (the sum of the weights is 1) and $p_i \geq 0$ for $1 \leq i \leq m$. Likewise, a (mixed) strategy of the follower is a probability vector $\delta^T = (q_1, \ldots, q_n)$, i.e., $\sum_{j=1}^{n} q_j = 1$ and $q_j \geq 0$ for $1 \leq j \leq n$. Where convenient, we read δ^T as a function and refer to the j^{th} column of δ by $\delta^T(j)$. For a given follower strategy $\delta^T = (q_1, \ldots, q_n)$, we define its support as the positions with non-zero probability, and denote it as $\mathsf{support}(\delta^T) = \{j \leq n \mid \delta^T(j) > 0\}$.

Note that we have formally introduced equilibrium concepts in Chapter 2. We only define here concepts that are not introduced before. As there is one follower, we denote with β the bribery vector for follower and represent it as $\beta^T = (\beta_1, \ldots, \beta_n)$, with non-negative entries $\beta_j \geq 0$ for all $j = 1, \ldots, n$, corresponding to each decision 1 through n of the follower, where the leader pays her follower the bribery β_j when he plays j. For an incentive strategy profile $(\langle \sigma, \delta \rangle, \beta)$, we denote the leader payoff by $\mathsf{lpayoff}(A; \langle \sigma, \delta \rangle, \beta) = \sigma A \delta - \ \delta^T \beta$ and the follower payoff by $\mathsf{fpayoff}(B; \langle \sigma, \delta \rangle, \beta) = \sigma B \delta + \delta^T \beta$.

We call strategies with singleton support *pure*. We are particularly interested in pure follower strategies, and abbreviate the strategy with $\delta^T(j) = 1$ by j, or by $\overrightarrow{j}$ to emphasise that j refers to a strategy. We call a bribery vector $\beta^T = (\beta_1, \ldots, \beta_n)$ a j-bribery if it incentivises follower for only playing j, that is, if $\beta_i = 0$ for all $i \neq j$. An incentive equilibrium $(\langle \sigma, \delta \rangle, \beta)$ is called simple, if δ is pure. For simple incentive equilibria, we would refer to bribery vector β with j-bribery β. We refer to the value of this j-bribery by $\iota = \beta_j$ as the incentive that the leader gives to her follower to solicit him to play j.

An incentive equilibrium is *friendly*, if, among all incentive equilibria, the follower return is maximal. A friendly incentive equilibrium is thus defined as follows.

Definition 3.1. An incentive equilibrium $(\langle\sigma, \delta\rangle, \beta)$ is called *friendly*, if $\mathsf{fpayoff}(B; \langle\sigma, \delta\rangle, \beta) \geq \mathsf{fpayoff}(B; \langle\sigma', \delta'\rangle, \beta')$ holds for all incentive equilibria $(\langle\sigma', \delta'\rangle, \beta')$.

Note that in a friendly incentive equilibrium we are not bothered about the follower's deviations (as the outcome is also in the favour of follower). This is unlike secure incentive equilibrium where for every follower deviation, either the leader must not lose or the follower loses strictly. We now give these requirements put on the incentive strategy profiles that should be met for an incentive strategy profile to be secure.

Definition 3.2. An incentive strategy profile $(\langle\sigma, \delta\rangle, \beta)$ is called a secure incentive strategy profile, if, for every follower deviation δ',

1. the follower loses strictly, i.e., $\mathsf{fpayoff}(B; \langle\sigma, \delta\rangle, \beta) > \mathsf{fpayoff}(B; \langle\sigma, \delta'\rangle, \beta)$ or,
2. the leader does not lose, i.e., $\mathsf{lpayoff}(A; \langle\sigma, \delta'\rangle, \beta) \geq \mathsf{lpayoff}(A; \langle\sigma, \delta\rangle, \beta)$.

As discussed in the Section 3.2, secure incentive strategy profile may not always exist. We establish this later in the section 3.8.3 (cf. Theorem 3.35). Therefore, to cover the broader case when secure incentive strategy profile do not exist, we define an ε-optimal incentive strategy profile for an $\varepsilon > 0$ as follows.

Definition 3.3. An incentive strategy profile $(\langle\sigma, \delta\rangle, \beta)$ is, for an $\varepsilon > 0$, called an ε-optimal incentive strategy profile if the leader payoff in $(\langle\sigma, \delta\rangle, \beta)$ is at most ε worse than in any other incentive strategy profile. Thus, for any other simple incentive strategy profile $(\langle\sigma', \delta'\rangle, \beta')$, it holds that $\mathsf{lpayoff}(A; \langle\sigma, \delta\rangle, \beta) \geq \mathsf{lpayoff}(A; \langle\sigma', \delta'\rangle, \beta') - \varepsilon$.

3.5 Incentive equilibria in bi-matrix games

We discuss here incentive equilibria in bi-matrix games. For an example, we refer to a bi-matrix game shown in the Table 3.1.

		Player II	
		I	II
Player I	I	$5, 0$	$0, 1$
	II	$0, 0$	$-1, 0$

TABLE 3.1: Equilibria in a bi-matrix game.

In this example, the only Nash equilibrium is the strategy profile (I, II). This strategy profile is also the only leader equilibrium. The leader payoff and the follower payoff in this strategy profile are 0 and 1 respectively. However, if we allow the leader to incentivise her follower for following a particular strategy profile, then she can incentivise her follower to play the pure strategy I by paying him a bribery value of 1. The strategy profile (I, I) with bribery value 1 is then an incentive equilibrium. The payoffs for the leader and the follower for this equilibrium are 4 and 1, respectively.

Like in leader equilibria, the leader can select a strategy profile where she benefits from deviation. We will discuss such a situation on the example of the Prisoner's dilemma [Tuc50]. As the entities who interact strategically are often called players, we use the term 'players' here instead of 'prisoners'. The game has the famous antinomy that both players do better if they both co-operate (C) with each other, while, both of them have the dominating strategy to defect (D). Consequently, (D, D) is the only Nash equilibrium in this game. We refer to the prisoner's dilemma payoff matrix from Table 3.2 that is a normal form representation of game from Figure 3.1.

The only Nash equilibrium here is the strategy profile (D, D) with a joint return of -16. Another observation is that leader equilibria are not powerful enough to overcome this antinomy. This is because the Nash or leader equilibrium in this game is largely based on a dominant strategy ('defect' in this example), that would remain dominant and always overpowers the other strategy. The dilemma here is that players get much better payoffs if both of them choose to co-operate. Mutual co-operation, however, is prevented by the presence of the dominant strategy to defect. Thus, the Nash or leader equilibrium here does not allow us to reach the social optimum. We observe that incentive equilibria help to overcome this shortcoming. The leader can assign a strategy profile (C, C) that gives overall utility of -2. This is because, with incentive equilibria, the leader can make use of her power to *incentivise*. She can bribe Player II into co-operation by offering a bribery value 1.

As a result, co-operating becomes an optimal choice for her follower. In this example (C, C) with bribery value 1 is the only incentive equilibrium. Note that the friendly incentive equilibrium is also *Pareto optimal* while the Nash equilibrium is not. A strategy profile is Pareto optimal when no player in the game can do better off without making atleast one player worse off. The strategy profile (C, C) is therefore Pareto optimal while the Nash equilibrium (D, D) does not meet this criterion here. Interestingly, the follower benefits more than the leader from the additional power of the leader to incentivise follower behaviour: after bribery, the leader return is -2, while the follower return is 0 in this symmetric game.

		Player II	
		C	D
Player I	C	$-1, -1$	$-10, 0$
	D	$0, -10$	$-8, -8$

TABLE 3.2: Prisoners' Dilemma.

However, this is not always the case. Table 3.3, for example, shows a bi-matrix game where the follower return is higher for leader and Nash equilibria: the strategy profile (II, II) is an incentive equilibrium with bribery value 1 and the strategy profile (I, I) is a leader equilibrium. The follower return in this leader equilibrium is 2, while, it is 1 in the only incentive equilibrium.

		Player II	
		I	II
Player I	I	$1, 2$	$0, 0$
	II	$-10, 1$	$10, 0$

TABLE 3.3: An example where follower does not benefit from incentive eqilibrium.

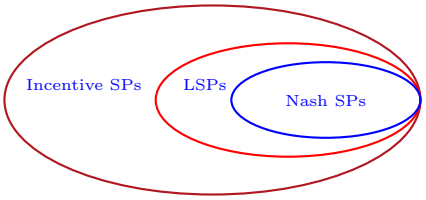

FIGURE 3.2: Incentive strategy profiles $\supseteq$ Leader strategy profiles $\supseteq$ Nash strategy profiles.

From the leader's point of view, it is easy to see that leader equilibria cannot be superior to incentive equilibria, and Nash equilibria cannot be superior to leader equilibria, not even if the leader can choose a Nash equilibrium. The reason is that the set of incentive strategy profiles includes the set of leader strategy profiles (as these are the incentive strategy profiles with bribery value 0), which in turn includes the pool for Nash equilibria (as these are the leader strategy profiles for which the leader has no incentive to deviate). The incentive strategy profiles can thus be selected from a wider base of choices, as shown in the Figure 3.2. The figure shows that incentive strategy profiles have less constraints than leader strategy profiles that in turn have less constraints than Nash strategy profiles. The optimum over smaller sets can, naturally, not be superior over the optimum of larger sets.

We now discuss the assumptions for the follower behaviour. Our first assumption is that of a friendly follower – a follower who follows the strategy profile assigned by the leader as it is. We argue that for a friendly follower, the leader has to be friendly ex aequo too. This can be exemplified by the bi-matrix game from Table 3.4.

		Follower	
		I	II
Leader	I	$1, 0$	$0, 0$
	II	$1, 1$	$1, 1$

TABLE 3.4: Leader behaves friendly when her follower is also friendly.

Here, a leader (and an incentive) equilibrium is (I, I). The technical definition seems fine: the follower has no incentive to deviate, as he has nothing to gain from playing II instead. However, it still seems pretty unlikely that this behaviour reflects a true behaviour in any game played by human players. It would assume that the pointless harm the leader inflicts on her follower (pointless in the sense that it does not incur an advantage for the leader) by playing I instead of II would not be met by a retaliation, which in turn would cost the follower nothing. In fact, this behaviour is not rational at all, as it invites non-consideration. Thus, the leader has to make follower friendly choices for a friendly follower.

Our next assumption is an adversarial follower who has a secondary objective to harm the leader. In the example from Table 3.4, when the leader plays I, the follower would play II: his own payoff is not affected by the choice, but his secondary objective is to minimise the leader return, which leads to playing II. The leader is now not obliged to make follower friendly choices and would assign a secure incentive strategy profile. An interesting observation is that it is in fact the loss of an unfriendly follower to act adversarial towards the leader.

As an example, we now refer to the bi-matrix game from Table 3.5. It shows a variant of the battle-of-sexes game where the follower (Player 2) has now three choices whereas the leader (Player 1) has only two. In a friendly incentive equilibrium, the leader would play the strategy I and assign the strategy III to her follower. The leader would pay an incentive of 1 for this and the return of both leader and follower for the friendly incentive equilibrium (I, III) is therefore 3. However, the strategy profile is not secure and therefore the leader cannot assign this strategy to an unfriendly follower. For an adversarial follower, the leader would in fact assign a strategy profile (II, II) that is also secure. It provides same return to the leader while the follower return now drops to 1. The friendly incentive equilibria are therefore in the interest of both the leader and her follower. On the other hand, secure incentive equilibria lies in the disinterest of the follower, costing nothing to the leader.

		Player 2		
		I	II	III
Player 1	I	$1, 3$	$0, 0$	$4, 2$
	II	$0, 0$	$3, 1$	$0, 0$

TABLE 3.5: An unfriendly follower suffers in a secure equilibria.

For the general case when secure incentive strategy profiles do not exist, we show that ε-optimal secure incentive strategy profiles exist. As an example, we again refer to the bi-matrix game from the prisoner's dilemma (Table 3.2). The only incentive equilibrium, where the leader co-operates and incentivises her follower to co-operate, too, by paying him a bribery value 1, is not secure: the follower could defect without affecting his payoff, while reducing the leader payoff from -2 to -10. The strategy profile (I, I) is an incentive equilibrium with bribery value 1, but this is not a secure strategy profile. However, the leader has an ε-optimal secure incentive strategy profile by suggesting (I, I) and offering a bribery value of $1 + \varepsilon$.

We now establish some of the nice properties of incentive equilibria in general to exemplify their simplicity, tractability and friendliness. We first discuss how much improvement one can obtain on the incentive equilibria as compared to the Nash or leader equilibria. If we norm the entries of the bi-matrix to be in $[0, 1]$, the improvement that the leader (and the follower) can obtain through incentive equilibria is ε close to 1 compared to leader (and Nash) equilibria. For this, we refer to the Table 3.6, a variant of the prisoner's dilemma with payoffs between 0 and 1. The social return in the incentive equilibrium is $2 - 2\varepsilon$, while in leader and Nash equilibria the social return is 2ε. Note

that, for any $\delta > 0$, we can choose an ε, e.g., $\varepsilon = \min\{0.1, \delta/3\}$, to obtain an improvement greater than $1 - \delta$, for the leader's return and the follower's return at the same time, using only values in $[0, 1]$ for the payoffs.

Prisoner I	Prisoner II: C	Prisoner II: D
C	$1-\varepsilon, 1-\varepsilon$	$0, 1$
D	$1, 0$	ε, ε

TABLE 3.6: A variant of the prisoner's dilemma.

Player 1	Player 2: I	Player 2: II
I	$1, 2$	$0, 0$
II	$0, 0$	$2, 1$

TABLE 3.7: A Battle-of-Sexes game.

Another well studied class of games are the battle of sexes games. We do not expand on these games as leader equilibria (and even Nash equilibria) are sufficient to obtain any optimal solution in such games. The incentive would therefore play no role in them. Note that the example Battle-of-Sexes game from Table 3.7 has only one leader/incentive equilibrium, but three Nash equilibria: the pure strategies where both players play I or both play II and a third mixed strategy equilibrium where Player 1 plays I with probability $\frac{1}{3}$, and Player 2 plays I with probability $\frac{2}{3}$. The outcome of these strategy profiles is $(1, 2)$, $(2, 1)$, and $(\frac{2}{3}, \frac{2}{3})$, respectively. Note that, even when one restricts the focus to those strategy profiles only, it needs to be negotiated which of them is taken.

The complexity of this negotiation is outside of the complexity to determine Nash equilibria in the first place, but it is yet another level of complexity that is reduced by using incentive (or leader) equilibria. The only leader or incentive equilibrium (with zero incentive) here is the strategy profile (II, II), which is a secure strategy profile, too.

In a leader resp. incentive equilibrium, it suffices for the leader to commit to a mixed strategy profile while the follower plays pure. A related equilibrium approach is correlated equilibrium [Aum74] that is another relaxed form of Nash equilibrium and does not require explicit randomisation on any of the player's part. By the use of correlated strategies one can achieve a payoff (for all players) that is not less than their Nash payoff. The correlated strategies [Aum74] are selected by random events rather than a mixed strategy. A randomised strategy in correlated equilibrium is therefore viewed as a random variable with values in the pure strategy space, rather than as a distribution over pure strategy space.

In correlated equilibrium a probability distribution is drawn (by a random event or a mediator) on the strategies of players. For a bi-matrix game, the probability distribution is given by a $m \times n$ matrix p that contains only non-negative entries of the type p_{ij} such that a strategy pair (i, j) is selected with probability p_{ij} and $\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1$. The distribution p_{ij} is a correlation strategy if the payoff of both players on playing the recommended strategy is no worse than playing any other strategy. Thus, a correlation

strategy has to satisfy the constraint that $\sum_{j=1}^{n} a_{ij}p_{ij} \geq \sum_{j=1}^{n} a_{kj}p_{ij}$ and $\sum_{i=1}^{m} b_{ij}p_{ij} \geq \sum_{i=1}^{m} b_{il}p_{ij}$ for each $1 \leq k \leq m$ and $1 \leq l \leq n$. These inequalities result in linear programmes and an objective function to maximise the social welfare is added. These linear programmes need to be solved to compute correlated equilibrium.

The computational difference with incentive equilibrium is that while in an incentive equilibrium the leader strategy is mixed, the strategies are randomised for both players in a correlated equilibrium. The objective in an incentive equilibrium is to maximise the leader's return while in a correlated equilibrium it is to maximise the social welfare.

Although a payoff vector (for both players) that is not achievable with the Nash equilibrium is achievable with correlated equilibrium, note that the payoff for the leader in correlated equilibrium cannot be higher than her payoff in an incentive equilibrium. For e.g., if we look at the Table 3.7, a payoff vector $(\frac{3}{2}, \frac{3}{2})$ is achievable with correlated strategies. As a random event, a coin is tossed, if it is heads, both players play strategy I and, if it is tails, they both play the strategy II. This allows them to secure a payoff of $(\frac{3}{2}, \frac{3}{2})$ that is not achievable with mixed Nash strategies. Note that the strategy profile is also an equilibrium as no player benefits from unilateral deviation. However, an incentive or leader equilibrium provides a higher payoff of 2 to the leader.

Use of correlated equilibrium in a leadership game (with two players) and where one player (the *leader*) is in a position to commit to a strategy profile is studied in [vSZ04]. They show that the lowest leader payoff in a (subgame perfect) leader equilibrium is atleast as large as any correlated payoff as any correlated payoff is as large as Nash payoff. The highest leader payoff in leader equilibrium is greater than or equal to highest payoff of the row player in correlated equilibrium ([vSZ04]). These observations together with our observation that leader payoff in an (friendly) incentive equilibrium can only be greater or equal to her payoff in a leader equilibrium implies that incentive equilibrium can only provide greater or equal payoff than any correlated payoff (for the leader).

Friendly incentive equilibria. Intuitively, it is clear from Figure 3.2 that incentive equilibria improve over leader or Nash equilibria w.r.t. the payoff of the leader. The reason is simple. The set of incentive strategy profiles form a broader set over the set of leader or Nash strategy profiles. The optimum selected over a larger set cannot be smaller than the optimum selected over smaller sets. Therefore, in any friendly incentive equilibrium, the leader can choose from strategies that only satisfy the side-constraint that the follower cannot improve over it by unilateral deviation. While selecting a strategy profile, the leader would consider to maximise follower's payoff as well. For incentive equilibria, the leader can use incentives, whereas leader equilibria would optimise only over strategies without incentive (or: with zero incentives). Thus, incentive equilibria are again an optimum over a larger base than leader equilibria. This optimisation may further leave a plateau of jointly optimal strategies, strategies with the same optimal payoff (after bribery) for the leader. We have argued in the Section 3.2 that and why

the leader can – and should – move a step ahead and choose an equilibrium that is optimal for her follower among these otherwise equivalent solutions. This is another advantage of a clear outcome for the leader – the option to choose ex-aequo an incentive equilibrium, which is good for the follower. Thus, a friendly incentive equilibrium is optimal for the leader, and assigns, among this class of equilibria, the highest payoff to the follower. Friendly incentive equilibria are therefore in favour of the follower as well. When the leader assigns strategies to herself and the other player, her primary objective is to maximise her own benefit and her secondary objective is that the follower receives a high gain, too. Thus, they refer to a situation where both, the leader and her follower, benefit, increasing the social quality of the result.

We give an algorithm for the computation of friendly incentive equilibria in the Section 3.6. The algorithm first constructs a set of constraint system, one for each pure follower strategy that can occur in an incentive equilibrium. The constraint system additionally requires the leader payoff to be optimal and maximises the follower payoff. As output it gives an optimal strategy profile, the bribery value, and the leader's and the follower's payoff.

We empirically evaluate the technique on randomly generated bi-matrix games. We considered 100,000 data-sets with continuous payoff values and integer payoff values for the evaluation of friendly incentive equilibria.

Tractability and purity of incentive equilibria. Another appealing property of general and friendly incentive equilibria is their simplicity: it suffices for the follower to consider pure strategies, or, similarly, it suffices for the leader to assign pure strategies to her follower. Another strong argument in favour of general and friendly incentive equilibria is that they are *tractable*, whereas the complexity of finding mixed strategy Nash equilibria is known to be PPAD-complete [DGP09].

3.6 Incentive Equilibria

We start with the discussion of friendly incentive equilibria in this section, followed by the discussion of ε-optimal incentive equilibria in the next section. We first show that ordinary and friendly incentive equilibria always exists. Moreover, there is always a *simple* friendly incentive equilibrium.

3.6.1 Existence of bribery stable strategy profiles

The existence of bribery stable strategy profiles is implied by the existence of Nash equilibria [Nas50, LH64], as Nash equilibria are special cases of bribery stable strategy profiles with the zero bribery vector.

Theorem 3.4 (See [Nas50, LH64])**.** *Every bi-matrix game $\mathcal{G}(A, B)$ has a Nash equilibrium.*

Corollary 3.5. *Every bi-matrix game $\mathcal{G}(A, B)$ has a bribery stable strategy profile.*

3.6.2 Optimality of simple bribery stable strategy profiles

Different to Nash equilibria, there are always simple incentive equilibria. In order to show this, we first show that there is always a simple bribery stable strategy profile. This is because, for a given bribery stable strategy profile $(\langle\sigma,\delta\rangle,\beta)$ and a j in the support of δ^T, $(\langle\sigma,j\rangle,\beta)$ is an incentive equilibrium, too.

Theorem 3.6. *For every bi-matrix game $\mathcal{G}(A,B)$ and every bribery stable strategy profile $(\langle\sigma,\delta\rangle,\beta)$ and $\mathsf{lpayoff}(A;\langle\sigma,\delta\rangle,\beta)=v$ for the leader, there is always a simple bribery stable strategy profile with $\mathsf{lpayoff}(A;\langle\sigma,j\rangle,\beta)\geq v$ for the leader.*

Proof. Let $S=\mathsf{support}(\delta^T)$ and let $s=\mathsf{fpayoff}(B;\langle\sigma,\delta\rangle,\beta)$ be the payoff for the follower in this simple bribery stable strategy profile. We first argue that, for all $j\in S$, $(\langle\sigma,j\rangle,\beta)$ is a simple bribery stable strategy profile with the same follower payoff as $(\langle\sigma,\delta\rangle,\beta)$. First, the follower payoff cannot be higher, as $(\langle\sigma,j\rangle,\beta)$ would otherwise not be a simple bribery stable strategy profile (the follower could improve his payoff by changing his strategy to j).

Assuming for contradiction that there is an $j\in S$ with $\mathsf{fpayoff}(B;\langle\sigma,j\rangle,\beta)<s$ implies, together with the previous observation that $\mathsf{fpayoff}(B;\langle\sigma,j\rangle,\beta)\leq s$ holds for all $j\in S$, that $\sum_{j\in S}\delta^T(j)\cdot\mathsf{fpayoff}(B;\langle\sigma,j\rangle,\beta)<s$, which contradicts $s=\mathsf{fpayoff}(B;\langle\sigma,\delta\rangle,\beta)$. Taking into account that the leader payoff $v=\sum_{j\in S}\delta^T(j)\cdot\mathsf{lpayoff}(A;\langle\sigma,j\rangle,\beta)$ is an affine combination of the leader payoffs for these simple bribery stable strategy profiles, there is some $j\in S$ with $\mathsf{lpayoff}(A;\langle\sigma,j\rangle,\beta)\geq\ v$. □

Note that the above theorem reflects the 'best response' [NRTV07] condition in a bi-matrix game. The 'best response' condition states that all pure strategies in the support of a mixed strategy must get maximal and also equal payoff. For a mixed strategy Nash equilibrium, any pure strategy in the support of a mixed strategy is a 'best response' in response to the strategy chosen by the other player. The above theorem asserts this condition for an incentive equilibrium. That is, for a mixed strategy for the leader, any pure strategy of the follower that is in support of the follower's mixed strategy, is a best response. It therefore suffices for the leader to assign a pure strategy to her subject.

3.6.3 Description of simple bribery stable strategy profiles

Theorem 3.6 allows us to seek incentive equilibria only among simple bribery stable strategy profiles. Intuitively, it implies that it suffices for the leader to assign pure strategy to her follower while she herself can play a mixed strategy. Note that this is in contrast to general Nash equilibria, cf. the rock-paper-scissors game. Simple bribery stable strategy profiles are defined by a set of linear inequations.

Theorem 3.7. *For a bi-matrix game $\mathcal{G}(A,B)$, $(\langle\sigma,j\rangle,\beta)$ is a simple bribery stable strategy profile if, and only if, $\sigma B\overrightarrow{j}+\beta\overrightarrow{j}\geq\sigma B\overrightarrow{i}+\beta\overrightarrow{i}$ holds for all pure strategies i=1,…,n of the follower.*

Proof. If $(\langle\sigma, j\rangle, \beta)$ is a simple bribery stable strategy profile, then in particular changing the strategy to a different pure strategy i cannot be beneficial for the follower. Consequently, it holds for all $i = 1, \ldots, n$ that $\sigma B \overrightarrow{j} + \beta \overrightarrow{j} \geq \sigma B \overrightarrow{i} + \beta \overrightarrow{i}$. If it holds for all $i = 1, \ldots, n$ that $\sigma B \overrightarrow{j} + \beta \overrightarrow{j} \geq \sigma B \overrightarrow{i} + \beta \overrightarrow{i}$, then we note that, for any follower strategy δ, the payoff under σ is an affine combination $\mathsf{fpayoff}(B; \langle\sigma, \delta\rangle, \beta) = \sum_{i \in S} \delta(i) \cdot \sigma B \overrightarrow{i} + \beta \overrightarrow{i}$ of the payoffs for the individual pure strategies. Using $\sigma B \overrightarrow{j} + \beta \overrightarrow{j} \geq \sigma B \overrightarrow{i} + \beta \overrightarrow{i}$, we get $\mathsf{fpayoff}(B; \langle\sigma, \delta\rangle) = \sum_{i \in S} \delta(i) \cdot \sigma B \overrightarrow{i} + \beta \overrightarrow{i} \leq \sum_{i \in S} \delta(i) \cdot \sigma B \overrightarrow{j} + \beta \overrightarrow{j} = \sigma B \overrightarrow{j} + \beta \overrightarrow{j} = \mathsf{fpayoff}(B; \langle\sigma, j\rangle, \beta)$. □

Theorem 3.8. *For a bi-matrix game $\mathcal{G}(A, B)$ with a bribery vector $\beta' = (\beta'_1, \ldots, \beta'_n)^T$, and for a simple bribery stable strategy profile $(\langle\sigma, j\rangle, \beta')$, we can replace β' by a j-bribery vector $\beta = (\beta_1, \ldots, \beta_n)^T$ with $\beta_j = \beta'_j$.*

Proof. According to Theorem 3.6, for a bi-matrix game $\mathcal{G}(A, B)$ and a bribery vector $\beta' = (\beta'_1, \ldots, \beta'_n)^T$, there is always a simple bribery stable strategy profile $\delta^T = \overrightarrow{j}$ with optimal payoff for the leader. We now choose $\delta^T = \overrightarrow{j}$ and $\beta = (\beta_1, \ldots, \beta_n)^T$, with $\beta_j = \beta'_j$ and $\beta_i = 0$ for all $i \neq j$. We first observe that $(\langle\sigma, j\rangle, \beta)$ provides the same leader and follower payoff as $(\langle\sigma, j\rangle, \beta')$. Next we observe that if $(\langle\sigma, j\rangle, \beta')$ is bribery stable, then so is $(\langle\sigma, j\rangle, \beta)$ by Theorem 3.7. In our equation systems, we can therefore focus on simple incentive equilibria $(\langle\sigma, j\rangle, \beta)$ with j-bribery β. The value of this j-bribery β can then be described by the value of the incentive $\iota = \beta_j$ that the leader gives to her follower to solicit him to play j. □

3.6.4 Computing incentive equilibria

Computing incentive equilibria invites the definition of a constraint system $\mathcal{C}_j^{\mathcal{G}(A,B)}$ for each pure follower strategy j, which describes the vectors σ by $\sigma = (p_1, \ldots, p_m)$. Theorem 3.7 and Theorem 3.8 allow to reflect the fact that $(\langle\sigma, j\rangle, \beta)$ can be assumed to be a simple bribery stable strategy profile with j-bribery β by using a constraint system. This constraint system $\mathcal{C}_j^{\mathcal{G}(A,B)}$ consists of $m+n+1$ constraints, where $m+1$ constraints describe that σ is a strategy,

- $\sum_{i=1}^{m} p_i = 1$ (the sum of the weights is 1) and
- the m non-negativity requirements $p_i \geq 0$ for $1 \leq i \leq m$,

and $n-1$ constraints reflect the conditions from Theorem 3.7 on an incentive equilibrium. That is, $\mathcal{C}_j^{\mathcal{G}(A,B)}$ contains $n-1$ constraints of the form

- $\sum_{k=1}^{m} (b_{jk} - b_{ik}) p_k + \iota \geq 0$,

one for each $i \neq j$ with $1 \leq i \leq n$, and a non-negativity constraint on the bribery value ι,

- $\iota \geq 0$,

As this reflects the conditions from Theorems 3.7 and 3.8, we first get the following corollary.

Corollary 3.9. *The solutions of* $\mathcal{C}_j^{\mathcal{G}(A,B)}$ *describe the set of leader strategy and* j*-bribery vector pairs* (σ, β) *such that* $(\langle \sigma, j\rangle, \beta)$ *is bribery stable.*

Example 3.1. *If we consider the first strategy of PrisonerII from the Prisoner's Dilemma from Table 3.2, then* $\mathcal{C}_1$ *consists of the constraints*

- $p_1 + p_2 = 1$,
- $\iota, p_1, p_2 \geq 0$, *and*
- $\iota - p_1 - 2p_2 \geq 0$.

Note that the constraints do *not* depend on the payoff matrix of the leader. What depends on the payoff matrix of the leader is the formalisation of the objective. We denote with $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ the linear programming problem that consists of the constraint system $\mathcal{C}_j^{\mathcal{G}(A,B)}$ and the objective function

$$\max \sum_{k=1}^{m} a_{jk} p_k - \iota$$

where $\sum_{k=1}^{m} a_{jk} p_k - \iota = \mathsf{lpayoff}(A; \langle \sigma, j\rangle, \beta)$ is the payoff the leader obtains for such a simple bribery stable strategy profile $(\langle \sigma, j\rangle, \beta)$ with $\sigma = (p_1, \ldots, p_m)$ and $\beta_j = \iota \geq 0$ is the bribery value of the j-bribery vector β.

Corollary 3.10. *The solutions to* $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ *describe the set of leader strategy and* j*-bribery vector pairs* (σ, β) *such that* $(\langle \sigma, j\rangle, \beta)$ *is bribery stable and the leader return* $\mathsf{lpayoff}(A; \langle \sigma, j\rangle, \beta)$ *is maximal among simple bribery stable strategy profiles with follower strategy* j *and a* j*-bribery vector.*

Example 3.2. *If we consider the first strategy of PrisonerII from the Prisoner's Dilemma from Table 3.2, then the* $\mathcal{LP}_1$ *consists of the constraints from* $\mathcal{C}_1$ *and the objective*

$$\max -p_1 - \iota.$$

This provides us with a simple algorithm for determining (simple) incentive equilibria.

Corollary 3.11. *To find an incentive equilibrium for a game* $\mathcal{G}(A, B)$*, it suffices to solve the linear programming problems* $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ *for all* $1 \leq j \leq n$*, to select an* i *with maximal solution among them, and to use a solution* $(\langle \sigma_i, i\rangle, \beta)$*, where* β *is a* i*-bribery vector with* $\beta_i = \iota$ *from the solution of* $\mathcal{LP}_i^{\mathcal{G}(A,B)}$*. This solution* $(\langle \sigma_i, i\rangle, \beta)$ *is an incentive equilibrium.*

Proof. Obviously, $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ has *some* solution iff $\mathcal{C}_j^{\mathcal{G}(A,B)}$ is satisfiable[1], and thus, by Corollary 3.9, if there is a leader strategy σ and a j-bribery vector β such that $(\langle \sigma, j\rangle, \beta)$

[1] Although it has no relevance for the proof, we would like to remark here that they all have a solution, as choosing a sufficiently high $\iota = \beta$, makes all the strategies $(\langle \sigma, j\rangle, \beta)$ with j-bribery vector bribery stable

is a simple bribery stable strategy profile. In this case, a solution to $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ reflects an optimal simple bribery stable strategy profile among the pure follower strategies that always play j; in particular, such an optimum exists. (Note that the leader return is bounded from above by the highest entry a_{ij} in her payoff matrix). Thus, the maximal return from some $\mathcal{LP}_i^{\mathcal{G}(A,B)}$ is the optimal solution among all simple bribery stable strategy profiles. Assuming that a better non-simple bribery stable strategy profile exists implies by Theorem 3.6 that a better simple incentive equilibrium exists as well and thus leads to a contradiction. Note that the existence of some simple bribery stable strategy profile is established by Corollary 3.5 and Theorem 3.6. □

As linear programming problems can be solved in polynomial time [Kar84, Kha79], finding an incentive equilibrium is tractable.

Corollary 3.12. *An optimal incentive equilibrium can be constructed in polynomial time.*

3.6.5 Friendly incentive equilibria

As discussed in the Section 3.2, the leader should follow a secondary objective of being benign to the follower. We have seen that it is cheap and simple to determine the value $v^l_{\max}$ that the leader can at most acquire in an incentive equilibrium. It is therefore an interesting follow-up question to determine the highest payoff $v^f_{\max}$ for her follower in an incentive equilibrium with leader payoff $v^l_{\max}$, i.e., to construct and evaluate *friendly* incentive equilibria. We first observe that friendliness does not come to the cost of simplicity.

Theorem 3.13. *For every bi-matrix game $\mathcal{G}(A,B)$ and every incentive equilibrium $(\langle\sigma,\delta\rangle,\beta)$ with payoff v for the follower, it holds for all $j \in \mathsf{support}(\delta^T)$ that $(\langle\sigma,j\rangle,\beta)$ is an incentive equilibrium with payoff v for the follower.*

Proof. Let $S = \mathsf{support}(\delta^T)$ be the support of δ^T and let $v^l_{\max} = \mathsf{lpayoff}(A;\langle\sigma,\delta\rangle,\beta)$ be the leader payoff for $(\langle\sigma,\delta\rangle,\beta)$. In the proof of Theorem 3.6 we have shown that, for all $j \in S$, $(\langle\sigma,j\rangle,\beta)$ is a simple bribery stable strategy profile with the same payoff $\mathsf{fpayoff}(B;\langle\sigma,j\rangle,\beta) = \mathsf{fpayoff}(B;\langle\sigma,\delta\rangle,\beta)$ for the follower. To establish that $(\langle\sigma,j\rangle,\beta)$ is also an incentive equilibrium, we first note that the leader payoff cannot be higher than in an incentive equilibrium, such that $\mathsf{lpayoff}(A;\langle\sigma,j\rangle,\beta) \leq v^l_{\max}$ holds for all $j \in S$. Assuming for contradiction that there is an $j \in S$ with $\mathsf{lpayoff}(A;\langle\sigma,j\rangle,\beta) < v^l_{\max}$ would, together with the previous observation that $\mathsf{lpayoff}(A;\langle\sigma,j\rangle,\beta) \leq v^l_{\max}$ holds for all $j \in S$, imply that the affine combination $\sum_{j\in S}\delta(j)\cdot\mathsf{fpayoff}(B;\langle\sigma,j\rangle,\beta) < v^l_{\max}$ of these values defined by δ is strictly smaller than $v^l_{\max}$, and would therefore lead to a contradiction. □

Theorem 3.14. *For a bi-matrix game $\mathcal{G}(A,B)$ and every incentive equilibrium $(\langle\sigma,\delta\rangle,\beta)$ with payoff v for the follower, there is always a pure follower strategy j with a j-bribery vector β' such that $(\langle\sigma,j\rangle,\beta')$ is an incentive equilibrium with follower payoff $\geq v$.*

Proof. First, according to Theorem 3.13, we can observe that there is always a pure follower strategy j, such that $(\langle \sigma, j \rangle, \beta)$ is an incentive equilibrium for the follower as it returns the maximal gain for the follower. Following the same argument as in the proof of Theorem 3.8, we can amend β by only incentivising the follower to play j, replacing all other entries β_i, $i \neq j$, by $\beta'_i = 0$ and setting $\beta'_j = \beta_j$. The payoff for the leader and follower are unaffected, and if $(\langle \sigma, j \rangle, \beta)$ is bribery stable, so is $(\langle \sigma, j \rangle, \beta')$: the follower return for playing j is unaffected, while the follower return for all other strategies is not increased. □

Like in the quest for ordinary incentive equilibria, we can therefore focus on pure follower strategies j and the respective j-bribery vectors when seeking friendly incentive equilibria. Recall that each constraint system $\mathcal{C}_j^{\mathcal{G}(A,B)}$ describes the set of leader strategies σ and gives a j-bribery vector β, such that $(\langle \sigma, j \rangle, \beta)$ is a simple bribery stable strategy profile. In order to be an incentive equilibrium, it also has to satisfy the optimality constraint

$$\sum_{k=1}^{m} a_{jk} p_k - \iota \geq v^l_{\max},$$

where $v^l_{\max}$ denotes the leader return for incentive equilibria. We refer to the extended constraint system by $\mathcal{E}_j^{\mathcal{G}(A,B)}$ By Corollary 3.9, the set of solutions to this constraint system is non-empty iff there is an incentive equilibrium of the form $(\langle \sigma, j \rangle, \beta)$.

Corollary 3.15. *The solutions to $\mathcal{E}_j^{\mathcal{G}(A,B)}$ describes the set of leader strategies σ and a bribery value ι, such that, for the j-bribery vector β with $\beta_j = \iota$, $(\langle \sigma, j \rangle, \beta)$ is an incentive equilibrium for $\mathcal{G}(A, B)$.*

Example 3.3. *Considering again the Prisoner's Dilemma from Table 3.2, $\mathcal{E}_1$ consists of the constraints from $\mathcal{C}_1$ plus the optimality constraint*

$$-p_1 - \iota \geq -2$$

We now extend the constraint system $\mathcal{E}_j^{\mathcal{G}(A,B)}$ to an extended linear programming problem $\mathcal{ELP}_j^{\mathcal{G}(A,B)}$ by adding the objective

$$\max \sum_{k=1}^{m} b_{jk} p_k + \iota$$

Corollary 3.16. *The solutions to $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ describe the set of leader strategy and j-bribery vector pairs (σ, β) such that $(\langle \sigma, j \rangle, \beta)$ is an incentive equilibrium that satisfies, if such a solution exists, that the follower return $\mathsf{fpayoff}(B; \langle \sigma, j \rangle, \beta)$ is maximal among these simple bribery stable strategy profiles with follower strategy j and j-bribery vectors β.*

Example 3.4. *If we consider the Prisoner's Dilemma from Table 3.2, then* $\mathcal{ELP}_1$ *consists of the constraints from* $\mathcal{E}_1$ *and the objective*

$$\max -p_1 - 10p_2 + \iota$$

Together with the observation of Theorem 3.13, Corollary 3.16 provides an algorithm for finding a friendly incentive equilibrium.

Corollary 3.17. *To find a friendly incentive equilibrium for a game* $\mathcal{G}(A, B)$*, it suffices to solve the linear programming problems* $\mathcal{ELP}_j^{\mathcal{G}(A,B)}$ *for all* $1 \leq j \leq n$*, to select an* i *with maximal solution among them, and to use a solution* $(\langle \sigma_i, i \rangle, \beta)$*, where* β *is an* i*-bribery vector with* $\beta_i = \iota$ *from the solution of* $\mathcal{ELP}_i^{\mathcal{G}(A,B)}$*. This solution* $(\langle \sigma_i, i \rangle, \beta)$ *is a friendly incentive equilibrium.*

Proof. $\mathcal{ELP}_j^{\mathcal{G}(A,B)}$ has *some* solution iff $\mathcal{E}_j^{\mathcal{G}(A,B)}$ is satisfiable, and thus, by Corollary 12, if there is a leader strategy σ such that $(\langle \sigma, j \rangle, \beta)$ is an incentive equilibrium. In this case, a solution to $\mathcal{ELP}_j^{\mathcal{G}(A,B)}$ reflects an incentive equilibrium with the maximal follower payoff among the incentive equilibria of the form $(\langle \sigma, j \rangle, \beta)$; in particular, such an optimum exists. Thus, the returned result is the optimal solution among all simple incentive equilibria. Assuming that a better non-simple friendly incentive equilibrium exists implies by Theorem 3.13 that a better simple friendly incentive equilibrium exists as well, and thus leads to a contradiction. Note that the existence of some simple optimal incentive equilibrium is implied by Corollary 3.11. □

As linear programming problems can be solved in polynomial time [Kar84, Kha79], finding friendly incentive equilibria is tractable.

Corollary 3.18. *A simple friendly incentive equilibrium can be constructed in polynomial time.*

3.6.6 Friendly incentive equilibria in zero-sum games

Here, we establish the friendliness of incentive equilibria in zero-sum games and show that the leader cannot gain anything by paying a bribery in zero-sum games. Zero-sum games are bi-matrix games where the gain of the leader is the loss of the follower and vice versa. They satisfy $a_{ij} = -b_{ij}$ for all $1 \leq i \leq m$ and all $1 \leq j \leq n$. Different to the general bi-matrix games, zero-sum games are determined in that they have determined expected payoffs for both players when both play rational. As in a zero-sum bi-matrix game, the payoff of player 1 is completely determined by the payoff of player 2 and vice versa. A rational behaviour for zero-sum games is therefore an acid (conclusive) test for new concepts: they do not have different levels of reasoning, and the opponent is predictable. We would like to make a few simple observations to show that incentive equilibria pass this acid test.

Theorem 3.19. *If $\langle \sigma, \delta \rangle$ is a Nash equilibrium in a zero-sum game and $j \in \mathsf{support}(\delta^T)$, then $\langle \sigma, j \rangle$ is a friendly incentive equilibrium with zero bribery vector β_0 and with* $\mathsf{lpayoff}(A; \langle \sigma, j \rangle, \beta_0) = \mathsf{lpayoff}(A; \langle \sigma, \delta \rangle, \beta_0)$.

Proof. We first establish that $(\langle \sigma, j \rangle, \beta_0)$ is a simple bribery stable strategy profile. To see this, we use that $(\langle \sigma, \delta \rangle, \beta_0)$ is a Nash equilibrium, and therefore $\mathsf{fpayoff}(B; \langle \sigma, j \rangle, \beta_0) \leq \mathsf{fpayoff}(B; \langle \sigma, \delta \rangle, \beta_0)$ holds for all $j \leq n$. Assuming that this inequation is strict for any $j \in \mathsf{support}(\delta^T)$ violates $\mathsf{fpayoff}(B; \langle \sigma, \delta \rangle, \beta_0) = \sum_{j=1}^{n} \delta(j) \cdot \mathsf{fpayoff}(B; \langle \sigma, j \rangle, \beta_0))$. Second, we observe that playing δ offers a return $\mathsf{fpayoff}(B; \langle \sigma', \delta \rangle, \beta_0) \geq \mathsf{fpayoff}(B; \langle \sigma, \delta \rangle, \beta_0)$, as otherwise the leader could benefit from deviating from $(\langle \sigma, \delta \rangle, \beta_0)$.

Now, to establish that this holds for all $j \leq n$, such that the strategy profile is simple, we use similar argument as above. I.e., using a similar "$\leq$, but not $<$ as it would contradict $\geq$ from the affine combination" argument, we can lead the assumption that $\mathsf{fpayoff}(B; \langle \sigma', \delta \rangle, \beta_0) > \mathsf{fpayoff}(B; \langle \sigma', j \rangle, \beta_0)$ holds to a contradiction for all $j \in \mathsf{support}(\delta^T)$. Consequently there is, for all leader strategies σ', a $j \in \mathsf{support}(\delta^T)$ such that $\mathsf{fpayoff}(B; \langle \sigma', j \rangle, \beta_0) \geq \mathsf{fpayoff}(B; \langle \sigma, \delta \rangle, \beta_0)$. The zero-sum property then provides the claim. □

Note that all incentive equilibria are for this reason (that these games are symmetric) friendly in zero-sum games. The definition of a simple bribery stable strategy profile now shows that the leader return $\mathsf{lpayoff}(A; \langle \sigma, j \rangle, \beta_0)$ can only be improved when the follower changes her strategy. (Note that this does not generally hold for non-zero-sum games.) Consequently, her incentive equilibrium provides her with the same guarantee as her rational strategy from zero-sum games. Her follower might be left exploitable, but in the selected strategy profile, he will receive the same payoff as with a rational strategy, but does not have to resort to randomisation. As these games are symmetric, this in particular implies that both players can play leader strategies in zero-sum games, and the leader can therefore not gain anything by paying a bribery (as it is applicable only when there is an asymmetry).

Corollary 3.20. *If $(\langle \sigma, j \rangle, \beta_0)$ is an incentive equilibrium in a zero-sum game $\mathcal{G}(A, B)$ and $(\langle \delta, i \rangle, \beta_0')$ is an incentive equilibrium in $\mathcal{G}(B, A)$, where β_0 and β_0' are the zero vectors, then $\langle \sigma, \delta \rangle$ is a Nash equilibrium in $\mathcal{G}(A, B)$.*

Naturally, this does not extend to general bi-matrix games.

3.6.7 Monotonicity and relative social optimality

As an interesting property of incentive equilibria, we observe that the payoff of the leader in incentive equilibria grows monotonously in A. A function grows *monotonously* in a matrix, if, for two matrices A and A' where each entry of A' is greater than or equal to the corresponding entry of A (intuitively: $a'_{ij} \geq a_{ij}$), then the value of the function is greater or equal.

Leader \ Follower	I	II
I	$-5, 1+\epsilon$	$-4, -1$
II	$5, 1$	$4, -1$
III	$0, 0$	$1, 0$

TABLE 3.8: Increasing the payoff matrix for the follower by ϵ

We show that the leader return in leader and incentive equilibria grows monotonously with A. Moreover, if each entry of A is strictly greater, then the leader return grows strictly.

Note that monotonicity does not hold for the follower in either case. Table 3.8 shows the effect of increasing the payoff matrix for follower by value ϵ. In the Table 3.8, for example, the friendly equilibrium changes from the pure leader strategy to play II and the pure follower strategy to play I for $\varepsilon = 0$ to the pure leader strategy to play III and the pure follower strategy to play II for $\varepsilon > 0$. As a result, the follower payoff is reduced from 1 to 0.

Theorem 3.21. *The leader payoff in incentive equilibria grows monotonously in the payoff matrix of the leader. If all entries grow strictly, so does the leader payoff.*

Proof. As observed earlier, the individual constraint systems do not depend on the payoff matrix of the leader, and the set of simple bribery stable strategy profiles is not affected by replacing a payoff matrix A by an entry-wise greater payoff matrix A'. An incentive equilibrium for A is thus a bribery stable strategy profile for A', too, and the payoff of this equilibrium for A' has the required properties. Consequently, an incentive equilibrium for A' has them as well. □

Theorem 3.22. *If $(\langle \sigma, j\rangle, \beta)$ is a friendly incentive equilibrium with j-bribery vector β, then j is the socially optimal response to σ.*

Proof. First, there is always a pure socially optimal response. For a pure follower response j to be socially optimal, the return for both players in a strategy profile $(\langle \sigma, j\rangle, \beta)$ should be better than any other pure follower response i. Let us assume for contradiction that there is a socially better pure response i. For i to be socially strictly better than j, it must hold that $\sigma A\overrightarrow{i} + \sigma B\overrightarrow{i} > \sigma A\overrightarrow{j} + \sigma B\overrightarrow{j}$. (Note that bribery is socially neutral.) This is equivalent to $\sigma A\overrightarrow{j} + \sigma B\overrightarrow{j} - \sigma A\overrightarrow{i} < \sigma B\overrightarrow{i}$. As $(\langle \sigma, j\rangle, \beta)$ is a friendly equilibrium and $\beta^T = (\beta_1, \ldots, \beta_n)$ is a j-bribery vector, we know that $\sigma B\overrightarrow{j} + \beta_j \geq \sigma B\overrightarrow{i}$. In order to incentivise the follower to play i, it suffices to choose an i-bribery vector $\beta' = (\beta'_1, \ldots, \beta'_n)^T$ such that $\sigma B\overrightarrow{i} + \beta'_i \geq \sigma B\overrightarrow{j} + \beta_j$, for which we can choose $\beta'_i = \beta_j + \sigma B\overrightarrow{j} - \sigma B\overrightarrow{i}$, which is non-negative as $(\langle \sigma, j\rangle, \beta)$ is bribery stable. Note that this immediately provides all inequalities from Theorem 3.7, such that $(\langle \sigma, j\rangle, \beta')$ is bribery stable. The leader payoff, however, would be $\mathsf{lpayoff}(A; \langle \sigma, i\rangle, \beta') = \sigma A\overrightarrow{i} - \beta'_i > (\sigma A\overrightarrow{j} + \sigma B\overrightarrow{j} - \sigma B\overrightarrow{i}) - \beta'_i = \sigma A\overrightarrow{j} - \beta_j = \mathsf{lpayoff}(A; \langle \sigma, j\rangle, \beta)$, which contradicts the optimality requirement of incentive equilibria. □

3.7 Secure incentive strategy profiles

As discussed in the Section 3.2, friendly incentive equilibria constitute a mutual considerate behaviour of the leader and follower. The extra computational effort for enforcing friendliness can be viewed as the price the leader has to pay for the consideration that she asks from her follower: to follow her suggestion unless it harms him. One can view this setting as a situation, where a rational follower has the main objective to maximise his return, and a secondary objective to maximise the return of the leader.

In this section, we discuss the impact of changing the follower model to one, where the follower is rational in that his main objective remains to maximise his return, but his secondary objective is reversed to harm the leader. Under this assumption, the leader no longer needs to be considerate to her follower, but she now has to *secure* her return against deviation.

An interesting observation is that a secure incentive equilibrium may not always exists (cf. Table 3.9), and when they do, a leader equilibrium also exists. We establish these results in Section 3.8. For the general case when they do not exist, we construct near optimal secure strategy profiles. Unsurprisingly, a near optimal strategy profile is easy to construct: when starting with a simple incentive equilibrium, the leader can secure it by raising the bribe by an arbitrarily small amount $\varepsilon > 0$. From a theoretical point of view, it is also interesting to determine whether or not there is also an optimal secure strategy profile.

Besides formalising the simple way of finding a secure ε-incentive equilibrium, we show that it suffices to seek secure incentive equilibria among leader equilibria: no incentive can be required for them. This provides a simple necessary criterion (equal leader return of incentive and leader equilibria), and a simple sufficient criterion (checking one such leader equilibrium for being secure). We show that checking (constructively) if a secure incentive equilibrium exists is tractable, though the algorithm is more intricate.

3.7.1 ε-optimal secure incentive strategy profiles

We start with the simple observation that there is, for all $\varepsilon > 0$, always a secure incentive strategy profile that provides a return to the leader which is at most ε worse for the leader than the return she obtains from an incentive equilibrium. This is because it suffices to increase the bribery value by ε to make an incentive strategy profile secure.

For ease of notation, we use, for a given bribery vector $\beta = (\beta_1, \ldots, \beta_n)$, with $\beta_j^\varepsilon = (\beta'_1, \ldots, \beta'_n)$ the bribery vector whose j^{th} entry is increased by ε ($\beta'_j = \beta_j + \varepsilon$) and whose other entries are not altered ($\beta'_i = \beta_i$ for all $i \neq j$).

Lemma 3.23. *For a simple incentive equilibrium $(\langle \sigma, j \rangle, \beta)$ and $\varepsilon > 0$, $(\langle \sigma, j \rangle, \beta_j^\varepsilon)$ is a simple secure incentive strategy profile.*

Proof. As $(\langle \sigma, j \rangle, \beta)$ is a simple incentive equilibrium, the follower has no incentive to deviate from j. In particular, the follower's return upon playing any other pure strategy is not better than pure strategy j.

Consequently, in $(\langle\sigma, j\rangle, \beta_j^\varepsilon)$, the follower will *lose* at least ε when deviating to any other pure strategy j', such that $j' \neq j$. He therefore loses strictly upon any deviation from j. □

Simple secure ε-equilibria are therefore simple to construct.

Theorem 3.24. *For a simple incentive equilibrium $(\langle\sigma, j\rangle, \beta)$ and $\varepsilon > 0$, $(\langle\sigma, j\rangle, \beta_j^\varepsilon)$ is a simple ε-incentive equilibrium.*

Proof. Lemma 3.23 shows that $(\langle\sigma, j\rangle, \beta_j^\varepsilon)$ is an incentive equilibrium, and it is obvious that the leader's return is ε lower than for the simple incentive equilibrium $(\langle\sigma, j\rangle, \beta)$.

Let us assume for contradiction that there is a secure incentive strategy profile $(\langle\sigma', \tau'\rangle, \beta')$, where the leader return exceeds the leader return by more than ε. Then $(\langle\sigma', \tau'\rangle, \beta')$ exceeds the leader return of $(\langle\sigma, j\rangle, \beta)$. But as secure incentive equilibria are in particular incentive equilibria, this contradicts the assumption that $(\langle\sigma, j\rangle, \beta)$ is an incentive equilibrium. □

It is therefore enough to focus on simple secure ε-incentive equilibria. Note that the above lemma also provides a recipe to construct near optimal secure strategy profiles: they can be obtained from simple incentive equilibria by increasing the bribery value slightly. By Lemma 3.23 and Corollary 3.18, their construction is therefore tractable.

Corollary 3.25. *A simple secure ε-incentive equilibrium can be constructed in polynomial time.*

Another corollary of Theorem 3.24 is that the value of ordinary and secure incentive equilibria – when they exist – cannot be different, as they cannot differ by more than ε for all $\varepsilon > 0$.

Corollary 3.26. *When a secure incentive equilibrium exists, then it is also an incentive equilibrium.*

3.8 Secure incentive equilibria

The easy way to construct near optimal simple secure incentive strategy profiles raises the immediate question if there always exists an optimal one. This is, unfortunately, not the case. Consider, for example, the bi-matrix game below.

		Follower	
		I	II
Leader	I	1, 0	0, 0

TABLE 3.9: A simple bi-matrix game without a secure incentive equilibrium.

In leader and ordinary incentive equilibria, the leader can influence the game by *suggesting* the follower to play I. Indeed, suggesting to play I while paying no incentive for doing so is the only incentive (and leader) equilibrium.

In a setting where the follower has a secondary objective to harm the leader, this suggestion has no avail. The only way for the leader to secure a payoff > 0 is to incentivise her follower to play strategy I by offering him a small bribery value of $\varepsilon > 0$. It is apparent that any value $\varepsilon > 0$ would do, while 0 itself is insufficient. Consequently, the leader can obtain any payoff < 1, but not 1, such that no optimal secure incentive strategy profile exists.

Lemma 3.27. *Secure incentive strategy profiles do not always exist.*

We will now show that, when a secure incentive equilibrium exists, there is one that is also a simple leader equilbrium.

Theorem 3.28. *If a secure incentive equilibrium exists, then there exists one, which is also a simple leader equilibrium.*

Proof. Let $(\langle \sigma, \tau \rangle, \beta)$ be a secure incentive equilibrium. We first observe that there is no pure strategy j such that the payoff of the follower would increase strictly when he plays j, and that, for all j where his payoff remains the same as for τ, the payoff for the leader would not decrease. Let us denote the set of pure strategies of the follower such that his payoff remains the same by J'. Let us denote the set of pure strategies of the follower such that both his payoff and the payoff of the leader remain the same by J. Note that J must include the support of τ.

We now distinguish three cases.

1. There is a strategy $j \in J$ such that $\beta_j = 0$.

   As for a pure follower strategy $j \in J$, the payoff of both the leader and the follower remains same and with $\beta_j = 0$, $(\langle \sigma, j \rangle, 0)$ is a simple secure incentive equilibrium (with zero bribery) and a simple leader equilibrium.

2. There is no strategy $j \in J$ with $\beta_j = 0$ and $J = J'$.

   As the follower loses on all pure strategies not in J, there is a minimal amount ε he loses on any of them. Let $\varepsilon' = \min\{\beta_j \mid j \in J\}$ and $\delta = \frac{1}{2}\min\{\varepsilon, \varepsilon'\}$. We then derive β' from β by setting $\beta'_j = \beta_j - \delta$ for all $j \in J$ and $\beta'_j = \beta_j$ otherwise.

   $(\langle \sigma, \tau \rangle, \beta')$ is then a secure incentive strategy profile with a higher payoff (by $\delta > 0$) for the leader compared to $(\langle \sigma, \tau \rangle, \beta)$.

   As a secure incentive equilibrium exists as an incentive equilibrium (Corollary 3.26), this contradicts the assumption that $(\langle \sigma, \tau \rangle, \beta)$ is an incentive equilibrium.

3. There is no strategy $j \in J$ with $\beta_j = 0$ and $J \neq J'$.

   As the follower loses on all pure strategies not in J', there is a minimal amount ε he loses on any of them. As the leader gains strictly more on all pure strategies in $J' \setminus J$, there is a minimal amount ε' on the increase of her gain among them. Let $j' \in J' \setminus J$ be a pure strategy of the follower for which the leader would gain this minimal ε'.

Let $\varepsilon'' = \min\{\beta_j \mid j \in J\}$ and $\delta = \frac{1}{2}\min\{\varepsilon, \varepsilon', \varepsilon''\}$. We then derive β' from β by setting $\beta'_j = \beta_j - \delta$ for all $j \in J$ and $\beta'_j = \beta_j$ otherwise.

$(\langle \sigma, j' \rangle, \beta')$ with $j' \in J' \setminus J$ is then a secure incentive strategy profile with a higher payoff (by $\varepsilon' > 0$) for the leader compared to $(\langle \sigma, \tau \rangle, \beta)$.

As a secure incentive equilibrium exists as an incentive equilibrium (Corollary 3.26), this contradicts the assumption that $(\langle \sigma, \tau \rangle, \beta)$ is an incentive equilibrium.

□

Note, however, that the bi-matrix game from Table 3.9 shows that leader equilibria with the same leader return as incentive equilibria is not a sufficient criterion.

Corollary 3.29. *For the existence of secure incentive equilibria, the existence of leader equilibria with the same leader return as incentive equilibria is a necessary, but not a sufficient criterion.*

3.8.1 Constructing secure incentive equilibria – outline

We give here an outline of a tractable and constructive test for the existence of secure incentive equilibria. Corollary 3.29 establishes that secure incentive equilibria can be sought among simple leader equilibria. It therefore suffices to consider leader strategy profiles when testing the existence of secure incentive equilibria.

We develop our test by making increasingly weaker assumptions on the knowledge we have. We start with knowing a solution: a secure leader strategy profile, which is also a secure incentive strategy profile (with zero incentive). We continue with assuming to have abstract information about such a solution, namely knowing on which deviation of the follower the leader would lose. We then close by assuming only knowledge about the pure follower strategy the leader advises.

Assume we already know a strategy profile $\langle \sigma, j \rangle$ that is a secure leader strategy profile. For the strategy profile $\langle \sigma, j \rangle$ to be a secure incentive equilibrium, we first identify the following important criterion that it needs to satisfy:

- The strategy profile $\langle \sigma, j \rangle$ is a leader strategy profile.
- $\mathsf{lpayoff}(\langle \sigma, j \rangle) \geq v^L_{IE}$, i.e., no ordinary incentive equilibrium exists with a greater return.
- The strategy profile $\langle \sigma, j \rangle$ is secure. Thus, for any follower deviation where the leader loses strictly, the follower should also lose strictly.

As we already know the strategy profile $\langle \sigma, j \rangle$, we define the sets J_{noLoss} and J_{loss} as follows.

Definition 3.30. The set $J_{\mathsf{noLoss}} = \{i \neq j \mid \mathsf{lpayoff}(A; (\sigma, i)) \geq \mathsf{lpayoff}(A; (\sigma, j))\}$ is the set of indices, where the leader does not lose when the follower deviates from the advised strategy j to i.

Definition 3.31. The set $J_{\mathsf{loss}} = \{i \neq j \mid \mathsf{lpayoff}(A;(\sigma,j)) > \mathsf{lpayoff}(A;(\sigma,i))\}$ is the set of indices, where the leader loses strictly when the follower deviates from the advised strategy j to i.

For any follower deviation from the strategy j to a different pure strategy i in the set J_{loss}, we have that the follower should also lose strictly. We denote by κ the minimal amount that the follower loses from any deviation to a different pure strategy in J_{loss}. We observe that the value of κ is strictly greater than 0. This is because, for all those strategies where the leader loses strictly, the follower is also bound to lose strictly, and the set J_{loss} of candidate strategies is finite.

We now assume that we only have abstract information about the strategy. That is, we know a pure follower strategy j and the set J_{loss}. For the set J_{loss} of strategies, we again have that the follower shall lose strictly. Thus, for all follower deviations to any strategy in J_{loss}, we have a constraint that the value of $\kappa > 0$. However, the strict constraint on a value cannot be formulated in a standard linear programming problem. Therefore, we encode it by an objective to maximise the minimal follower loss.

Finally, assume that we do not know a simple secure leader equilibrium and we know only about the pure follower strategy j. Note that there are only n many pure candidate strategies for the follower, and we can check them all.

As opposed to that, for n pure follower strategies, we have a total of $n \cdot 2^{n-1}$ strategy combinations of j and J_{loss}. The number of linear programmes formed are too many such that the approach to solve all of these many linear programmes is not tractable.

We can, however, estimate the value of κ rather than computing it. We know that we are only interested in solutions with a positive κ. For our estimation, we use the smallest positive $\kappa > 0$ that can be computed by Karmarkar's algorithm in the running time it needs to solve the linear programmes. This estimation is good enough, as the estimation of κ is written in polynomial time and thus has polynomial length.

3.8.2 Existence of secure incentive equilibria

Based on the above discussion, we will now discuss a more involved, but tractable, technique for checking whether or not secure incentive equilibria exist. The test is constructive, and provides a simple secure leader equilibrium (which is also a secure incentive equilibrium) in case secure incentive equilibria exist. As Theorem 3.28 allows for seeking the secure incentive equilibria only among simple leader equilibria, we adjust the constraint system needed for the computation of leader equilibria accordingly. The constraints required for the computation of a leader equilibrium are given in Appendix A.1.1.

Exploiting Theorem 3.28, a natural step when checking the existence of secure incentive equilibria is therefore to compare the value of leader and incentive equilibria. This can be done using the algorithms from Section 3.6 (for incentive equilibrium) and the techniques given in Appendix A.1 (for leader equilibrium). If they differ, we do not have to look further. In case the leader return from friendly incentive equilibrium and

leader equilibrium are equal, it is worth checking if there are finitely many such strategy profiles. In this case, we can simply check if one of them is secure.

We start with the assumption that we already know a secure leader strategy profile, and, therefore, a simple secure incentive equilibrium (with zero incentive). We then also get the sets J_{loss} and J_{noLoss}. Before we study the adjusted constraint system, we observe that, if we already know a simple secure leader equilibrium $\langle \sigma, j \rangle$, then this leader equilibrium also satisfies an additional side constraint[2].

Lemma 3.32. *Let $\mathcal{G}$ be a bi-matrix game with simple secure leader equilibrium $\langle \sigma, j \rangle$ and, therefore, with the simple secure incentive equilibrium $(\langle \sigma, j \rangle, 0)$. Then there is a $K' \geq 0$, such that, for all $i \neq j$ and for all $K \geq K'$, $\mathsf{lpayoff}(A; (\sigma, i)) + K \cdot \mathsf{fpayoff}(B; (\sigma, j)) \geq \mathsf{lpayoff}(A; (\sigma, j)) + K \cdot \mathsf{fpayoff}(B; (\sigma, i))$ holds.*

Proof. The proof almost follows from the definition of secure leader strategy profiles. For the individual $i \neq j$, we have that,

1. For a given strategy j and the set J_{noLoss}, we observe that (σ, j) is a simple leader strategy profile. Therefore $\mathsf{fpayoff}(B; (\sigma, j)) \geq \mathsf{fpayoff}(B; (\sigma, i))$ holds.

   With the definition of J_{noLoss}, we get that $\mathsf{lpayoff}(A; (\sigma, i)) + K \cdot \mathsf{fpayoff}(B; (\sigma, j)) \geq \mathsf{lpayoff}(A; (\sigma, j)) + K \cdot \mathsf{fpayoff}(B; (\sigma, i))$ holds for all $K \geq 0$.

2. Note that the secure leader equilibrium condition states that, if the leader loses strictly from any deviation, then the follower should also lose strictly. We therefore find the follower loss upon deviation from the pure strategy j to any other strategy in the set J_{loss}. We first note that there is nothing to show when J_{loss} is empty. Otherwise, we denote by κ the minimal loss that the follower might incur from all the possible deviations to J_{loss}. That is,

   $$\kappa = \min \left\{ \mathsf{fpayoff}(B; (\sigma, j)) - \mathsf{fpayoff}(B; (\sigma, i)) \mid i \in J_{\mathsf{loss}} \right\}$$

   and observe that $\kappa > 0$, because the definition of J_{loss} requires for simple secure leader strategy profiles that the follower loses strictly on deviation to an index in J_{loss}.

When we select $K' = \frac{\|A\|}{\kappa}$, where $\|A\|$ denotes the maximal absolute difference between two entries of $\mathcal{G}$, then the inequation holds for all $K \geq K'$. □

We now give the constraint system using a constant $K \geq 0$ for computing secure leader equilibria.

Note that the construction of the linear programme is based on the knowledge of a suitable constant K, e.g., the one given above. We will, however, see that, irrespective of the $K \geq 0$ used, all solutions to the constraint system are secure leader equilibria. Only the guarantee that there is a solution (provided that there exists a secure leader equilibria) depends on choosing a sufficiently large K.

[2]Note that, technically we do not need to establish these properties when we already have such a strategy profile $\langle \sigma, j \rangle$, but we need these properties later in our proofs.

Linear programming problems for constructing secure leader equilibria We give here a constraint system $\mathcal{SLE}_j^{\mathcal{G}(A,B)}$ to compute the secure leader equilibria using such a constant K. For this, we extend the solution from leader equilibria to use the value of K. This gives us a secure leader equilibrium that is also the secure incentive equilibrium. The constraint system $\mathcal{SLE}_j^{\mathcal{G}(A,B)}$ consists of $m+n+1$ constraints, where $m+1$ constraints describe that σ is a strategy,

- $\sum_{i=1}^{m} p_i = 1$ (the sum of the weights is 1) and
- the m non-negativity requirements $p_i \geq 0$ for $1 \leq i \leq m$,

and $n-1$ constraints reflect the conditions on a secure leader equilibrium using the value of constant K. That is, for each $i \neq j$ with $1 \leq i \leq n$, we add the following constraint.

That is deviation to pure strategy i would cost the follower atleast the value κ.

$$\mathsf{fpayoff}(B;(\sigma,j)) \geq \mathsf{fpayoff}(B;(\sigma,i)) + \kappa$$

$$\sum_{k=1}^{m}(a_{ki}-a_{kj})p_k + K \cdot \sum_{k=1}^{m}(b_{kj}-b_{ki})p_k \geq 0$$

Additionally, we have the same constraint here that the leader's reward is not less than her reward from an incentive equilibrium.

$$\sum_{k=1}^{m} a_{kj}p_k \geq v_{IE}^{L}.$$

Where the value v_{IE}^{l} is the leader's reward from an incentive equilibrium. We denote with $\mathcal{SLP}_j^{\mathcal{G}(A,B)}$ the linear programming problem that consists of the constraint system $\mathcal{SLE}_j^{\mathcal{G}(A,B)}$ and the objective function is to maximise the leader's return, that is, $\max \sum_{k=1}^{m} a_{kj}p_k$.

Theorem 3.33. *For any given $K \geq 0$, any solution to the above constraints is a proper solution: (σ, j) is a secure leader strategy profile, that is also a secure incentive strategy profile $(\langle \sigma, j \rangle, 0)$.*

Proof. We start by observing that any such strategy profile (σ, j) is a leader strategy profile with a proper return value for the leader. That is, the leader return from (σ, j) is no less than her return from any incentive equilibrium.

What remains to be shown is that the strategy profile is also secure. To show this, we have the following observations. For all i in J_{noLoss}, there is nothing to show.

For all i in J_{loss}, the leader loses strictly. That is, the strict inequation for the leader payoff $\mathsf{lpayoff}(A;(\sigma,i)) < \mathsf{lpayoff}(A;(\sigma,j))$ is satisfied. But at the same time, for these strategies, $\mathsf{lpayoff}(A;(\sigma,i)) + K \cdot \mathsf{fpayoff}(B;(\sigma,j)) \geq \mathsf{lpayoff}(A;(\sigma,j)) + K \cdot \mathsf{fpayoff}(B;(\sigma,i))$ has to be satisfied, too.

This naturally implies $K \cdot \mathsf{fpayoff}(B;(\sigma,j)) > K \cdot \mathsf{fpayoff}(B;(\sigma,i))$ also holds. For any $K \geq 0$, this implies $\mathsf{fpayoff}(B;(\sigma,j)) > \mathsf{fpayoff}(B;(\sigma,i))$. □

Corollary 3.34. *For all bi-matrix games $\mathcal{G}$, there is a $K \geq 0$ such that, iff $\mathcal{G}$ has a secure incentive equilibrium, then it has a secure incentive equilibrium $(\langle\sigma,j\rangle,0)$, which is also a simple leader equilibrium and, for all $i \neq j$, it satisfies $\mathsf{lpayoff}(A;(\sigma,i)) + K \cdot \mathsf{fpayoff}(B;(\sigma,j)) \geq \mathsf{lpayoff}(A;(\sigma,j)) + K \cdot \mathsf{fpayoff}(B;(\sigma,i))$.*

Example 3.5. *We consider a bi-matrix game from Table 3.10. The game is a variant of the battle-of-sexes game where the follower has now three options to select from, while the leader has only two. If we consider the strategy profile (II,II) from Table 3.10 and assume it to be a secure strategy profile, then we note that the set J_{loss} and J_{noLoss} consists of the following pure follower strategies*

- $J_{\mathsf{loss}} = \{I\}$
- $J_{\mathsf{noLoss}} = \{III\}$

		Player 2		
		I	II	III
Player 1	I	$1,3$	$0,0$	$4,2$
	II	$0,0$	$3,1$	$3,-3$

TABLE 3.10: A variant of Battle-of-Sexes game

We first check that the leader reward from the strategy profile (II,II) is not less than her reward from an incentive equilibrium. The strategy profile (I,III) is a (friendly) incentive equilibrium with leader reward 3. This constraint is therefore satisfied. We then check the other necessary constraints. Note that the minimal loss of the follower upon deviation from strategy II to any strategy in the set J_{loss} is 1 and therefore the value of κ and K for this strategy set are 1 and 7 respectively. For this value of κ and K, the following constraint from Corollary 3.34 is satisfied for all $i \neq j$.

$$\mathsf{lpayoff}(A;(\sigma,i)) + K \cdot \mathsf{fpayoff}(B;(\sigma,j)) \geq \mathsf{lpayoff}(A;(\sigma,j)) + K \cdot \mathsf{fpayoff}(B;(\sigma,i))$$

Example 3.6. *If we consider the strategy profile (I,III) from the bi-matrix game from Table 3.10, then we note that the set J_{noLoss} is empty while the set J_{loss} consists of the following pure follower strategies*

- $J_{\mathsf{loss}} = \{I,II\}$

The value of κ is -1 and therefore the strategy profile is not secure. Note that the strategy profile (I,III), which is a friendly incentive equilibrium, is not secure as the follower might deviate to the pure strategy I causing the leader to lose. We have to restrict ourselves to only the strategy profiles where the value of $\kappa > 0$ is satisfied.

3.8.3 Construction of secure incentive equilibria – Given a strategy j and a set J_{loss}

We give here the detailed techniques for the construction of a secure incentive equilibrium, if one exists. This subsection is therefore concerned with estimating a constant from Corollary 3.34 and proving it to be big enough. The estimation of such a constant is oriented at the proof of Lemma 3.32. We start with lifting the assumption that we know a simple secure leader equilibrium, but we do know that a simple secure leader equilibrium $\langle \sigma, j \rangle$ with a set J_{loss} exists. For this, we also need to estimate κ from the second case of the proof of Lemma 3.32.

We are now equipped with the following information. We have the strategy j that we assign, the leader payoff (as from any incentive equilibrium), J_{loss}, and J_{noLoss}. We now write a constraint system using all this information and set an objective to maximise κ. That is, we now compute a maximal κ. For the construction of κ, we simply expand the constraint system for simple leader equilibria that assign the strategy j in three ways. We give an adjusted constraint system $\mathcal{LA}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ here that is an adjustment of the constraint system for simple leader equilibria (cf., A.1.1).

1. We add a constraint that the payoff of the leader is the same as for the leader equilibria (cf., Appendix A.1.1 for the constraint system for computing the leader equilibria) and for the incentive equilibria (cf., Section 3.6 for the constraint system for computing the incentive equilibria).

2. We add, for all $i \in J_{\mathsf{noLoss}}$, a constraint $\mathsf{lpayoff}(A;(\sigma,i)) \geq \mathsf{lpayoff}(A;(\sigma,j))$

3. We adjust, for all $i \in J_{\mathsf{loss}}$, the constraints for the follower to $\mathsf{fpayoff}(B;(\sigma,j)) \geq \mathsf{fpayoff}(B;(\sigma,i))+\kappa$ (i.e., we require that deviation to i costs the follower at least κ).

Extended constraint system $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ For a known set J_{loss}, we now turn to the linear programming problem denoted by $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ that consists of the constraint system from $\mathcal{LA}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ and the objective function is to maximise the value of κ. We define a constraint system $\mathcal{LA}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ for each pure follower strategy j and the set J_{loss}. The constraint system is an extension of $\mathcal{LP}^{\mathcal{G}(A,B)}_{j}$ (given in Appendix A.1) and assigns the strategy j as described above. We give the extended constraint system as follows.

That is, $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ first contains a constraint on the reward of the leader. We denote by v^l_{IE} the leader's reward from an incentive equilibrium that we can get from the Algorithm 1. We therefore add the following constraint,

$$\sum_{k=1}^{m} a_{kj} p_k \geq v^l_{IE}.$$

Where $\sum_{k=1}^{m} a_{kj} p_k$ is the leader's reward from the leader equilibrium. We then add the following constraints.

For all pure strategies $i \in J_{\mathsf{noLoss}}$, we add a constraint

$$\sum_{k=1}^{m}(a_{ki} - a_{kj})p_k \geq 0$$

For all pure strategies $i \in J_{\mathsf{loss}}$, we add a constraint

$$\sum_{k=1}^{m}(b_{kj} - b_{ki})p_k \geq \kappa$$

For the linear programming problem $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$, we now give the following objective function

$$\max \kappa$$

We recall Example 3.5 here to note that the value of κ is strictly greater than 0 for a secure incentive equilibrium. For all other strategy profiles (cf., Example 3.6), where the value of $\kappa \leq 0$, the constraint system would not give an optimal solution, that is, a secure incentive equilibrium does not exists.

Theorem 3.35. *If no solution with a $\kappa > 0$ exists, then there exists no secure leader strategy profile (σ, j) and a secure incentive strategy profile $(\langle\sigma, j\rangle, 0)$ with the given set J_{loss} of deviations, where the leader loses strictly.*

If a solution with a $\kappa > 0$ exists, then there exists a secure leader strategy profile (σ, j) that satisfies the constraints from above. $(\langle\sigma, j\rangle, 0)$ is then also a secure incentive strategy profile.

Note that the set of strategies $J^{(\sigma,j)}_{\mathsf{loss}}$ obtained from the strategy profile (σ, j), is not necessarily the set J_{loss} used when defining the constraint system, as there is no guarantee that the leader loses when the follower deviates to a strategy $i \in J_{\mathsf{loss}}$. However, note that, $J^{(\sigma,j)}_{\mathsf{loss}} \subseteq J_{\mathsf{loss}}$ holds.

Corollary 3.36. *For all such solutions where the value of $\kappa > 0$ exists, a constant $K = \frac{\|A\|}{\kappa}$ exists, that is suitable for use in Lemma 3.32.*

3.8.4 For an unknown set J_{loss}

As can be observed, computing such a maximal κ requires to solve $n \cdot 2^{n-1}$ linear programmes. For a given pure follower strategy j and a set J_{loss}, we have a total of $n \cdot 2^{n-1}$ strategy combinations, and therefore, these many constraint systems. On the first glance, this seems of limited use, as solving all these linear programmes is not a tractable solution. Computing a maximal κ requires to solve too many linear programmes such that the complexity is exponential in the size of bi-matrix. We, therefore, estimate the smallest minimal value of $\kappa > 0$ that serves as an estimation for all of them.

Lemma 3.37. *The estimation of κ can only be approximated and we would therefore get an approximate solution to the linear programme $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$.*

We first observe from the examples 3.5 and 3.6 that any value of κ where $\kappa' \geq \kappa > 0$ holds, would serve our purpose. Here, κ' is one such value that is greater than 0, if a solution exists. We, therefore, do not need to compute a maximal κ and thus do not need to solve all these linear programmes. For our estimation, we simply use the smallest κ that is greater than 0 and smaller than any of the solutions to one of these linear programmes.

Estimating the value of κ We use an estimated value of κ to solve the linear programmes from above. We note that the estimation is good enough for the representation of the input length in polynomial size. The running time of Karmarkar algorithm is $\mathcal{O}(m^{3.5}L^2)$ [Kar84], where L is the number of bits of input to the algorithm, and m is the number of variables. Input variable in our case is the number of leader strategies and one extra variable for the value of κ. L and m are now polynomial in the bi-matrix and are easy to estimate. We only need to have an estimation for the running time of the largest constraint system. Once we have a polynomial estimate of the running time of K, we can have a polynomial estimate of writing κ in any of the $n \cdot 2^{n-1}$ linear programmes.

Therefore, the technique remains tractable when using an estimated value of κ instead of computing κ by solving too many linear programmes. We observe that we are only interested in the value of $\kappa > 0$. Reconsidering this from Lemma 3.37, we can simply ignore all other κ, such that for our purpose, we may use any value of κ that satisfy $\kappa' \geq \kappa > 0$. Here κ' is one such $\kappa > 0$ such that a solution to one of the linear programme exists [3]. Using such an estimation of κ from above, we can therefore solve linear programmes $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$. These polynomial time algorithms [Kar84, Kha79] need to write κ, such that we only use the smallest $\kappa > 0$ that can be written in the running time of Karmarkar's algorithm. This therefore also gives us an estimation for the running time of the largest constraint system. The resulting κ then provides for a sufficiently large $K = \frac{\|A\|}{\kappa}$, which is big enough for the constant K in Corollary 3.34.

Remark 3.38. We only estimate the value of κ and use such an estimation to determine the value of a suitable constant K.

Computing a suitable constant K The existence of a simple secure leader equilibrium implies that, unless J_{loss} is empty (in which case we can choose $K = 0$), we can estimate a suitable minimal κ, and therefore a suitable K. This estimation would give us a value of κ such that, if any of the linear programmes $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ has a solution $\kappa' > 0$, then it is fine to use the value of κ that satisfy $\kappa' \geq \kappa > 0$, and use this value to determine a suitable $K = \frac{\|A\|}{\kappa}$. Note that using such an estimation of κ that can be written

[3] Note that κ' refer to any value of $\kappa > 0$ and we get this value only if a solution to one of the linear programme $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$ exists. Once we have a polynomial estimate of the running time of K, we can have a polynomial estimate of the writing κ in any of the $n \cdot 2^{n-1}$ linear programmes. We can then estimate $\kappa = \min\{\kappa > 0, \kappa'\}$.

For the cases where no such κ' exists, we may use the smallest positive number that the Karmarkar algorithm can write within its running time for any of the $\mathcal{LAP}^{\mathcal{G}(A,B)}_{j,J_{\mathsf{loss}}}$.

in Karmarkar's algorithm, we also get an estimation of the resulting constraint system. Note that, even for a small bi-matrix from Example 3.5, the value of constant K would be hundred of digits. Using this largest value of K, the size of the linear programme can be determined. This provides us with the following theorem.

Theorem 3.39. *For a bi-matrix game $\mathcal{G}$ and a given pure follower strategy j, we can construct a $K \geq 0$ in polynomial time, such that the following holds. If $\mathcal{G}$ has a simple leader equilibrium $\langle \sigma, j \rangle$, which is also a secure incentive equilibrium, then it has a simple leader equilibrium $\langle \sigma', j \rangle$, which is also a secure incentive equilbrium and satisfies $\mathsf{lpayoff}(A; (\sigma', i)) + K \cdot \mathsf{fpayoff}(B; (\sigma', j)) \geq \mathsf{lpayoff}(A; (\sigma', j)) + K \cdot \mathsf{fpayoff}(B; (\sigma', i))$ for all $i \leq n$.*

The resulting K is big enough for use in Corollary 3.34. We can therefore use such a K as determined from the estimated smallest κ for extending the constraint system from Appendix A.1.1 by the inequations from Lemma 3.32.

Theorem 3.40. *Let $\mathcal{G}$ be a bi-matrix game with a similar value for incentive and leader equilibria. Then we can construct a K in polynomial time such that the extended linear programme described above has a solution if, and only if, $\mathcal{G}$ has a secure incentive equilibrium for some pure follower strategy j. A solution defines a strategy profile $\langle \sigma, j \rangle$, which is a simple leader equilibrium, such that $(\langle \sigma, j \rangle, 0)$ is a secure incentive equilibrium for $\mathcal{G}$.*

Testing the existence of such a solution, and determining one in case one exists, can be done in time polynomial in $\mathcal{G}$.

3.9 Evaluation

In this section, we give details of our proof-of-concept implementation of friendly incentive equilibrium and the results obtained. We have randomly generated two sets of benchmarks with uniformly distributed entries in the bi-matrices. One set of benchmarks uses continuous payoff values in the range from 0 to 1 (Table 3.11), and the other set of benchmarks uses integer payoff values in the range from -10 to 10 (Table 3.12). They contain samples of 100,000 games for each matrix form covered.

An incentive equilibrium (*IE*) of a bi-matrix game $\mathcal{G}(A, B)$ can be computed by solving the linear programming problems from Corollaries 3.10 and 3.16. The result is a simple strategy profile $(\langle \sigma, j \rangle, \beta)$, where j is a pure strategy of the follower, σ is given as a tuple of probabilities that describe the likelihood the leader chooses her individual strategies, and β is a j-bribery vector. We have implemented Algorithm 1 for computing friendly incentive equilibrium, using the LP solver [MKP], taken from `http://lpsolve.sourceforge.net/5.5/`. The algorithm returns a friendly *IE* in form of a strategy profile $(\langle \sigma, j \rangle, \beta)$, as well as the payoffs obtained by the follower and leader in the friendly *IE* under the j-bribery vector returned for the given bi-matrix game. We have implemented our algorithm in C and our implementation is available at

`http://cgi.csc.liv.ac.uk/~anshul/BIMatrix`. We have used GAMBIT [MMT13] to compute the Nash equilibria (NE). The data size is given in terms of number of follower strategies ($\#FSt$) and the number of leader strategies ($\#LSt$).

Experimental results We analysed the outcome of the random games along the parameters 'average optimal return value of the leader ($LeadR$)', 'average optimal return value of the follower ($FollR$)', 'average bribery value', and the confidence interval radius for both leader and follower return for a 95% confidence interval. We also gave the execution time for 100,000 games. The highest average execution time observed, which is obtained for friendly incentive equilibria, is 21.5 milliseconds.

Algorithm 1: The Algorithm outputs a (pure) friendly incentive equilibrium $(\langle\sigma, j\rangle, \beta)$, a j-bribery vector β, leader payoff $\mathsf{lpayoff}(A; \langle\sigma, j\rangle, \beta)$ and follower payoff $\mathsf{fpayoff}(B; \langle\sigma, j\rangle, \beta)$

Input: A bi-matrix game $\mathcal{G}(A, B)$
Output: A friendly incentive equilibrium $\langle\sigma, j\rangle$, a j-bribery vector β (represented by the value $\iota = \beta_j$), $\mathsf{lpayoff}(A; \langle\sigma, j\rangle, \beta)$, and $\mathsf{fpayoff}(B; \langle\sigma, j\rangle, \beta)$

$max \leftarrow \min\{a_{ij} \mid i \leq m, j \leq n\}$
// initialise the leader payoff to minimal entry of A
$opt \leftarrow \emptyset$ // initialise optimal pure follower strategies
for $j \leftarrow 1$ ***to*** n *//for each pure follower strategy* **do**
    write linear programme $\mathcal{LP}_j^{\mathcal{G}(A,B)}$;
    call LPsolver();
    get objective value obj_val
    if $obj\_val > max$ **then**
        $max \leftarrow obj\_val$
        $opt \leftarrow \{j\}$
    if $obj\_val = max$ **then**
        $opt \leftarrow obj \cup \{j\}$
$max' \leftarrow \min\{b_{ij} \mid i \leq m, j \leq n\} - 1$
for $j \in opt$ *//for pure follower strategies with IE* **do**
    write linear programme $\mathcal{ELP}_j^{\mathcal{G}(A,B)}$;
    call LPsolver();
    get objective value obj_val and solution σ
    if $obj\_val > max'$ **then**
        $max' \leftarrow obj\_val$
        $IE \leftarrow (\langle\sigma, j\rangle, \beta)$
return IE, max', max, *and* ι

We summarise the results for friendly incentive equilibria (IE) and leader equilibria (LE), respectively, in Table 3.11 for continuous variables in the 0 to 1 range and in Table 3.12 for integer variables in the -10 to 10 range. The respective table shows leader return (avg), follower return (avg), bribery value (avg), confidence interval radius (leader), confidence interval radius (follower), total execution time for 100,000 samples in incentive equilibria (left) and leader equilibria (right). The results indicate that the bribery value falls with the number of leader strategies and, less pronounced, with the number of follower strategies. This is not very surprising: the limit value for infinitely

		Incentive Equilibria						Leader Equilibria				
#FSt	#LSt	Lead R average	Foll R average	Bribery average	conf I leader	conf I follower	ET (minutes)	Lead R average	Foll R average	conf I leader	conf I follower	ET (minutes)
2	2	0.726498	0.621347	0.031252	0.001188	0.001705	4.2175	0.698048	0.612999	0.001332	0.001774	2.7368
2	3	0.798458	0.585959	0.020400	0.000926	0.002014	4.6354	0.785525	0.577285	0.001009	0.002068	3.8177
2	5	0.868985	0.529380	0.010426	0.000642	0.002537	4.5459	0.865601	0.523492	0.000667	0.002555	4.0808
2	10	0.929643	0.434452	0.003310	0.000365	0.003249	6.4351	0.929367	0.432039	0.000367	0.003250	4.5798
3	2	0.752211	0.697569	0.043221	0.001051	0.001526	5.0878	0.710645	0.683939	0.001274	0.001623	4.5323
3	3	0.818569	0.662023	0.027266	0.000809	0.001893	6.5401	0.800711	0.647836	0.000926	0.001970	4.9813
5	3	0.856830	0.558777	0.026226	0.000646	0.003482	11.1242	0.817277	0.725141	0.000843	0.001852	8.0363
10	10	0.967437	0.184819	0.003880	0.000160	0.005209	35.8585	0.954231	0.684151	0.000207	0.003089	34.0453

TABLE 3.11: Values using continuous payoffs in the range 0 to 1.

		Incentive Equilibria						Leader Equilibria				
#FSt	#LSt	Lead R average	Foll R average	Bribery average	conf I leader	conf I follower	ET (minutes)	Lead R average	Foll R average	conf I leader	conf I follower	ET (minutes)
2	2	4.748595	3.030555	0.661109	0.024802	0.030519	3.5780	4.227947	2.982509	0.027691	0.030597	3.5780
2	3	6.272594	2.880576	0.432742	0.019294	0.030389	3.9091	6.041986	2.797888	0.020913	0.030337	3.3044
2	5	8.723898	2.894013	0.217244	0.013183	0.029920	4.1927	7.666887	2.831469	0.013692	0.029740	3.3426
2	10	8.982101	3.039720	0.068128	0.006878	0.029555	4.9849	8.977945	3.013178	0.006930	0.029482	4.1737
3	2	5.295267	4.623187	0.901498	0.021894	0.025617	4.9897	4.552556	4.473518	0.026218	0.025743	3.6521
3	3	6.680038	4.444846	0.575029	0.016841	0.025512	5.3623	6.356434	4.286942	0.019193	0.025418	3.9804
5	3	7.4977325	6.740712	0.548419	0.013251	0.032891	7.6684	6.728477	5.887561	0.017346	0.019969	5.7481
10	10	9.727398	7.82520	0.065963	0.002783	0.046590	37.3025	9.41755	7.587172	0.003934	0.013824	24.9537

TABLE 3.12: Values using integer payoffs in the range -10 to 10.

many strategies is 0, as there is, with limit probability 1, an entry arbitrarily close[4] to the social return 2 in the continuous case, and with the social return 20 in the integer case. For the same reason, it is not surprising that the leader benefits from an increase in her own strategies and in the strategies of the follower alike, whereas the follower does not seem to benefit from an increase in the number of leader strategies.

Table 3.13 compares incentive, leader, and Nash equilibria for leader return and follower return, respectively, for same datasets, where payoff entries are uniformly distributed integer entries in the range from -10 to 10. We have used Gambit [MMT13] to compute Nash equilibria. If there is more than one *NE* in a data-set, we have considered the optimal one with the maximum payoff for the leader, using the follower payoff as a tie-breaker. Our results show that, for both leader and follower, the return is always higher in *IE* as compared to *LE* and even more so when compared to *NE*. Table 3.13 also shows the gain for the leader in *IE* as compared to *NE*. Its value is given by $i\_gain = (LeadRet_{bribery} - LeadRet_{Nash})/(Max_{leadvalue} - LeadRet_{Nash})$ to describe the improvement obtained.

The data set used there is tiny, 10 samples each. This is because it is expensive to compute optimal Nash equilibria. The unsurprisingly large differences to the values from Tables 3.11 and 3.12 confirm that the values have to be read with caution. They suffice to give an impression on the advantage obtained over Nash equilibria as shown in Table 3.13.

The improvement gain, i_gain as shown in Table 3.13, is a measure of how much of the potential improvement gain has been realised. The improvement obtained (numerator) is the difference between the leader payoff in the *IE* and the best *NE*, while the maximum improvement possible (denominator) is the difference between the maximal entry in the payoff matrix of the leader and her payoff in her best *NE*. The value thus norms the leader's gain if she pays bribery to the follower. The higher the value, the more is leader's gain in *IE* by paying bribery as compared to *NE*. An i_gain of value 0 would refer to

[4] ε close for an arbitrarily small, but fixed, $\varepsilon > 0$

		Leader Return				Follower Return		
#FSt	#LSt	Incentive	Leader	Nash	*i_gain*	Incentive	Leader	Nash
2	2	5.96	4.28	1.34	0.56	5.24	4.05	3.53
2	3	6.13	4.78	2.45	0.54	3.77	2.61	1.9
2	5	5.33	5.33	2.81	0.48	2.87	0.94	0.14
2	10	8.81	8.81	7.1	0.66	5.69	5.26	1.89
3	2	4.93	3.43	1.23	0.54	5.31	5.3	5.3
3	3	5.94	4.71	0.67	0.57	5.56	4.3	4.3
5	3	5.95	4.34	1.12	0.52	6.24	6.12	5.2
10	10	9.13	9.01	6.72	0.78	7.27	6.7	1.96

TABLE 3.13: Average leader return and follower return in different equilibria.

no gain, while an *i_gain* of value 1 would refer to freely choosing a strategy profile that does not have to comply with any stability requirement. For the follower, her gain with bribery is also always higher or equal to her gain without bribery, or in *NE*. Thus, the friendly *IE* guarantees a local social optimum in the form of a socially optimal follower return.

Note that the execution time given for each data-size is the total time required for the complete data-set, i.e., 100,000 games. The execution time rises faster with an increasing number of follower strategies than with an increasing number of leader strategies. This was to be expected, as the number of follower strategies determines the number of linear programmes that need to be solved to find a friendly incentive equilibrium. Our expectation for randomly drawn examples was to find roughly a quadratic growth in the number of follower strategies and a linear growth in the number of leader strategies. The actual growth seems to be a bit lower, but this may well be due to noise and random effects. The execution time is tiny in all instances.

Symbolic analysis of relevant classes. As discussed in the Section 3.2, prisoners dilemma / arms race games are one standard class of problems, where our technique provides very nice results. We give a short overview on their symbolic solution. A general bi-matrix game of this class are games in the form of Table 3.14 that satisfy the following constraints:

- $d > b > h > f$ and $e > a > g > c$, and
- $\min\{b-g, a-g\} > \max\{d-b, h-f, e-a, g-c\}$.

		Player II	
		1	2
Player I	1	a, b	c, d
	2	e, f	g, h

TABLE 3.14: "Prisoners dilemma" payoff matrix.

It is easy to see that the only Nash and leader equilibrium is to play (2,2), with leader return g and follower return h. In an incentive equilibrium, however, the leader can incentivise her follower to play 1 when she pledges to play 1 herself and promises to

pay him a bribery of $d - b$ for playing 1. The follower return then increases to d, while the leader return increases to $b + a - d$.

As mentioned in the Section 3.2, the class of battle of sexes games – these are the games satisfying $g > a > \max\{c, e\}$ and $b > h > \max\{d, f\}$ – is a class of games, for which incentive equilibria provide optimal result for the leader, namely the strategy (2,2) with bribery 0. This is, however, also a leader equilibrium.

Finally, when we consider the games where the leader has no choice except for the selection of the bribery value, we note that an incentive equilibrium provides the social optimum, without effecting the outcome for the follower. If, e.g., the individual values for the leader and follower are $N(0, 1)$ normal distributed, then the follower's expected outcome for n columns is the expected maximal value of n independent $N(0, 1)$ distributed variables, v_n. The social outcome for each pair is $N(0, \sqrt{2})$ normal distributed, such that the expected social return increases from v_n to $\sqrt{2} \cdot v_n$, while the expected leader return increases from 0 (as for leader and Nash equilibria) to $(\sqrt{2} - 1) \cdot v_n$.

3.10 Discussion

We have introduced incentive equilibria as a consequential generalisation of leader equilibria. However, the general setting used in leader equilibria is extended by allowing the leader to facilitate her communication with her follower (she needs to communicate at least her strategy in all leader models) to announce an incentive she would pay for a favourable response of her follower. The incentive used there is modelled as a payment that reduces the payoff of the leader and increases the payoff of her follower.

The results are manifold. One outcome is the philosophical discussion on the implications of different assumptions on the behaviour of the follower. We distinguish the classical optimistic assumption that the follower is, ex aequo, friendly to the leader, and the conservative assumption that he is not. We also discussed the implications that these different follower's behaviour have on the the behaviour of the leader.

It is quite interesting to review those differences and their implications. The cost inconsiderate followers incur for the leader is – if any – arbitrarily small, as it suffices to increase the incentive ever so slightly to secure her own return. The only 'cost' for the leader attached to considerate followers is to take their requirements on board as a secondary optimisation criterion.

An interesting observation is that the follower would suffer from being inconsiderate, and not the leader. The occasional increase by an arbitrarily small amount needs to be set against potential losses when the leader has choice, alongside with guaranteed losses when the secure incentive equilibria exist, but are not friendly, as shown in the example below.

In this example, the leader does not need to incentivise her follower. She plays I in the only friendly incentive equilibrium, but II in the only secure incentive equilibrium,

		Follower I	Follower II
Leader	I	$1, 1$	$0, 1$
	II	$1, 0$	$-1, -1$

TABLE 3.15: Loss of inconsiderate follower.

while her follower plays I in both of them. Naturally, the outcome of the leader is unaffected, while the payoff of the follower drops from 1 to 0.

As observed for the prisoners' dilemma (cf. Table 3.2), incentivising seems to be individually better for *both* players, and the follower may benefit more from friendly incentive equilibria than the leader. From the leader's perspective, the unsurprising observation *incentive equilibria* beat *leader equilibria* beat *Nash equilibria* is confirmed. While these (not necessarily strict) advantages of the leader always hold, the case for the follower is slightly weaker, as examples exist where the gain of the leader is accompanied by a loss for her follower. It is good news that our experimental results suggest that the follower would benefit on average from friendly incentive equilibria.

The results are also interesting in that all problems discussed were shown to be tractable. Tractability is arguably a prerequisite for applicability, as the strategies need to be computed to be of use, and the tractability of all occurring problems – calculating incentive, friendly incentive, leader, and friendly leader equilibria as well as deciding the existence of secure incentive equilibria and calculating them (if they exist) or secure ε-incentive equilibria (otherwise) – suggests that their implementation is not hindered by excessive computation costs.

Incentive equilibria therefore provide a natural explanation for bribery in asymmetric non-zero sum bi-matrix games.

Chapter 4

Mean-payoff Games

This chapter is mainly based on the results from [GS14] and [GST+16]. In this chapter, we establish the existence of optimal leader and incentive strategy profiles in multi-player mean-payoff games. Note that we have introduced mean-payoff games in general in Chapter 1 (cf. Section 1.2).

This chapter is organised as follows. We start with the discussion of few motivational examples in Section 4.2.1 to show how multi-player mean-payoff games form a natural choice to study the model with quantitative settings. Section 4.3 formally introduces multi-player mean-payoff games and other terminology used throughout this chapter. Section 4.4 introduces leader equilibria in these games and state important results in this context. We extend the technical details from Section 4.4 to establish the existence of incentive equilibria in Section 4.5. We establish NP-hardness of the related decision problem in Section 4.6. We state implementation details and results obtained from a tool in Section 4.7. We finally summarise this chapter with a discussion in Section 4.8.

4.1 Abstract

We study optimal equilibria in turn based multi-player mean-payoff games under game settings where we have a distinguished player–called the leader–who can assign strategies to all other players, referred to as her followers. Note that Nash equilibria are a standard way to define the rational behaviour of different players in multi-player games that treat all players equally. Since in our settings a leader has additional power over the game to assign strategies to all participating players(including herself), we introduce *leader* and *incentive equilibria* for these game settings. A strategy profile is a *leader strategy profile* if no player, except for the leader, can improve his payoff by changing his strategy unilaterally. An optimal leader strategy profile with the maximal return for the leader is called as *leader equilibria*. We further allow the leader to additionally influence the behaviour of her followers by transferring parts of her payoff to her followers. The ability to incentivise her followers provides the leader with more freedom in selecting strategy profiles, and we show that this can indeed improve the leader's payoff in such games. We call these strategy profiles as *incentive strategy profiles* and optimal strategy profiles as

incentive equilibria. Incentive equilibria are therefore a natural generalisation of leader and Nash equilibria.

We establish the existence of leader and incentive equilibria in multi-player mean-payoff games. We show that the leader always has an optimal strategy in this setting and that no Nash equilibrium can be superior to it. We further show that the decision problem related to constructing incentive equilibria is NP-complete. However, we show that, when the number of players is fixed, the complexity of the problem falls in the same class as two-player mean-payoff games. We discuss algorithm for the computation of these optimal strategy profiles and give our experience with implementing these algorithmic techniques.

4.2 Introduction

This chapter studies the existence of optimal strategy profiles in multi-player mean-payoff games. Mean-payoff games [ZP96, CHJ05b] have been widely studied as a class of perfect information games.

Mean-payoff games are in particular interesting to study because of the diversity of their application fields. Mean-payoff games enjoy a special status in verification, since μ-calculus model checking and parity games can be reduced in polynomial-time to solving mean-payoff games. Mean-payoff objectives also occur when temporal objectives of the players are represented as deterministic Büchi automata (DBA) with their modern quantitative semantics [Hen13], where we are interested in the limit-average share of occurrences of accepting states rather than merely in whether or not infinitely many accepting states occur. Finally, another remarkable application of mean-payoff games is in optimal controller synthesis in the framework of Ramadge-Wonham [RW89, PR89] where the goal of the game is to find a control strategy that maximises the average reward earned during the evolution of the system.

We consider non zero-sum multi-player mean-payoff games here. In these games, a finite number of players control various vertices and move a token along the edges to collectively produce an infinite run. There is a player-specific reward function that, for every edge of the graph, gives an immediate reward to each player. The payoff to a player associated with a play is the limit average of the rewards in the individual moves. Mean-payoff games are positionally determined games, i.e., starting at any vertex, optimal positional strategies exist for both the players. The way each player plays can be captured by a strategy. A set of strategies, one for each player, is called a strategy profile. A strategy profile is in a Nash equilibrium if no player has an incentive for unilateral deviation, i.e., if all other players adhere to their strategy, a player cannot increase her payoff by changing her strategy. Nash equilibria [Nas50, Leh90, Umm08, OR94] are a common way to describe stable strategies with the intuition that only if no player gains from changing her strategy unilaterally, the strategy will be maintained.

However, we consider game settings, where we allow one player to assign strategies to herself and to all other players in the game alike. We refer to this special player as *leader* and the other players who merely follow the strategy profile assigned by the leader as her *followers*. These game settings are then considered as asymmetric as one player is given more power in defining the rules of the game such that she can define the strategy profile. Given that we allow the leader to select the complete strategy profile, another natural question that arises is whether achieving a Nash equilibrium is the right target for her? This more power over the game naturally allows the leader to 'discriminate' against herself. The constraints on the strategies of other players are clearly a pre-requisite for a stable strategy, but not necessarily for her. We therefore allow leader to select strategies that she can improve over. This gives her more leeway when selecting a strategy profile. We show later in section 4.4.1 that the leader may actually suffer from restricting her strategy in the Nash sense.

These strategy profiles are considered asymmetric in that leader may now select a strategy profile where no other player but the leader may have an incentive to deviate. These strategy profiles needs to be *stable* in that no other player would benefit from deviation (as this player would then not follow the suggestion put forward by the leader). Among these stable strategy profiles, leader can then choose an optimal one. In this chapter, we discuss the techniques to compute optimal strategy profiles that provides optimal return to a special player. We study the existence of optimal Nash and leader equilibria in multi-player mean-payoff games. We show the existence of optimal *leader strategy profiles* in these games. The optimal strategy profiles among this class that gives maximal return to the leader is known as *leader equilibria*.

We further show that leader can also use her power to define more stable and more optimal strategy profiles. We therefore extend this setting to the one where we also allow leader to incentivise her followers to follow a particular strategy profile. The leader therefore has even more powerful strategies, where she not only puts forward strategies that describe how the players move, but also gives non-negative incentives to the followers for compliance. These incentives are then added to the overall rewards the respective follower would receive in each move of the play, and deduced from the overall reward of the leader.

Like for leader equilibria, a strategy profile is stable if no *follower* has an incentive to deviate. Similar to leader strategy profiles, incentive strategy profiles are also assigned by the leader, where leader also pays an incentive to her followers. These incentives are then added to the overall payoff of a player in a given play. We call these strategy profiles as *incentive strategy profiles* and the one with optimal return for the leader as an *incentive equilibrium*. An *incentive equilibrium* is a stable strategy profile with maximal reward for the leader. Note that Nash equilibria cannot beat leader equilibria, as the leader can choose from a wider range of strategy profiles. Likewise, leader equilibria cannot beat incentive equilibria, as the leader can, again, choose from a wider range of strategy profiles (leader equilibrium can be viewed as an incentive equilibrium with 0

incentive). In this chapter, we therefore study the potential that a rational leader has if she is allowed to incentivise her followers for a particular strategy profile.

4.2.1 Motivational Examples

We first discuss how multi-player mean-payoff games [ZP96, Jur98, CHJ05b, BV07, Sch08, BEF$^+$11, BCD$^+$11] form a natural choice to study the model with quantitative settings. First, there has been a recent trend to replace traditional model checking by quantitative model checking (see [Hen13] for a recent survey). In traditional model checking, a qualitative property such as 'a system is always eventually granted access to a resource', ($\square\Diamond$access), is checked.

In quantitative model checking, the qualitative measure like whether a DBA would accept an infinite play is replaced by a quantitative measure, where the quality of a path would be measured by the limit average share of accepting states occurring in a run of the DBA. This naturally defines a mean-payoff condition. In qualitative model-checking, however, a DBA would accept an infinite play if the accepting state is visited infinitely many times, and all of these paths would then be of equal quality.

Qualitative Nash equilibria have been used to refine a worst case analysis. In [Hen13], for example, the distributed development of a system is considered, where teams develop components that try to establish individual specifications. This is a symmetric setting where a component can safely be assumed not to be malicious to the extent that, for harming others, it would sacrifice compliance with its main objective. Within this constraint, however, it is conservatively considered to be adversarial. When we view the leader as adversarial, we can use the same techniques to determine how she can coordinate an attack on the system to minimise the payoff.

To reflect the traditional qualitative properties, we refer to the deterministic Büchi automata (DBA) from Figures 4.1 and 4.2. The first automaton, $\mathcal{A}$, shown in Figure 4.1 is in an accepting state whenever the process requests (and is granted) access to a resource, while the second automaton $\mathcal{B}$, shown in Figure 4.2 is in an accepting state whenever the player is granted access. Conditions from Figure 4.1 and Figure 4.2 reflects a two-property automata.

Its quantitative counterpart is to measure the quality of path by the limit average of accepting states occuring in a run of the DBA. Note that it also defines a mean-payoff condition. In this setting, $\mathcal{B}$ refers to the limit average share of the time that a system's critical resource is used, while $\mathcal{A}$ refers to the limit average frequency with which a process asks for (and receives) access on an infinite path.

This inspired us to consider a setting, where different selfish players follow different objectives defined by such DBAs. In our example from Figures 4.3 and 4.4, the environment consists of two selfish players who want to maximise the frequency in which they are granted access to a critical resource (using $\mathcal{A}$ for their respective objective), while the control objective of the system is to maximise the utilisation of the system (using $\mathcal{B}$ for this objective).

In some states of this model the players have choices. In our example, the two players have the choice to make two different kinds of requests, r_a (resp. r'_a for the second player), which shall trigger an access for just one time unit, represented by g_a (resp. g'_a for the second player), or a request r_b (resp. r'_b), which shall trigger an access for three time units, represented by three g_b (resp. g'_b). They can also use a local ε move.

In order to keep the model simple, we focus on models where, on each state, there is one player who resolves the choice. The states in which a player resolves the choice are depicted as squares. Slightly more general, we consider mean-payoff games (MPGs).

Figure 4.5 depicts the multi-player mean-payoff game defined by players from Figures 4.3 and 4.4 with their respective properties. The nodes are labelled by the players who own them, e and e' for the rational environment players, and r for the system player. The payoff is shown in the order payoff for e, e', r.

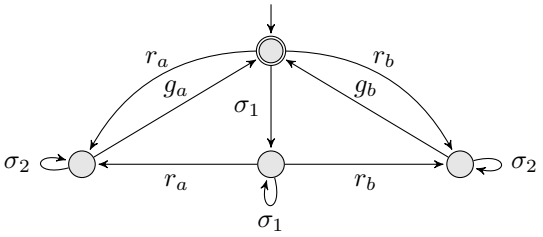

FIGURE 4.1: $\sigma_1 = \neg(r_a \vee r_b)$, $\sigma_2 = \neg(r_a \vee r_b \vee g_a \vee g_b)$.

In mean-payoff games based on quantitative specifications, studying a leader of this type has numerous natural justifications. For example, we might seek optimal control of a system that is used by the external players, who are not under our control, but to whom we can communicate the rules of its use. It is natural to assume that the rules will only be complied with, if the external players have no incentive to deviate, while the controller can take a higher perspective and take the indirect effect (in form of non-compliance by the external players) of her deviation into account when setting the rules. Similarly, an adversarial leader has to take the rationality of the external players involved into account, but she can herself resolve the remaining non-determinism in the system in any way that complies with this constraint.

In our example, the leader will seek to maximise (in the controller model) resp. minimise (in the attacker model) the time any process is using the critical resource, while the individual processes attempt to maximise their own access to it.

In software engineering, the environment is often regarded as antagonistic. This relates to a two-player zero-sum game, where the environment forms a monolithic block

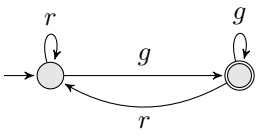

FIGURE 4.2: $g = g_a \vee g_b \vee g'_a \vee g'_b$, $r = r'_a \vee r_b{}' \vee r_a \vee r_b \vee \varepsilon$.

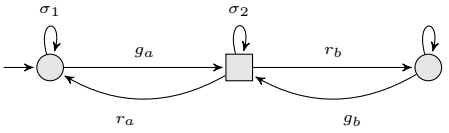

FIGURE 4.3: $\sigma_1 = r'_a \vee r'_b \vee g'_a \vee g'_b \vee \varepsilon$, $\sigma_2 = \sigma_1 \vee g_a \vee g_b$.

whose gains and losses are the losses and gain, respectively, of the system, represented by the leader in our setting.

These settings therefore reflect how an interested party, called a *leader*, can capitalise on setting an equilibrium strategy. The question can therefore be phrased as:

How should a reflective leader control a system if given a chance to do so?

Note that the advantages that the leader can get does not occur in the qualitative setting, where such an equilibrium would also be Nash. We next show that a rational leader can further improve her payoff if along with the power to define strategy profiles, she is also allowed to incentivise her followers. Leader can do so by paying some of her own utility to her followers. On the first glance, it seems not to be in the interest of leader, but careful observation shows that this allows leader to define even more powerful and stable strategy profiles. We give few examples to exemplify the role that incentives can play to achieve good stable solutions of multi-player mean-payoff games. In these examples, incentive equilibria are strictly better than Nash and leader equilibria.

Example 4.1. *Consider the multi-player mean-payoff game shown in Figure 4.6. Here we have three players: Player 1, Player 2 (the leader), and Player 3. The vertex labelled* 1 *is controlled by Player 1, while the vertex labelled* 2 *is controlled by Player* 2*. All other vertices are controlled by Player* 3*. We further annotate the rewards of various players on the edges of the graph by giving a triple, where the reward of the players* 1*,* 2*, and* 3 *are shown in that order. We omit the labels when the rewards of all players are* 0*. An incentive equilibrium would be given by (a strategy profile leading to) the play* $\langle 1, 2, 3^\omega \rangle$*,*

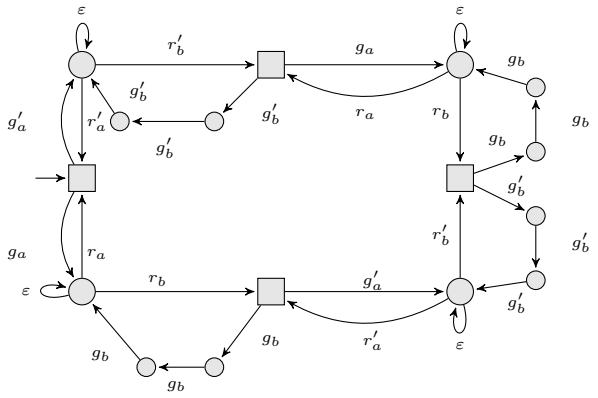

FIGURE 4.4: The rational environments (Figure 4.3) and the system (Figure 4.4), shown as automata that coordinate on joint actions.

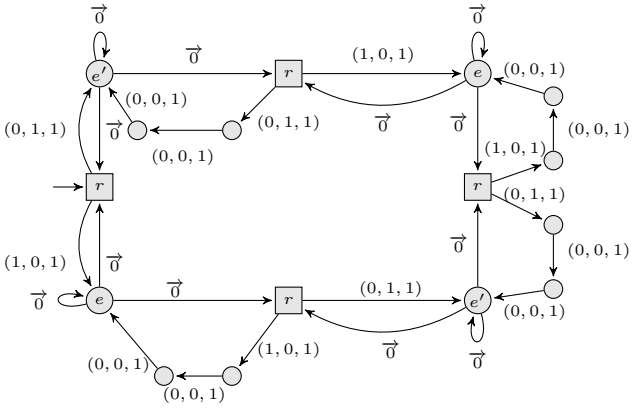

FIGURE 4.5: The multi-player mean-payoff game from the properties from Figures 4.1,4.2 and 4.3,4.4.

where the leader pays an incentive of 1 *to Player* 1 *for each step and* 0 *to Player* 3*. By doing this, she secures a payoff of* 8 *for herself. The reward for the players* 1 *and* 3 *in this incentive equilibrium are each* 1 *and* -9*, respectively. A leader equilibrium would result in the play* $\langle 1, 2, 5^\omega \rangle$ *(with rewards) of* 1 *for Player 1 and the leader and* -2 *for Player 3: when the leader cannot pay any incentive to Player* 1*, then the move from Vertex* 2 *to Vertex* 3 *will not be part of a stable strategy. The only Nash equilibrium in this game would result in the play* $\langle 1, 4^\omega \rangle$ *with the rewards of* 1 *for Player 1,* 0 *for the leader, and* -1 *for Player 3. This example therefore shows how the leader can benefit from her additional choices in leader and incentive equilibria.* □

Example 4.2. *Consider the multi-player mean-payoff game shown in the Figure 4.7 with five players—Player 0 (the leader) and Player 1 to 4 (followers). For* $i \in \{1, 2, 3, 4\}$*, Player* i *controls the vertex labelled* i *in the game and gets a reward of* 1 *whenever token is at vertex* i*. (To keep the rewards on the edges, one could encode this by giving this reward whenever vertex* i *is entered.) Player* 0 *gets a reward of* 1 *in all of these vertices. The payoff of all other players is* 0 *in all other cases. Notice that the only play defined by Nash or leader equilibria in this example is* $\langle (1, 5, 6)^\omega \rangle$*, which gives provides a payoff of* $\frac{1}{3}$ *to Player* 0 *and Player* 1*, and a payoff of* 0 *to all other players. For incentive equilibria, however, the leader can give an incentive of* $\frac{1}{12}$ *to all followers when they follow the play* $\langle (1, 2, 3, 4)^\omega \rangle$*. It is easy to see that such a strategy profile is stable under incentives. The leader will then receive a payoff of* $\frac{2}{3}$*, i.e., her payoff from the cycle,* 1*, minus the incentives given to the other players,* $4 \cdot \frac{1}{12}$*. All other players receive a payoff of* $\frac{1}{3}$*, consisting of the payoff from the cycle,* $\frac{1}{4}$*, plus the incentive they receive from the*

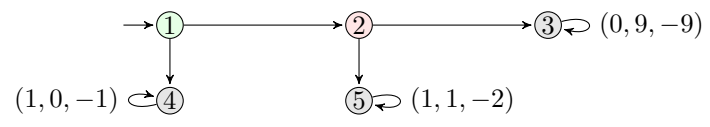

FIGURE 4.6: Incentive equilibrium beats leader equilibrium beats Nash equilibrium.

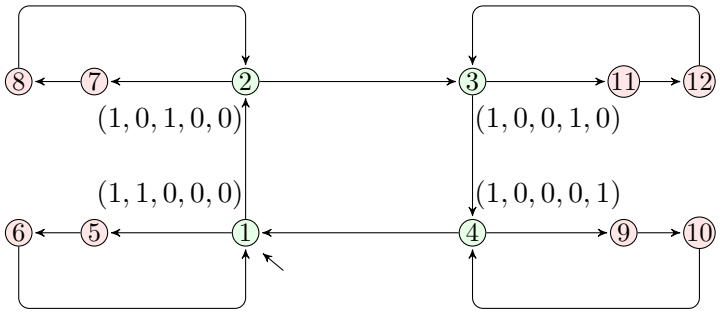

FIGURE 4.7: Incentive equilibrium gives much better system utilisation.

leader, $\frac{1}{12}$*. Notice that this payoff is not only better from the leader's point-of-view, the other players are also better off in this equilibrium.* □

4.2.2 Related Work

The existence of Nash equilibria in multi-player mean-payoff games has been established in [TR97]. Ummels and Wojtczak [UW11] studied the complexity of determining the existence of Nash equilibria, where each reward falls into a given closed interval in multi-player mean-payoff games. Both sides of the NP completeness proofs are closely related to ours. They considered Nash equilibria for mean-payoff games and showed that the decision problem of finding a Nash equilibria is NP-complete for pure (not allowing randomisation) strategy profiles, while the problem is undecidable for arbitrary randomised strategies. The undecidability result of [UW11] for Nash equilibria in arbitrary randomised strategies can be easily extended to leader equilibria. For this reason, we focus on non-randomised strategies throughout this chapter. Ummels [Umm08] analysed the complexity of Nash equilibrium in infinite multi-player games with co-Buchi or parity winning conditions. He established the NP completeness of Nash equilibrium in infinite games with these objectives.

In [Umm06], Ummels has studied the concept of subgame perfect equilibrium for the case of infinite games. He has given simple examples to show that subgame perfect equilibrium, where choice of strategy should be such that it is optimal for initial history of the game and not for just initial vertex, exists in the case of infinite games.

The reward and punish strategy profiles that we study for mean-payoff games are inspired by [UW11, BDS13] and similar strategies in stateless games [Fri77]. Leader equilibria were introduced by von Stackelberg [vS34] and were further studied in [Fri77]. Incentive equilibria have recently been introduced for bi-matrix games [GS15], but have, to the best of our knowledge, not been used in infinite games.

Two-player mean-payoff games can be solved in pseudo-polynomial time [ZP96, BCD$^+$11], smoothed polynomial time [BEF$^+$11], PPAD [EY10], and randomised subexponential [BV07] time. Their decision problem is also known to be in UP∩co-UP [Jur98, ZP96]. Our tool builds on the optimal strategy improvement algorithm for finding the mean-partition from [Sch08]. Their tractability is still an open problem.

4.3 Preliminaries

A *multi-player mean-payoff game* (MMPG) is a tuple $\langle P, V, \{V_p \mid p \in P\}, v_0, E, \{r_p : E \to \mathbb{Q} \mid p \in P\}\rangle$, where

- P is a set of players with a distinguished leader player $l \in P$,
- V is a set of vertices with a designated initial vertex $v_0 \in V$,
- $\{V_p \mid p \in P\}$ is a partition of the vertices V characterising vertices controlled by the respective players,
- $E \subseteq V \times V$ is a set of edges, such that each vertex has a successor ($\forall v \in V \; \exists v' \in V, \; (v, v') \in E$), and
- $\{r_p \mid p \in P\}$ is a family of reward functions $r_p : E \to \mathbb{Q}$, that assign, for each player $p \in P$, a reward to each transition to p.

We use two-player zero-sum mean-payoff games (2MPGs) to determine the outcome of MMPGs when, from some point onwards, one player, say p, is playing against all others, where the objective of p is inherited from the multi-player MPG, while the objective of the remaining players is to minimise her reward. As the objective of the remaining players is defined by the objective of p, we use only r_p to describe the objective of the game. The 2MPG *for* p of an MMPG $\mathcal{M} = \langle P, V, \{V_p \mid p \in P\}, v_0, E, \{r_p : E \to \mathbb{Q} \mid p \in P\}\rangle$, denoted $\mathsf{2mpg}(\mathcal{M}, p)$, is therefore the game $\langle P, V, \{V_p, V \setminus V_p\}, v_0, E, r_p\rangle$. 2MPGs have optimal memoryless strategies for both players, and the outcome is, when starting in any vertex $v \in V$, determined [ZP96]. By abuse of notation, we denote this value for a vertex v by $r_p(v)$.

A finite play $\pi = \langle v_0, v_1, \ldots, v_n\rangle$ of the game $\mathcal{G}$ is a sequence of vertices such that v_0 is the initial vertex, and for every $0 \leq i < n$, we have, $(v_i, v_{i+1}) \in E$. An infinite play is defined in an analogous manner. To select an initial vertex, we can allow probability distribution over the vertices and that we do for discounted sum games in Chapter 5. However, throughout this chapter we select initial vertex with probability 1 and denote it with an incoming arrow. A multi-player mean-payoff game is played on a game arena $\mathcal{G}$ among various players by moving a token along the edges of the arena. The game begins by placing a token on the initial vertex. Each time the token is on a vertex controlled by a player $p \in P$, the player p chooses an outgoing edge and moves the token along this edge. The game continues in this fashion forever, and the players thus construct an infinite play of the game. The (raw) payoff $r_p(\pi)$ of a player $p \in P$ associated with a play $\pi = \langle v_0, v_1, \ldots\rangle$ is the limit average reward of the path, given as

$$r_p(\pi) \stackrel{\text{def}}{=} \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} r_p\big((v_i, v_{i+1})\big).$$

We refer to this value as the raw payoff of the player p to distinguish it from the payoff for the player that also includes the incentive given to the player by the leader.

If the reward functions sum up to 0, i.e., if $\sum_{p \in P} r_p(e) = 0$ holds for all edges $e \in E$, then we call the MMPG a zero-sum game.

Note that we have formally introduced equilibrium concepts in Chapter 2. For all those definitions we refer the reader to Chapter 2. We only define here concepts that have not been introduced before. A strategy of a player is a recipe for the player to choose the successor vertex. It is given as a function $\sigma_p : V^* V_p \to V$ such that $\sigma_p(\pi)$ is defined for a finite play $\langle v_0, \ldots, v_n \rangle$ when $v_n \in V_p$ and it is such that $(v_n, \sigma_p(\pi)) \in E$. A strategy profile σ defines a unique play π_σ, and therefore a raw payoff $r_p(\sigma) = r_p(\pi_\sigma)$ for each player p.

We recall that a leader equilibrium is a maximal leader strategy profile (w.r.t. the raw payoff of the leader). Thus, a leader strategy profile allows for solutions, where the leader could improve upon her reward by changing her strategy. An incentive strategy profile satisfies the stability requirements of the leader equilibria and allows the leader to give incentives to her followers. An optimal strategy profile in this class of strategy profiles that provides maximal reward to the leader as an incentive equilibrium. The way incentives are given to players in an infinite play is as follows.

Definition 4.1 (Incentives and Overall Payoffs)**.** An incentive to a player p is a function $\iota_p{:}V^*V_p{\to}\mathbb{R}_{\geq 0}$ from the set of histories to incentives. Incentives can be extended to infinite play $\pi = \langle v_0, v_1, \ldots \rangle$ in the usual mean-payoff fashion:

$$\iota_p(\pi) \stackrel{\text{def}}{=} \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \iota_p(v_0 \ldots v_{n-1}).$$

The overall payoff $\rho_p(\pi)$ to a follower in run π is the raw payoff plus all incentives, $\rho_p(\pi) \stackrel{\text{def}}{=} r_p(\pi) + \iota_p(\pi)$, while the overall payoff of the leader $\rho_l(\pi)$ is her raw payoff after deducting all incentives, $\rho_l(\pi) \stackrel{\text{def}}{=} r_l(\pi) - \sum_{p \in P \setminus \{l\}} \iota_p(\pi)$.

We write $\bar{\iota}_p(\overline{\sigma})$ for the incentive to player p for the unique run $\pi_{\overline{\sigma}}$ under incentive profile $\bar{\iota}$.

Our observation that incentive strategy profiles form a broader class over leader strategy profiles that in turn form a broader class over Nash strategy profiles together with Example 4.1, where we showed an example of a game where incentive equilibria (IE) are superior over leader equilibria (LE) and LE are superior over Nash equilibria (NE) for the leader, yield the following result.

Theorem 4.2 (IE $\geq$ LE $\geq$ NE)**.** *Incentive equilibria do not provide smaller return than leader equilibria, and leader equilibria do not provide smaller return than Nash equilibria. Moreover, there are games, for which these three values are different.*

Here, we refer to the leader's return (or: payoff) in different equilibria. The above theorem therefore establishes that $r_l(IE) \geq r_l(LE) \geq r_l(NE)$. Revisiting the example from Figure 4.6 illustrates this relation between various sets of strategy profiles. The example established that optimum when selected over smaller sets cannot be better for

the leader when selected over a larger set. For this reason, the leader's return from incentive equilibrium is better than her return from leader equilibrium that is better than her return from a Nash equilibrium.

We now discuss in detail the existence and construction of leader equilibria in Section 4.4 and give an extension of these techniques in Section 4.5 to discuss the construction of incentive equilibria.

4.4 Leader equilibria

A leader strategy profile that provides the maximal reward for the leader among all leader strategy profiles is called a *leader equilibrium*. In the remainder, we show that

1. leader equilibria are generally better (for the leader) than Nash equilibria (Theorem 4.4),

2. determining if there is a strategy profile σ with $r_l(\sigma) = 1$, such that the strategy profile σ is a Nash equilibrium or leader strategy profile is NP-hard even for zero-sum MMPGs with reward functions whose domain is $\{-1, 1\}$ (Theorem 4.29), and the optimal reward of the leader cannot be approximated efficiently (Corollary 4.30), and

3. leader equilibria (and optimal Nash equilibria) always exist, and, for a fixed set of players, finding them in MMPGs is polynomial time reducible to solving two-player MPGs (Corollary 4.17).

For social optima, it suffices to add a social reward to the reward function, e.g., the sum of the individual rewards, without letting the respective player own any vertex. The technique introduced in this chapter can then be used to optimise the social payoff.

We start with a trivial inference of the existence of leader strategy profiles.

Lemma 4.3. *Leader strategy profiles exist for all multi-player mean-payoff games.*

Proof. It is shown in [BDS13] that Nash equilibria for multi-player mean-payoff games exist. By definition, any strategy profile in Nash equilibrium, is also a leader strategy profile. □

What remains to be shown is that optimal leader strategy profiles (i.e., leader equilibria) exist, but that the optimum can be taken is implied by the construction from Section 4.4.2 (cf. Corollary 4.12).

4.4.1 Superiority of leader equilibria

In this subsection, we show that leader equilibria are superior over Nash equilibria: a benign leader who assigns strategies in such a way that she only makes sure that no *other* player has an incentive to deviate, while allowing for the use of 'modest' strategies that she can improve upon when the other players stick to their strategies, is more

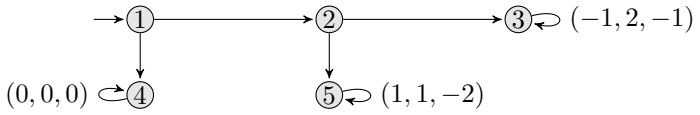

FIGURE 4.8: An MMPG, where the leader equilibrium is strictly better than all Nash equilibria.

successful than a leader who follows the short sighted egoistic approach to chose only among strategies she cannot improve upon herself.

On first glance, it may not seem to be in the interest of the leader to be benign. To the contrary, it would seem that the leader could improve upon such strategy profiles by simply adjusting her strategy. A second look, however, reveals that she has to comply with less constraints and can, consequently, choose from a larger pool of strategy profiles. In particular, this implies that the optimum cannot be worse.

The MMPG from Figure 4.8 exemplifies why this may lead to an increased payoff. It shows a simple MMPG with five vertices, 1 through 5, where the *leader* owns vertex 2, and a player *first* owns the initial vertex, vertex 1. The other vertices have exactly one successor (themselves), such that it does not matter who owns them. We assume that they are assigned to a third player, player *passive*. The rewards are depicted in the order first player, leader, passive player. The rewards on the edges $(1, 2)$, $(1, 4)$, $(2, 3)$, and $(2, 5)$ is not shown, because these edges can only be taken once in a play. Their rewards therefore have no impact on the payoff for any player.

Initially, player *first* can either play to vertex 4, or to vertex 2. When playing to vertex 4, every player will receive a payoff of 0. When he plays to vertex 2, the *leader* can either move on to vertex 3, securing herself a payoff of 2, to the cost of the *first* and the *passive* player, who both receive a payoff of -1. Alternatively, she could play to vertex 5, where both the *leader* and the *first* player receive a payoff of 1, to the cost of the *passive* player, whose payoff is -2. Clearly, the leader has an incentive to move to vertex 3 but she chooses to go to vertex 5 in a leader strategy profile such that it is preferable for the *first* player to move to vertex 2 and not to vertex 4.

The only play produced by a Nash equilibrium in this game is the play $1 \cdot 4^\omega$ and we note that this is the only Nash equilibrium in this example. However, the leader has a better leader equilibrium: she can benignly waive her option to move to vertex 3, and instead move to vertex 5. Then, it becomes preferable for the *first* player to move to vertex 2. This results in an improved reward for the leader.

Theorem 4.4. *Nash equilibria cannot be superior to leader equilibria, and leader equilibria can be strictly better than all Nash equilibria, w.r.t. the leader's payoff.*

4.4.2 Reward and punish strategy profiles for leader equilibria

Let us consider a leader strategy profile σ. We first show that we can obtain a leader strategy profile with a similar payoff for the leader by applying a punishment to the first

player who deviates from σ. The power to define the equilibrium allows the leader to use the power of all remaining players to punish this deviator.

That is, we use a strategy profile where all players co-operate to produce π_σ. Note that, the *leader* solicits co-operation from every player who owns some vertex in the game. Further, the strategy profile σ offers the reward $r_p(\pi_\sigma)$ to a player p, which is at least as good as the reward that player p would have received in the two-player game $\mathsf{2mpg}(\mathcal{M}, p)$ from any vertex in $\mathsf{ver}(\pi_\sigma)$. But, if a player deviates from σ, all other players co-operate to harm this player, throwing their own interests to the wind.

Thus, not complying with the requirement to produce π_σ will lead to a payoff of the deviating player, which equals the payoff of this player in a two-player game that starts at the point of her deviation, i.e., at the vertex owned by her, where she is supposed to play in accordance with σ. We denote by $r_p(v)$ the reward player p would receive from the two-player game that starts at any vertex v. We call any such strategy profile σ a *reward and punish strategy profile* and define it as follows.

Definition 4.5. A strategy profile σ is a reward and punish strategy profile if it offers a reward $r_p(\pi_\sigma)$ to every player p and any deviation from σ by a player p will eventually lead her to get a payoff $r_p(v)$ that is a lower payoff (not necessarily strictly) than $r_p(\pi_\sigma)$.

Note that, for reward and punish strategy profile, π_σ essentially defines σ.

Lemma 4.6. *If σ is a leader strategy profile, then there is a reward and punish strategy profile σ', which is also a leader strategy profile and defines the same play $\pi_\sigma = \pi_{\sigma'}$. If σ is a Nash equilibrium, so is σ'.*

Proof. We first observe that π_σ alone defines the reward of all players for the strategy profile σ and thus, due to $\pi_\sigma = \pi_{\sigma'}$, of σ'.

Let us assume for contradiction that a player $p \in P$ for Nash equilibria resp. $p \in P \setminus \{l\}$ for leader strategy profiles has an incentive to deviate from her strategy in σ'. Then her payoff in σ' will be determined by the result of the two-player zero-sum MPG 'her against the rest' as defined by the reward and punish strategy profiles. Note that the initial play up to this point has no impact on the limit reward.

But she can deviate from her strategy in σ at the same position with at least the same reward, by simply assuming that she plays against all other players in the same game. Consequently, she has an incentive to deviate in the strategy profile σ, too, which contradicts the assumption that σ is a Nash equilibrium resp. leader strategy profile. □

This observation allows us to concentrate on reward and punish strategy profiles only. Let $\mathsf{ver}(\pi)$ be the set of vertices that occur in a play, and let $\mathsf{own}(S) = \{p \in P \mid S \cap V_p \neq \emptyset\}$ be the set of players that own some vertex in S. With these terms, it is simple to characterise reward and punish strategy profiles.

Lemma 4.7. *For an MMPG $\mathcal{M}$, a play π_σ is the outcome of a reward and punish strategy profile σ, which is a Nash equilibrium resp. leader strategy profile, if, and only*

if, for all vertices $v \in \mathsf{ver}(\pi)$ *and all players* $p \in \mathsf{own}(\mathsf{ver}(\pi_\sigma))$ *resp.* $p \in \mathsf{own}(\mathsf{ver}(\pi_\sigma)) \setminus \{l\}$ *that control a vertex that occurs in the play, it holds that* $r_p(\pi_\sigma) \geq r_p(v)$.

Proof. To show the 'if' direction, we assume for contradiction that $r_p(\pi_\sigma) < r_p(v)$ holds for some vertex $v \in \mathsf{ver}(\pi)$, which is owned by p (resp. owned by $p \neq l$). Then player p can improve on her strategy by following her strategy until v is reached, and henceforth follow the strategy from $\mathsf{2mpg}(\mathcal{M}, p)$. As the initial play does not influence the limit inferior, her payoff would be at least $r_p(v)$, which is strictly greater than $r_p(\pi_\sigma)$ (contradiction).

To show the 'only if' direction, we assume for contradiction that $r_p(\pi_\sigma) \geq r_p(v)$ holds, but no reward and punish strategy profile defines π_σ. Assume that player p, deviates in vertex v from π_σ. Then the other players will join to diminish her payoff henceforth. Taking into account that the initial sequence up to this point has no influence on the payoffs, they can follow the optimal strategy of the opponents of p from $\mathsf{2mpg}(\mathcal{M}, p)$, restricting the payoff of player p to $r_p(v)$ (contradiction). □

The above lemma explains the use of reward and punish strategy profiles in forming an optimal play and has a flavour of popular *folk theorem* in infinitely repeated games. The theorem discussed thoroughly in [Fri71] explains that for an infinitely repeated game, a strategy profile σ is in a Nash equilibrium if each player receives strictly more than his minimax payoff (the minimum payoff that the other players can enforce upon him). All players therefore continue to follow σ if no deviation occurs. However, if a player i deviates from σ at some point, all other players attempt to minimise the payoff of the deviating player i such that it is not beneficial for player i to deviate. The reward and punish strategy profiles are inspired from this.

In the next step, we now show that we can determine the existence of a *well behaved* reward and punish strategy profile that satisfies constraints over players reward as seen from above lemma and where every edge (that forms a part of a strongly connected component) has a limit share of the run. A strategy profile is *well behaved* if the ratio in which every edge occurs has a limit, that is, if, for all edges $(s,t) \in E$, there is a $p_{(s,t)} = \lim_{n\to\infty} \frac{\#_n^{(s,t)}(\pi_\sigma)}{n}$, where $\#_n^{(s,t)}(v_0, v_1, v_2 \ldots) = \big|\{i < n \mid (v_i, v_{i+1}) = (s,t)\}\big|$ is the number of edges (s,t) among the first n edges that occur in a play $v_0, v_1, v_2 \ldots$. (This limit does not necessarily exist for general strategy profiles). Note that well behaved reward and punish strategy profiles optimises the leader's reward such that it suffices to refer to this class of strategy profiles only. This is because the supremum of general strategies that are not well behaved cannot be higher than the supremum of well behaved strategy profiles. We now discuss the constraint system needed for well behaved reward and punish strategy profiles in detail.

Linear programs for well behaved reward and punish strategy profiles We give here a constraint system $\mathcal{C}_S^\sigma$ that defines a strategy profile σ to be well behaved and utilises the information on the set S of strongly connected component. The first central observation is that, if we already know

- the set of vertices Q visited in π_σ and
- a (strongly connected) set S of vertices such that $S \subseteq Q$ contains all vertices that are visited infinitely often (and is therefore strongly connected),

then we can infer a constraint system $\mathcal{C}^\sigma_S$ by Lemma 4.7, which characterises necessary and sufficient conditions for a strategy profile σ to be a well behaved reward and punish strategy profile. The constraint system consists of two parts. One part is the ratios, where we use the $p_{(s,t)}$ from above for edges $(s,t) \in E \cap S \times S$, and similarly p_v for the limit ratio of each vertex v in S. The limit ratio p_v of a vertex should be equal to the limit ratios of its incoming edges and equal to the limit ratios of all outgoing edges from that vertex (see below). I.e., we derive information p_v for a vertex v once we have information p_e for all its incoming and outgoing edges. Obviously, the limit ratio p_v of each vertex not in S and of each edge not in $S \times S$ must be 0.

This provides a first part of a constraint system, namely

- the ratio of vertices and edges that are not in S resp. $S \times S$ is 0,
  - $p_v = 0$ for all $v \in V \setminus S$ *and*
  - $p_e = 0$ for all $e \in E \setminus S \times S$,
- the ratio of vertices and edges that are in S resp. $S \times S$ is ≥ 0,
  - $p_v \geq 0$ for all $v \in S$ *and*
  - $p_e \geq 0$ for all $e \in E \cap S \times S$,
- the sum of the ratio of vertices is 1, $\sum_{v \in V} p_v = 1$,
- the ratio of a vertex is the sum of the ratios of its incoming edges and the ratio of a vertex is the sum of the ratios of its outgoing edges,
  - $p_s = \sum_{t.(s,t) \in E} p_{(s,t)}$ for all $s \in S$ *and*
  - $p_t = \sum_{s.(s,t) \in E} p_{(s,t)}$ for all $t \in S$.

Note that for the limit ratio of a vertex, the ratios of its incoming and outgoing edges should be equal. The second part of the constraint system stems from Lemma 4.7. For a well behaved strategy profile σ, $r_p(\pi_\sigma) = \sum_{e \in E} p_e r_p(e)$ is simply the weighted sum of the rewards of the individual edges. This provides us with a constraint

$$\sum_{e \in E} p_e r_p(e) \geq \max_{v \in Q}(r_p(v))$$

for all $p \in \mathsf{own}(Q)$ for Nash equilibria, and for all $p \in \mathsf{own}(Q) \setminus \{l\}$ for a leader strategy profile. Note that we make the use of p_e from above to find the weighted sum of the rewards of the edges. Before we define the objective function, we state a simple corollary from Lemma 4.7.

Corollary 4.8. *Every well behaved reward and punish strategy profile satisfies constraints $\mathcal{C}^\sigma_S$, and every well behaved strategy profile σ, whose play π_σ satisfies constraints $\mathcal{C}^\sigma_S$, defines a reward and punish strategy profile.*

The objective of the leader is obviously to maximise $r_l(\pi_\sigma) = \sum_{e \in E} p_e r_l(e)$. Once we have the objective function, we can define linear programming problem $\mathcal{LP}^\sigma_S$ using constraint system $\mathcal{C}^\sigma_S$ and it is simple to determine a solution in polynomial time [Kar84, Kha79].

The relevant points are first to establish that a well behaved reward and punish strategy profile exists for each such solution, and second, to show that non-well behaved reward and punish strategy profiles cannot be preferable for the leader.

From Q, S, and a solution to the linear programs to a well behaved reward and punish strategy profile We start with the simple case that the vertices and edges with non-0 ratio are strongly connected.

We design π_σ as follows. We first go from the initial vertex v_0 through states in Q to some state in S. (Note that this initial path has no bearing on the limit inferior that defines the payoff of the individual players.)

Once we have reached S, we intuitively keep a list for each vertex in S. In this list, we keep the number of times each outgoing edge with non-0 ratio has been taken. We also apply an arbitrary (but fixed) order on the outgoing edges. Each time we are in this vertex, we choose the first edge (according to this order) that has been taken less often (from this vertex) than $\frac{p_e}{p_v}$, the ratio p_e of the edge divided by the ratio p_v of this vertex, suggests. If no such edge exists, we take the first edge.

Remark 4.9. An implementation of such a list is finite: let r_e be the ratio of an outgoing edge e of a vertex v divided by the ratio of the vertex v it emerges from, and let d be the least common denominator of these ratios for a vertex v. Then we can re-set the counters for the outgoing edges to 0 after d steps.

The result is obviously a well behaved strategy profile and the first part of the constraint system is clearly satisfied. It therefore suffices to convince ourselves that the second part is satisfied as well.

Now assume for contradiction that this is not the case. Let q_v and q_e be the real ratio of the vertices and edges, respectively. Note that our simple rule for the selection of vertices implies that $\frac{p_e}{p_v}$ is correct for all edges $e = (v, v') \in E \cap S \times S$. Then there must be a vertex $v \in S$, which has the highest factor $\frac{q_v}{p_v}$. As it is the highest factor, none of its predecessors in $E \cap S \times S$ can have a higher ratio; consequently, they must have the same ratio. By a simple inductive argument, this expands to the complete strongly connected set of non-0 vertices. As $\sum_{v \in S} p_v = 1 = \sum_{v \in S} q_v$ holds, this implies $p_v = q_v$ for all $v \in S$.

To extend this argument to the general case, we first observe that the non-0 vertices and edges form islands of (maximal) SC parts C_1, through C_k. We use this observation to compose a play as follows.

We start with an initial part, a transfer from v_0 to C_1 as in the simple case. We then continue by playing a C_1^1 part, a transfer, a C_2^1 part, a transfer, ..., a C_k^1 part, transfer C_1^2, and so forth. To achieve a well behaved strategy profile we do the following.

1. We fix the ratio $\sum_i C_1^i : \sum_i C_2^i : \ldots : \sum_i C_k^i$ according to the the sum of the p_v for vertices v in the respective component. This ratio never changes, and it is given by natural numbers $c_1, c_2, \ldots, c_k$, such that $c_1 : c_2 : \ldots : c_k$ satisfies this ratio.

2. We let C_j^i grow slowly with i. We can, for example, use $i \cdot c_j$.

   Note that the transfer part has constant length, bounded by $|S|$. Thus the limit ratio of transfer is 0.

3. We let the transfer to C_{j+1}^i go to the vertex, in which C_j^i was left. Note that the transfer may contain vertices of various components, but as the overall ratio of the transfer is 0, this does not affect the limit probability.

   Consequently, we can use the controller from the simple case of one SCC for the sequence $C_i^1, C_i^2, C_i^3 \ldots$, which only focuses on the relevant part of the i^{th} component.

In effect, we have simple controllers for the individual components, and a single counting controller that manages the transfer between the components.

It is easy to see that the resulting controller inherits the correct ratios from the simple individual controllers. Together with Corollary 4.8 we get:

Theorem 4.10. *If the linear program $\mathcal{LP}_S^\sigma$ for sets Q of reachable states and S of states visited infinitely often has a solution, then there is a well behaved reward and punish strategy profile that meets this solution.*

Finally, we show that non-well behaved reward and punish strategy profiles cannot provide a better solution than the one provided by the previous theorem.

Theorem 4.11. *For given sets Q and S, non-well behaved reward and punish strategy profiles cannot provide better rewards for the leader than the reward r_l for the leader obtained by the well behaved reward and punish strategy profiles described above.*

Proof. We have shown in Lemma 4.7 that there exists a well defined constraint system obeyed by all reward and punish strategy profiles with set Q of reachable states and all $p \in \mathsf{own}(Q)$ for Nash, and for all $p \in \mathsf{own}(Q) \setminus \{l\}$ for the leader strategy profile. Let us assume for contradiction that there is a reward and punish strategy profile σ that defines a play π_σ with a strictly better reward $r_l(\pi_\sigma) = r_l + \varepsilon$ for some $\varepsilon > 0$.

Let k be some position in π_σ such that, for all $i \geq k$, only positions in the infinity set S of π_σ occur. Let π be the tail $v_k v_{k+1} v_{k+2} \ldots$ of π_σ that starts in position k. Obviously $r_p(\pi) = r_p(\pi_\sigma)$ holds for all players $p \in P$.

We observe that, for all $\delta > 0$, there is an $l \in N$ such that, for all $m \geq l$, $\frac{1}{m} \sum_{i=0}^{m-1} r_p\big((v_i, v_{i+1})\big) > r_p(\pi) - \delta$ holds for all $p \in P$, as otherwise the limit inferior property would be violated.

We now fix, for all $a \in \mathbb{N}$, a sequence $\pi_a = v_k v_{k+1} v_{k+2} \dots v_{k+m_a}$, such that $v_{k+m_a+1} = v_k$ and $\frac{1}{m} \sum_{i=0}^{m_a-1} r_p\big((v_i, v_{i+1})\big) > r_p(\pi) - \frac{1}{a}$ holds for all $p \in P$.

Let $\pi_0 = v_0 v_1 \dots v_{k-1}$. We now select $\pi' = \pi_0 {\pi_1}^{b_1} {\pi_2}^{b_2} {\pi_3}^{b_3} \dots$, where the b_i are natural numbers big enough to guarantee that $\frac{b_i \cdot |\pi_i|}{|\pi_{i+1}| + |\pi_0| + \sum_{j=1}^{i} b_j \cdot |\pi_j|} \geq 1 - \frac{1}{i}$ holds.

Letting b_i grow this fast ensures that the payoff, which is at least $r_p(\pi) - \frac{1}{i}$ for all players $p \in P$, dominates till the end of the first iteration[1] of $|\pi_{i+1}|$.

The resulting play belongs to a well behaved (as the limit exists) strategy profile, and can thus be obtained by a well behaved reward and punish strategy profile by Lemma 4.7. It thus provides a solution to the linear program from above, which contradicts our assumption. □

In particular, Theorems 4.10 and 4.11 imply together with Lemma 4.3 the existence of a leader equilibrium.

Corollary 4.12. *Leader equilibria exist for all multi-player mean-payoff games.*

Decision & optimisation procedures The *decision problem* related to the construction of optimal equilibria asks whether or not, for a given threshold r_{thld}, there exists a strategy profile σ, which is a Nash equilibrium resp. leader strategy profile and provides a reward $r_l(\pi_\sigma) \geq r_{\mathsf{thld}}$ for the leader.

In Lemma 4.7 and Theorem 4.11 we have established that it is enough to consider well behaved reward and punish strategy profiles. The relevant behaviour of these strategy profiles is captured by the set of reachable vertices, the set of infinitely visited vertices S, and the ratio of the edges in $E \cap S \times S$.

We use this observation in various algorithms, starting with a non-deterministic one.

Theorem 4.13. *For an MMPG $\mathcal{M}$ and a threshold r_{thld}, the respective decision problem for leader strategy profiles and Nash equilibria is NP-complete, both in the general case and when restricted to zero-sum games with payoffs in $\{-1, 1\}$.*

Proof. We use non-determinism to first guess a set Q of visited vertices, a set S of vertices visited infinitely often and then the linear program defined by them and a solution thereof. Note that the linear program is polynomial in $\mathcal{M}$ and, consequently, has a polynomial solution.

After having a closer look at the sets Q and S, we can check that there is a possible path from the initial vertex to S, that S is strongly connected, that Q and S define the guessed linear program, its constraint system is satisfied by the solution and the reward of the leader is at least the threshold r_{thld} given. All of these tests can obviously be performed in polynomial time.

The respective hardness results are established in Theorem 4.28. □

[1] Including the first iteration of π_{i+1} is a technical necessity, as a complete iteration of π_{i+i} provides better guarantees, but without the inclusion of this guarantee, the π_j's might grow too fast, preventing the existence of a limes.

Although there is no perfectly fitting lemma or theorem for citing in, the inclusion in NP could have been inferred from [UW11], and the techniques used there are quite similar to ours. We re-proved it as we need the intermediate results below.

The hardness result uses a polynomial number of players. This raises a question if the complexity is better for a bounded number of up to k players.

We first assume that we are already provided with solutions to the 2MPGs to $\mathcal{M}$. To devise a decision procedure, we start with a simple observation:

Lemma 4.14. *For a given MMPG $\mathcal{M}$ with k players and n vertices, there are at most $(n+1)^k$ many different thresholds in the related linear programs.*

Proof. For each player p, there is either the threshold $r_p(v)$ for some vertex v of $\mathcal{M}$, or no restriction on the threshold at all in Part II of the constraint system of a linear program. □

Consequently, we only have to consider the most liberal constraint systems.

Lemma 4.15. *For a given MMPG $\mathcal{M}$ with k players and n vertices and a threshold as referred to in the proof of Lemma 4.14, it suffices to refer to up to n first parts of the constraint system of the linear programming problem.*

Proof. For each Part II of the constraint system as referred to in the proof of Lemma 4.14, it is easy to determine the maximal set Q of nodes that can be visited. For this maximal Q, we can determine the strongly connected components $S_1, S_2, \ldots$ of $(V, E) \cap Q$ that are reachable from the initial vertex v_0. Obviously, there are at most n of them.

It is now easy to see that, for all Q', S' that define Part II of the constraint system, Q' is contained in Q and S' is contained in one SCC S_i from above. Now Q and S_i define a more liberal Part I of a constraint system than Q' and S'. Thus, every solution for Q' and S' is a solution for Q and S_i, too. □

Thus, for a given k, there are only polynomially many linear programming problems to consider, and they are easy to construct. Solving linear programming problems requires only polynomial time [Kha79, Kar84]. We thus obtain the following theorem.

Theorem 4.16. *If we are provided with the solutions to the 2MPGs defined by an MMPG with a fixed number k of players, then we can determine an optimal solution in polynomial time.*

Corollary 4.17. *MMPGs with a fixed number of players can be solved in polynomial time by a machine with an oracle for solving two-player zero-sum MPGs. If 2MPGs are solvable in polynomial time, so are MMPGs with a fixed number of players.*

4.4.3 Reduction to two-player mean-payoff games

Thus, finding optimal strategy profiles in MMPGs with a fixed number of players can be derived from solutions to 2MPGs. Various works have been published on solving 2MPGs.

In [BCD+11], the authors give an improved pseudopolynomial procedure to solve two-player mean-payoff games. [BV07] provides a randomised strongly subexponential and pseudopolynomial algorithm, [EY10] show the complexity of solving 2MPG in PPAD, and [Jur98, ZP96] contain an UP∩CoUP algorithm for the respective decision problems. There are wilder reductions like one to symbolic linear programming [Sch09] and a smoothed polynomial time complexity [BEF+11].

Corollary 4.17 therefore provides the following:

Corollary 4.18. *MMPGs with a fixed number of players can be solved in UP∩coUP, in pseudo polynomial time, in smoothed polynomial time, in PPAD, and in randomised subexponential time.*

4.5 Incentive Equilibria

We are now ready to extend the technical details from Section 4.4 to establish the existence of incentive equilibria in multi-player mean-payoff games. We first introduce a canonical class of incentive strategy profiles–the *perfectly incentivised strategy profiles* (PISPs)–that corresponds to the Stackelberg version of the classic subgame perfection. Note that not all PISPs are valid incentive strategy profiles (ISPs). On the other hand, we show that every ISP has a corresponding PISP (which is also an ISP) with the same leader reward, such that it suffices to consider this class to construct incentive equilibria.

In the following subsection, we show that, for PISPs that are ISPs, it suffices to find a maximum in a *well behaved* class of strategy profiles: strategy profiles where every edge has a limit share of the run—by showing that the supremum of general strategies cannot be higher than the supremum of these well behaved ones.

We then show how to construct well behaved PISPs that are ISPs based on a family of constraint systems that depend on the occuring and recurring vertices on the play. At the same time, we show that no general ISP that defines a play with this set of occuring and recurrent vertices can have a higher value. The set of occuring and recurrent vertices can be guessed and the respective constraint system can be build and solved in polynomial time, which also provides inclusion of the related decision problem in NP.

Finally we use another stability criterion of our canonical form—they are secure in each subgame after deviation—to turn IEs into secure and subgame-perfect ε IEs.

4.5.1 Canonical incentive equilibria

We define a canonical form of an incentive equilibrium with this play that we call *perfectly incentivised strategy profiles* (PISP). In a PISP, a deviator (a deviating follower) is punished, and the leader incentivises all other followers to collude against the deviator. While the larger set of strategies and plays that define them (when compared to Nash and leader equilibria) lead to a better value, this incentive scheme leads to a higher stability: the games are subgame perfect relative to the leader. Before defining subgame perfect incentive strategy profile, we first define a reachable subgame for a play π that

starts at initial vertex v_0 with history hv that contains reachable vertices as a subgame that starts at vertex v and defines the same play.

Definition 4.19 (Subgame Perfect)**.** A strategy profile $(\overline{\sigma}, \bar{\iota})$ is a subgame perfect incentive strategy profile, if every reachable subgame is also an incentive strategy profile.

This term adjusts the classic notion of subgame perfect equilibria to our setting. Subgame perfection refers to believable threats: broadly speaking, when a player threatens to play an action that harms herself, then it may happen that the other players do not believe this player and therefore deviate. In a subgame perfect Nash equilibrium, it is therefore required that the subgame started on each history also forms a Nash equilibrium.

Note that the leader is allowed to benefit from deviation in our setting.

The means to obtain subgame perfection after deviation is to make all players harm the most recent deviator. Thus, we essentially resort to a two-player game. For a multi-player mean-payoff game $\mathcal{G}$, we define, for each follower p, the two-player mean-payoff game where p keeps his reward function, while all other players have the same antagonistic reward $-r_p$. Two-player mean-payoff games are memoryless determined, such that every vertex v has a value, which we denote by $r_p(v)$. This value clearly defines a minimal payoff of a follower: when he passes by a vertex in a play, then he cannot expect an outcome below $r_p(v)$, as he would otherwise deviate.

PISP strategy profiles are in the tradition of reward and punish strategy profiles (cf. Section 4.4.2). In any 'reward and punish' strategy profile, the leader facilitates the power of all remaining followers to punish a deviator. If a player p chooses to deviate from the strategy profile at history h, the game would turn into a two-player game, where all the other followers and the leader forsake their own interests, and jointly try to 'punish' p. That is, player p may still try to maximise his reward and his objective remains exactly the same, but the objective of rest of the players have changed to minus the reward of player p. As they form a coalition with the joint objective to harm p, this is an ordinary two-player mean-payoff game that starts at the vertex $\mathsf{last}(h)$.

For a strategy profile $\overline{\sigma}$ and a history h, we call h a deviating history, if it is not a prefix of $\pi_{\overline{\sigma}}$. We denote by $\mathsf{dev}(h, \overline{\sigma})$ the last player p, who has deviated from his or her strategy $\overline{\sigma}_p$ on a deviating history h.

Definition 4.20 (Perfectly Incentivised Strategy Profile)**.** A perfectly incentivised strategy profile is defined as a strategy profile (PISP) $(\overline{\sigma}, \bar{\iota})$ with the following properties. For all prefixes h and h' of π_σ and for all followers p, it holds that $\bar{\iota}_p(h) = \bar{\iota}_p(h')$. We also refer to this value by $\bar{\iota}_p$. For deviator histories h', the incentive $\bar{\iota}_p(h')$ is 0 except for the cases explicitly referred to below.

On every deviating history h with deviating player $p = \mathsf{dev}(h, \overline{\sigma})$, the player p' who owns the vertex $v = \mathsf{last}(h)$ follows the strategy from the 2MPG $\mathcal{G}_p$. If, under this strategy, player p' selects the successor v' at a vertex v in the 2MPG $\mathcal{G}_p$ (and thus $\overline{\sigma}_{p'}(h) = v'$), p' is a follower, *and* $p' \neq p$, then player p' receives an incentive, such that $r_{p'}(v, v') + \bar{\iota}_{p'}(h \cdot v') = r_{\max} + 1$.

Intuitively, these are the special kind of 'incentive strategy profiles' that are subgame perfect. PISPs are perfectly incentivised and corresponds to reward and punish strategies and are subgame perfect. A PISP where no follower has deviated is subgame perfect as it gives to every follower atleast the maximal value he would receive from any 2MPG (because if not the follower would deviate). Now for the case where a follower has deviated, the leader can make strategy profile subgame perfect as follows. She would award an incentive to rest all players she wants to incentivise at a moment to an amount, such that they receive their maximal reward from 2MPG (plus one) in each step. Therefore only PISPs and not all incentive strategy profiles are subgame perfect. All PISP are however not neceessarily ISP as every PISP does not guarantee an optimal play.

Note that, technically, the leader punishes herself in the definition of PISP. This is only to keep definitions simple; she is allowed to have an incentive to deviate, and the subgame perfection does not impose a criterion upon her. Note also that a PISP is not necessarily an incentive strategy profile, as it does not guarantee anything about $\pi_{\overline{\sigma}}$. A PISP satisfies the constraints of reward and punish strategy profiles and has a corresponding ISP such that it suffices to consider PISPs only (cf. Theorem 4.21). PISPs essentially defines the constant value (the incentive and the reward) that the leader should pay to her followers for it to be an ISPs. The deviating histories of a PISP (cf. Lemma 4.22) are therefore not part of an incentive strategy profile as it does not guarantee that the leader would receive maximal reward. To summarise, a PISP is subgame perfect (relative to the leader) and is an ISP if it results in an optimal play and is well behaved.

We now give the following theorem that states the importance of PISPs in constructing an incentive equilibrium.

Theorem 4.21. *Let $(\overline{\sigma}, \bar{\iota})$ be an ISP that defines a play $\pi_{\overline{\sigma}}$. Then we can define a PISP $(\overline{\sigma}, \bar{\iota})$, which is also an ISP, with the same reward that defines the same play.*

The proof of this theorem follows from Lemma 4.22 and Lemma 4.23.

Lemma 4.22. *Let $(\overline{\sigma}', \bar{\iota}')$ be a strategy profile that defines a play $\pi_{\overline{\sigma}'}$, which contains precisely the vertices Q. Let $(\overline{\sigma}', \bar{\iota}')$ satisfy that, for all followers p and all vertices $v \in Q$ owned by p, $\bar{\iota}_p(\overline{\sigma}') + r_p(\overline{\sigma}') \geq r_p(v)$. Then we can define a PISP $(\overline{\sigma}, \bar{\iota})$ with the same reward, which defines the same play.*

Proof. We note that a PISP $(\overline{\sigma}, \bar{\iota})$ is fully defined by the play $\pi_{\overline{\sigma}}$ and the $\bar{\iota}$ restricted to the prefixes of $\pi_{\overline{\sigma}}$. We now define the PISP $(\overline{\sigma}, \bar{\iota})$ with the following property: $\pi_{\overline{\sigma}} = \pi_{\overline{\sigma}'}$, that is the play of the PISP equals the play defined by the ISP we started with. For all followers p and all prefixes h of $\pi_{\overline{\sigma}}$, we have $\bar{\iota}_p(h) = \bar{\iota}'_p(\overline{\sigma})$. It is obvious that $(\overline{\sigma}', \bar{\iota}')$ and $(\overline{\sigma}, \bar{\iota})$ yield the same reward for all followers and the same reward for the leader. We now assume for contradiction that the resulting PISP is not an incentive strategy profile. If this is the case, then a follower p must benefit from deviation at some history h. Let us start with the case that h is a deviator history. In this case, the reward for

p upon not deviating is $r_{\max} + 1$, while it is the outcome of some game upon deviation, which is clearly bounded by $r_{\max}$.

We now turn to the case that h is not a deviator history, and therefore a prefix of $\pi_{\overline{\sigma}}$. Let p be the owner of $v = \mathsf{last}(h)$. If p is the leader, we have nothing to show. If p is a follower and does not have an incentive to deviate in $(\overline{\sigma}, \bar{\iota})$, we have nothing to show. If p is a follower and has an incentive to deviate in $(\overline{\sigma}, \bar{\iota})$, we note that his payoff after deviation would be bounded from above by $r_p(v)$. Thus, he does not have an incentive to deviate (contradiction). □

Lemma 4.23. *Let $(\overline{\sigma}, \bar{\iota})$ be an ISP that defines a play $\pi_{\overline{\sigma}}$, which contains precisely the vertices Q. Then, for all followers p and all vertices $v \in Q$ owned by p, $\bar{\iota}_p(\overline{\sigma}') + r_p(\overline{\sigma}') \geq r_p(v)$ holds.*

Proof. Assume that this is not the case for a follower p and a vertex $v \in Q$ owned by p. Then p would benefit upon deviating when visiting v. □

4.5.2 Existence and construction of incentive equilibria

A second observation is that we can resort to a simple type of strategy profiles which are *well behaved* in that each edge occurs with a limit probability (the limes inferior and superior of the share of its occurence on $\pi_{\overline{\sigma}}$ are equal). This is similar to the case of leader equilibiria (cf. Section 4.4.2). We first discuss how to find optimal ISPs among well behaved PISPs, and then show that no superior ISPs exist.

Linear programs for well behaved PISPs that are also ISPs. The first part of the constraint system is similar to the constraints given in Section 4.4.2 (cf. Constraint system $\mathcal{C}^{\sigma}_{\mathcal{S}}$), i.e., for every vertex and every edge that occurs in $\pi_{\overline{\sigma}}$, we put a constraint that their vertex and edge ratio has a limit. The first part of the constraint system for strategy profile $\pi_{\overline{\sigma}}$ is thus akin to first part of the constraint system for strategy profile π_{σ} from $\mathcal{C}^{\sigma}_{\mathcal{S}}$.

The second part of the constraint system gives constraints over the reward functions of the players, i.e., for the strategy profile $\pi_{\overline{\sigma}}$ and for all followers p, we now use raw reward plus incentives given to each follower. The second part of the constraint system therefore stems from the proof of Lemma 4.22: the reward for every follower p given by $r_p(\overline{\sigma})$ is simply $\sum_{e \in E} p_e r_p(e)$ plus $\bar{\iota}_p$, that is, it is the weighted sum of the raw rewards of the individual edges plus the incentive given to the followers over that edge. We therefore get the following constraints:

$$\bar{\iota}_p + \sum_{e \in E} p_e r_p(e) \geq \max_{v \in Q}(r_p(v))$$

for all followers p for an ISP. Before we define the objective function, we state a simple corollary from the proof of Lemma 4.22.

Corollary 4.24. *Every well behaved PISP that is an ISP satisfies these constraints, and every well behaved strategy profile $(\overline{\sigma}, \overline{\iota})$, whose play $\pi_{\overline{\sigma}}$ satisfies these constraints, defines a PISP, which is then an ISP.*

Note that the resulting PISP is an ISP even if $(\overline{\sigma}, \overline{\iota})$ is not. This is because the satisfaction of the constraints are enough for the final contradiction in the proof of Lemma 4.22.

The objective of the leader is obviously to maximise $r_l(\overline{\sigma}) - \sum_{p \in P \setminus \{l\}} \overline{\iota}_p = \sum_{e \in E} p_e r_l(e) - \sum_{p \in P \setminus \{l\}} \overline{\iota}_p$. Once we have this linear programming problem, it is simple to determine a solution in polynomial time [Kar84, Kha79]. Using the similar description from Section 4.4.2 (cf. Constraint system $\mathcal{C}^{\sigma}_{S}$), we first observe that it is standard to construct a play defining a PISP from a solution.

Another key observation is that, if the linear program detailed above for sets Q of reachable states and S of states visited infinitely often has a solution, then there is a well behaved reward and punish strategy profile that meets this solution.

Theorem 4.25. *Non-well behaved PISPs that are also ISPs cannot provide better rewards for the leader than those from well behaved PISPs that are also ISPs.*

Proof. The proof is closely related to the proof of the Theorem 4.11. Corollary 4.24 shows that there exists a well defined constraint system obeyed by all well behaved PISPs that are also ISPs with a set Q of reachable states and a set S of recurrent states.

Let us assume for contradiction that there is a reward and punish strategy profile $(\overline{\sigma}, \overline{\iota})$ that defines a play $\pi_{\overline{\sigma}}$ with the same sets Q and S of reachable and recurrent states, respectively, that provides a strictly better reward $r_l(\overline{\sigma}) - \sum_{p \in P \setminus \{l\}} \overline{\iota}_p$, which exceeds the maximal reward obtained by the leader in well behaved PISPs that are also ISPs by some $\varepsilon > 0$.

We now construct a well behaved PISPs that are also ISPs and that also provides a better return. First, we take a $\overline{\iota}'$ with $\overline{\iota}_p = \overline{\iota}'_p$ for all followers p. This allows us to focus on the raw rewards only.

Let k be some position in $\pi_{\overline{\sigma}}$ such that, for all $i \geq k$, only positions in the infinity set S of $\pi_{\overline{\sigma}}$ occur. Let π be the tail $v_k v_{k+1} v_{k+2} \dots$ of $\pi_{\overline{\sigma}}$ that starts in position k. Obviously $r_p(\pi) = r_p(\overline{\sigma})$ holds for all players $p \in P$.

We observe that, for all $\delta > 0$, there is an $l \in N$ such that, for all $m \geq l$, $\frac{1}{m} \sum_{i=0}^{m-1} r_p\big((v_i, v_{i+1})\big) > r_p(\pi) - \delta$ holds for all $p \in P$, as otherwise the limes inferior property would be violated.

We now fix, for all $a \in \mathbb{N}$, a sequence $\pi_a = v_k v_{k+1} v_{k+2} \dots v_{k+m_a}$, such that $v_{k+m_a+1} = v_k$ and $\frac{1}{m} \sum_{i=0}^{m_a-1} r_p\big((v_i, v_{i+1})\big) > r_p(\pi) - \frac{1}{a}$ holds for all $p \in P$.

Let $\pi_0 = v_0 v_1 \dots v_{k-1}$. We now select $\pi' = \pi_0 {\pi_1}^{b_1} {\pi_2}^{b_2} {\pi_3}^{b_3} \dots$, where the b_i are natural numbers big enough to guarantee that $\frac{b_i \cdot |\pi_i|}{|\pi_{i+1}| + |\pi_0| + \sum_{j=1}^{i} b_j \cdot |\pi_j|} \geq 1 - \frac{1}{i}$ holds.

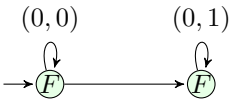

FIGURE 4.9: Secure equilibria.

Letting b_i grow this fast ensures that the payoff, which is at least $r_p(\pi) - \frac{1}{i}$ for all players $p \in P$, dominates till the end of the first iteration[2] of $|\pi_{i+1}|$.

The resulting play belongs to a well behaved (as the limit exists) strategy profile, and can thus be obtained by a well behaved PISP by Corollary 4.8. It thus provides a solution to the linear program from above, which contradicts our assumption. □

Consequently, it suffices to guess the optimal sets Q of vertices that occur and S of vertices that occur infinitely often to obtain a constraint system that describes an incentive equilibrium, which is well behaved and a PISP—and therefore subgame perfect.

Corollary 4.26. *The decision problems 'is there a (subgame perfect) incentive equilibrium with leader reward $\geq r$' is in NP, and the answer to these two questions is the same.*

Recall that a (subgame perfect) incentive equilibrium refers to those strategy profiles that are subgame perfect relative to the leader (this criterion is not for the leader as she is allowed to deviate).

Note that, if we have a fixed number of players, the number of possible constraint systems is polynomial. Like for leader equilibria, there are only polynomially many (for n vertices and k followers $O(n^k)$ many) second parts (the constraints on the follower rewards) of the constraint systems. For them, it suffices to consider the most liberal sets Q (which is unique) and S (the SCCs in the game restricted to Q, at most n). For a fixed number of players, finding incentive equilibiria is therefore in the same class as solving 2MPGs.

4.5.3 Secure ε incentive strategy profiles

We take a short detour to another class of equilibria that make a solution stable: secure equilibria. Secure equilibria [CHJ06] have been defined as Nash equilibria with the additional property that each player would, upon unilateral deviation, either lose strictly, or no other player would lose. Naturally, we have to adjust this definition appropriately. We say that a strategy profile $(\overline{\sigma}, \bar{\iota})$ is a secure incentive strategy profile, if, upon unilateral deviation, every follower either receives a strictly lower reward, or an equal reward. In the latter case, all other players have to receive at least the same reward as before.

We show that we can obtain subgame perfect secure ε incentive equilibria (i.e., every subgame is a secure incentive strategy profile) by simply increasing the individual

[2]Including the first iteration of π_{i+1} is a technical necessity, as a complete iteration of π_{i+i} provides better guarantees, but without the inclusion of this guarantee, the π_j's might grow too fast, preventing the existence of a limes.

incentives from the strategy we have constructed by $\frac{\varepsilon}{|P|}$. The payoff between secure ε incentive equilibria and general incentive equilibria is therefore arbitrarily small. This is in contrast to leader and Nash equilibria, where security can come to a high cost. In the simple example from Figure 4.9 (rewards are shown in the order (follower, leader)), where the left vertex is owned by the follower, the leader can incentivise the follower to move to the right vertex by an arbitrarily small incentive ε, resulting in a secure incentive strategy profile and payoffs of $1 - \varepsilon$ and ε and for the leader and her follower, respectively. A secure leader (and Nash) equlibrium would require the follower to stay forever in the left vertex, resulting in a payoff of 0 for the leader and her follower alike. This is in contrast to 'normal' leader (or Nash) equilibria, which would allow for the follower moving the token to the right, resulting in a payoff of 1 and 0 for the leader and her follower, respectively.

Theorem 4.27. *We can obtain a secure subgame perfect ε incentive equilibrium $(\overline{\sigma}, \overline{\iota})$.*

Proof. Using Theorem 4.25, we can produce a well behaved incentive equilibrium, which is also a PISP. Re-visiting the proof of Lemma 4.22, this PISP satisfies the requirements of a secure incentive equilibrium in every subgame that starts in a deviating history.

For non-deviating histories, however, we have that no follower benefits from deviation, but will normally lack the security property (e.g., in the example from Figure 4.6). We now produce a new PISP $(\overline{\sigma}, \overline{\iota}')$, such that $\overline{\iota}'$ is obtained from $\overline{\iota}$ by selecting $\overline{\iota}'_p = \overline{\iota}_p + \frac{\varepsilon}{|P|}$ for all followers. The subgames that start in a deviating history are not affected by this change, such that the resulting PISP also satisfies the requirements of a secure incentive equilibrium from these positions. For non-deviating histories, however, we have now increased the value for following slightly, such that the pre-requisite of secure equilibria is satisfied here, too. (The deviating follower would strictly decrease his reward.) □

4.6 NP-hardness

In this section, we establish that the related decision problem whether there exist 'a leader resp. incentive equilibrium with payoff of the leader greater or equal to a threshold' is NP-complete.

Theorem 4.28. *The problem of deciding whether a leader resp. incentive equilibrium σ with reward $r_l(\sigma) \geq 1$ resp. $r_l(\sigma) \geq 1 - \frac{1}{2n}$ of the leader exists in games with rewards in $\{0, 1\}$, is NP-complete.*

Proof. The proof is closely related to the NP-hardness proof from [UW11]. We consider reduction of the 3SAT satisfiability formula over n atomic propositions with m conjuncts to solve a MMPG, in order to establish NP-hardness. We assume the game graph has $2n + 1$ players and $5m + 4n + 2$ vertices with payoffs 0 and 1 only. The $2n + 1$ players consist of $2n$ players for the $2n$ literals corresponding to the n variables, and the *leader* who intuitively tries to validate the formula. The number of variables n therefore relate

to the number of players (that are $2n + 1$). For reduction, we consider an example of a 3SAT formula $C_1 \wedge C_2 \wedge C_3$ with $C_1 = p \vee q \vee \neg r$, $C_2 = p \vee \neg q \vee \neg r$, and $C_3 = \neg p \vee q \vee r$. We have $2n$ players for the $2n$ literals that corresponds to the n variables and there is one leader player. The game consists of three phases – an initial assignment phase in which leader would make either truth or false assignment to the literal players. Then, we have a validation phase, in which leader intuitively tries to validate the 3SAT formula as per the assignment done in assignment phase. In the third evaluation phase, payoffs are rewarded to every player including leader. Assignment phase would consist of $2n$ literal vertices and m leader vertices as leader chooses literals corresponding to n variables for the formula assignment. In Validation phase, there are 3 vertices in every conjunct of the 3SAT formula, i.e., $3m$ vertices. In the evaluation phase, we again have $2n$ literal vertices and a leader vertex where game goes round in a cycle of length n. At every point in this cycle, payoffs are given to the players and the leader. There is additionally one sink vertex in the game graph, that has only one outgoing edge to itself. The sink vertex has a payoff of 1 for all literal players but a payoff of 0 for the leader. Game is terminated at sink vertex.

It is in the evaluation phase that by choosing the payoff for the players, it can be decided whether there exists a leader resp. incentive equilibrium in the game with payoff of ≥ 1 for the leader.

Starting in the assignment phase, for each of the m conjuncts, there are m leader vertices. For each conjunct, leader would select a literal vertex for each variable. Like, for a variable $'Z'$, leader would either choose an assignment $'z'$ or $'\neg z'$. The selected literal vertex then has two choices – to continue the game by going to the next leader vertex or to terminate the game by choosing to go to sink vertex. The sink vertex has only one self loop that has only one outgoing edge with payoff of 1 for all literal players and a payoff of 0 for the leader. If literal vertex chooses to go to next leader vertex, leader would go with further assignments in the assignment phase.

In the validation phase, leader intuitively tries to validate whether the 3SAT formula is satisfiable or not according to the chosen assignments in the assignment phase. For each conjunct and for each variable $'Z'$, leader either goes to $'z'$ where literal $'\neg z'$ receives a payoff of 0 and every other player and leader would receive a payoff of 1. Here, also, at every literal vertex, player may opt to continue the game by going to the next leader vertex or to terminate the game by going to the sink vertex. In our example formula, the formula is satisfiable if leader in the assignment phase would select literals p, $\neg q$ and r. The leader resp. incentive equilibrium would be a path $(p, \neg q, r)^\omega$ with an optimal leader return 1.

If the formula is not satisfiable, any run might have to path by both $'z'$ and $'\neg z'$ for a conjunct and a literal player at any point who receives a payoff of 0 might terminate the game by going to sink vertex – that results in leader receiving a payoff of 0. Thus, for all strategy profiles that end up in the sink vertex and for the unsatisfiable formulas, a leader resp. incentive equilibrium has a payoff of 0 for the leader. In the validation

phase, if a literal player deviates to sink vertex, all other player receive a payoff of 1. Leader, therefore, incentivise all remaining players to form a coalition and act against the deviating player. Leader can achieve this by promising to pay a small incentive $1 - \frac{1}{2n}$ to every other non-deviating follower in the game.

While, if the formula is satisfiable, game further goes to the evaluation phase, where nodes are owned by the leader. Here, for a variable $'Z'$, leader either moves to $'z'$, where a payoff of 1 is given to every player but 0 is given to $'\neg z'$or leader goes to $'\neg z'$ where a payoff of 1 is given to $'\neg z'$ and all other players, while a payoff of 0 is given to $'z'$. Additionally, an incentive of $\frac{1}{|2n|}$ is given to all other players. The leader's payoff in any leader resp. incentive equilibrium, is therefore, greater or equal to 1 resp. $1 - \frac{1}{2n}$. The leader return is therefore better than the payoff of 0 at the sink vertex for the leader.

The proof is now complete. □

A simple (polynomial-time) reduction from SAT to establish NP-hardness of Nash equilibrium in a game with qualitative objectives where payoffs are just 0 and 1 and each player either lose or win the game is given by Ummels [Umm05]. Their settings are simple as they allow only one player to make non-trivial choices and their reduction establish that there exist a Nash (subgame perfect) equilibrium where all players win. To obtain NP-hardness for leader resp. incentive equilibrium from there ([Umm05]), we first assign edge weights for each player. Player 0 (we now assume her as the *leader*) control the intermediate vertices and for each clause there are literal vertices X_i and $X_{\neg i}$ that belongs to player i. The vertex m after m clauses now contains an edge back to the start vertex. The game arena is now extended by adding a sink vertex that contains an edge to itself and a player i could terminate the game from some literal vertex X_i or $X_{\neg i}$ owned by him by taking an edge to the sink vertex.

From the intermediate leader vertices the leader can move to X_i or $X_{\neg i}$ that are owned by player i. At X_i player i receives a reward of 1 if it occurs in the formula and all other players (including the leader) receive 0. At $X_{\neg i}$ leader receives a reward of 1 while all other players (including player i) receive 0. At the sink vertex every literal player would receive a payoff of 1 while the leader receives a payoff of 0. The leader would receive a payoff of 1 if the formula is satisfiable. Clearly, leader can construct an equilibrium path if a satisfiable formula exists as the leader at the intermediate vertices would only take an edge to the literal vertex that occurs in the formula such that no literal player has an incentive to deviate to the sink vertex. Thus a satisfiable formula here implies the existence of an incentive resp. leader equilibrium with payoff ≥ 1 for the leader.

Zero-sum games To progress from proof of Theorem 4.28 to zero sum games, we can simply replace the rewards of 0 to -1 and add $(n-1)$ additional players who own no vertex and always have a reward of -1. The result that these games cannot be approximated clearly carries over.

Theorem 4.29. *The decision problem of whether or not a leader resp. incentive equilibrium σ with reward $r_l(\sigma) \geq 1$ of the leader exists in games with rewards in $\{0,1\}$ resp. zero sum games with rewards in $\{-1,1\}$, such that the reward of the leader is always in $\{0,1\}$ resp. $\{-1,1\}$ is NP-complete.*

The NP-hardness proof from Theorem 4.28 together with the Corollary 4.26 that gives its inclusion in NP, gives us the NP-completeness.

Corollary 4.30. *Unless P=NP, no tractable algorithm can approximate the optimal reward of the leader closer than* 0.5.

4.7 Implementation

In this section, we give details about the algorithmic techniques that we used further in the implementation of a tool to evaluate multi-player mean-payoff games (MMPGs). We have implemented a tool [mpg] in C++ to evaluate the performance of the proposed algorithms for MMPGs for a small number of players. For an efficient implementation we restricted our rewards to 0 and 1. This class of MMPG is sufficient to solve quantitative Buchi game problems where the goal of each player is to maximise the limit share of time spend in accepting states. Our main algorithm consists of two main steps: one is finding the base values from evaluating the underlying two-player mean-payoff games (2MPGs), and the other is to infer (and solve) a small number of linear programming problems. The key step to solving MMPGs with few players is the reduction to solving the underlying 2MPGs. We first give the algorithmic details and then discuss our experimental results.

For solving the 2MPGs, we have implemented a strategy improvement algorithm for identifying 0-mean partitions and coupled it with a logarithmic search to find the individual mean partitions to determine value at the vertices. During the logarithmic search, the vertices are successively cut into mean partitions of decreasing size. Our main algorithm consists of two main steps: one is finding the base values from evaluating the underlying 2MPGs, and the other is to infer (and solve) a small number of linear programming problems. The key step to solving MMPGs with few players is the reduction to solving the underlying 2MPGs. The number of different solutions to these games is usually small, and, consequently, the number of linear programming problems to solve is small, too. There are many algorithms for solving 2MPGs. The most promising candidates in our view are efficient algorithms [BV07]. An α-mean partition of a 2MPG is the subset of vertices, for which the return is $\geq \alpha$.

We first describe an optimal strategy improvement algorithm from [Sch08] for finding the 0-mean partitions and then we describe how we can expand it to evaluate 2MPGs.

4.7.1 Strategy Improvement Algorithm

This section summarises the strategy improvement algorithm from [Sch08] for mean-payoff games that gives an optimal improvement in every step by selecting the best

combination of updates from a set of all profitable changes. For full proofs of the results related to this algorithm we refer the reader to [Sch08]. We have implemented this algorithm for solving two-player games and give its quantitative expansion in the Section 4.7.2. The algorithm allows us to take the global effects of different local modifications to the strategy into account. The algorithm is primarily based on computing estimations at every step. We give the technical definitions of Escape games from [Sch08] and game arena that are being used in the algorithm.

Escape Games Escape games are total reward games that are tailored for the optimal improvement method. They allow player 0 to terminate every play immediately on each of her positions. Technically this is done by extending the arena with a fresh *escape position*, which forms a sink of the extended arena, and can be reached from every position of player 0. Every play of an *escape game* either eventually reaches the escape position and then terminates, or it is an infinite play in the non-extended arena.

Bipartite Arena The arena is assumed to be bipartite for technical convenience.

Extended Arena In an escape game, the finite arena $\mathcal{A} = (V_0, V_1, E)$ where V_0 is the set of Player 0 vertices, V_1 is the set of Player 1 vertices, and E is the set of edges, is extended to the directed graph $\mathcal{A}' = (V_0, V_1', E')$, which extends the arena $\mathcal{A}$ by a fresh position $\perp$ of player 1 ($V_1' = V_1 \uplus \{\perp\}$) that is reachable from every position of player 0 ($E' = E \cup V_0 \times \{\perp\}$). The escape position is a sink in $\mathcal{A}'$.

Finite Plays Since the escape position is a sink, every play terminates when reaching $\perp$. The set of plays is therefore extended by the finite plays $\pi = p_0 p_1 p_2 p_3 \ldots \perp \in (V_0 \uplus V_1)^* \{\perp\}$.

Escape Games An *escape game* is a game $\mathcal{E} = (V_0, V_1, E, w)$, where $\mathcal{A} = (V_0, V_1, E)$ is a finite arena, and $w : \mathbb{E} \to \mathbb{Z}$ is a weight function that assigns an integer value to every edge. An escape game is played on the extended arena $\mathcal{A}' = (V_0, V_1', E')$. For the extended arena weight function is extended to $w' : \mathbb{E} \to \mathbb{Z}$ with $w'(e) = w(e)$ for all $e \in E$, $w'((v_0, V_1')) = 0$ for all $v_0 \in V_0$. An infinite play $\pi = p_0 p_1 p_2 \ldots \in V^\omega$ of an escape game is evaluated to ∞ if $\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} w'\big((p_i, p_{i+1})\big) = \infty$ and to $-\infty$ if $\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} w'\big((p_i, p_{i+1})\big) = -\infty$. A finite play $\pi = p_0 p_1 p_2 \ldots p_n \perp$ is evaluated to $v(\pi) = \frac{1}{n} \sum_{i=0}^{n-1} w'\big((p_i, p_{i+1})\big)$. While the objective of player 0 is to maximise this value, the objective of player 1 is to minimise it.

Estimations An *estimation* is defined as a valuation function $v : V_0 \cup V_1' \to \mathcal{N}$ for an escape game $\mathcal{E} = (V_0, V_1, E, w)$ as witnesses for the existence of a memoryless strategy f of player 1, which guarantees that every f-conform play π starting in some position p is evaluated to $\rho(\pi) \geq v(p)$. Formally, an estimation v has to satisfy the following side conditions:

- $v(\perp) = 0$ (the function maps all outgoing edges from $v_0 \in V_0$ to V_1' to 0),
- for every position $p \in V_0$ of player 0 there is an edge $e = (p, p') \in E'$ such that $v(p) \leq v(p') + w'(p')$ is true,
- for every position $q \in V_1$ of player 1 and every edge $e = (q, q') \in E$ such that $v(q) \leq v(q') + w'(q')$ is true, and
- player 0 has a strategy f_∞ that maps every position $p \in V_0$ of player 0 with $v(p) = \infty$ to a position $p' = f_\infty(p)$ with $v(p') = \infty$, and which guarantees that every f_∞-conform play π starting in p is evaluated to $\rho(\pi) = \infty$.

A trivial estimation is simple to construct: We denote with v_0 the estimation that maps the escape position to $v_0(\perp) = 0$, every position $p \in V_0$ of player 0 to $v_0(p) = 0$, and every position $q \in V_1$ to $v_0(q) = min\{0 + w'(q') \mid (q, q') \in E\}$.

Remark The simple construction of the trivial estimation is the only position where the restriction to bipartite games is used.

Solving Escape Games Escape games are are tailored for the stepwise optimal improvement method. The strategy improvement algorithm is a stepwise optimal algorithm in that it gives an improvement of estimations in every step till a fixed point is reached. Every estimation (for example, the trivial estimation v_0) can be used as a starting point for the algorithm.

Optimal Improvement The estimations intuitively refer to strategies of player 0 for the extended arena. (Although estimations are a more general concept; not all estimations refer to a strategy.) The edges of the improvement arena $\mathcal{A}_v$ of an escape game $\mathcal{E} = (V_0, V_1, E, w)$ and an estimation v that originate in positions of player 0 refer to all promising strategy updates, that is, all strategy modifications that *locally* lead to a – not necessarily strict – improvement (profitable and stale modifications). We call an improvement v' of v *optimal* if it dominates all other estimations $\widehat{v}$ that refer to (memoryless) strategies of player 0 that contain only improvement edges. Finding this optimal improvement thus relates to solving an *update game*, which deviates from the full escape game $\mathcal{E}$ only by restricting the choices of player 0 to her improvement edges.

Basic Update Step Instead of computing the optimal improvement v' of an estimation v directly, the optimal update $u = v' - v$ is computed.

For a given escape game $\mathcal{E} = (V_0, V_1, E, w)$ with estimation v, we define the *improvement potential* of an edge $e = (p, p') \in E_v$ in the improvement arena $\mathcal{A}_v$ as the value $P(e) = v(p') + w(p') - v(p) \geq 0$ by which the estimation would locally be improved when the respective player chose to turn to p' (disregarding the positive global effect that this improvement may have). To construct the optimal update, the improvement arena is constructed, and the optimal update of the escape position is evaluated to $u(\perp) = 0$.

The improvement of the remaining positions is then evaluated successively by applying the following evaluation rule:

1. if there is a position $p \in V_1$ of player 1 that has only evaluated successors, we evaluate the improvement of p to $u(p) = min\{u(p') + P\big((p,p')\big) \mid (p,p') \in E\}$,
2. else if there is a position $p \in V_1$ of player 1 that has an evaluated successor p' with $u(p') = P\big((p,p')\big) = 0$, we evaluate the improvement of p to $u(p) = 0$,
3. else if there is a position $p \in V_0$ of player 0 that has only evaluated successors, we evaluate its improvement to $u(p) = max\{u(p') + P\big((p,p')\big) \mid (p,p') \in E_v\}$[3],
4. else we choose a position $p \in V_1$ of player 1 with minimal intermediate improvement $u'(p) = min\{u(p') + P\big((p,p')\big) \mid p'$ is evaluated and $(p,p') \in E\}$ and evaluate the improvement of p to $u(p) = u'(p)$. (Note that $min\{\emptyset\} = \infty$.)

Correctness The basic intuition for the stepwise optimal improvement algorithm is to re-estimate the value of a position only *after* all its successors have been re-estimated. In this situation, it is easy to determine the optimal decision for the respective player. In a situation where all unevaluated positions do have a successor, we note that every cycle in $\mathcal{A}_v$ has non-negative weight (weight ≥ 0), and every infinite play in $\mathcal{A}_v$ is evaluated to ∞. An optimal strategy of player 1 will thus turn, for some position of player 1, to an evaluated successor. It is safe to chose a transition such that the minimality criterion on the potential improvement u' is satisfied, because, independent of the choice of player 1, no better potential improvement can arise at any later time during this update step. Following these evaluation rules therefore provides an optimal improvement.

Complexity In spite of the wide variety of strategies that are considered simultaneously, the update complexity of the algorithm is surprisingly low. The stepwise optimal improvement algorithm generalises Dijkstra's single source shortest path algorithm to two-player games. The critical part of the algorithm is to keep track of the intermediate update u', and the complexity of the algorithm depends on the used data structure. The default choice is to use binary trees, resulting in an update complexity of $O(m \log n)$. However, using advanced data structures like 2-3 heaps reduces this complexity slightly to $O(m + n \log n)$.

4.7.2 Quantitative evaluation of mean-payoff games

For finding the individual mean partition, we first used the algorithm outlined above in the Section 4.7.1 to find 0-mean partitions, and in this section we give in detail how to expand it to find the value of 2MPGs.

[3] We could also choose $u(p) = max\{u(p') + P\big((p,p')\big) \mid p'$ is evaluated and $(p,p') \in E\}$.

Solving 2MPGs

Recall that in two-player mean-payoff games, both players have optimal memoryless strategies. Under such strategies, the game will follow a 'lasso path' from every starting vertex: a finite (and possibly empty) path, followed by a cycle, which is repeated infinitely many times. The value of a game position is defined by the average of the edge weights on this cycle. In our context, the edge weights are either 0 or 1. As both players would have a memoryless optimal strategy for 2MPGs, the value of playing them in our setting with payoffs 0 and 1 is a cycle of length l. The values of the vertices are therefore fractions $\frac{a}{l}$ with $0 \leq a \leq l \leq n$, where n is the total number of vertices in the game graph, and a is the number of 'accepting' events in the DBA that refers to the objective of the respective player, i.e., the edges with value 1, occurring on this cycle.

Note that such an algorithm only needs to use integer values: to determine the $\frac{a}{l}$ mean partition, we can replace edge weights 1 by a and edge weights 0 by $a - l$, and then determine the 0-mean partition. Conceptually, to find the $\frac{a}{l}$-mean partition, one would simply subtract $\frac{a}{l}$ from the weight of every edge and look for the 0-mean partition. However, to stay with integers, it is better to use integer values on the edges, e.g., by replacing the 0s by $-a$, and the 1s by $l - a$. For games with n vertices, there are only $O(n^2)$ values for the fraction $\frac{a}{l}$ to consider, as optimal memoryless strategies always lead to lasso paths and only the cycle at the end of the lasso determines the values for a and l, where $0 < a < l \leq n$.

We start by narrowing down the set of values by classifying the mean partition in a logarithmic search. After determining the $\frac{1}{2}$ mean partition, we know which values are < 0.5 and ≥ 0.5, respectively. The two parts of the game can then be analysed further, determining the $\frac{1}{4}$ and $\frac{3}{4}$ mean partition, respectively. After s such partitioning, all values in a partition of the game are either known to be in an $[k \cdot 2^{-s}, (k+1) \cdot 2^{-s}[$ interval for some $k < 2^s - 1$, or in the interval $[1 - 2^{-s}, 1]$. We stop to bisect when the size p of a partition is at most 2^s. In this case, the respective interval has $f \leq p$ fractions with a denominator $\leq p$. We determine them, store them in a balanced tree, and use it to determine the correct value of all vertices of the partition in $\lceil \log_2 f \rceil$ steps.

Solving multi-player mean-payoff games

Now we determine a order in a binary search manner in order to find the relevant mean partition.

1. To start with, we find a $\frac{1}{2}$ mean partition. For the $\geq$ part of the game, we then continue with a $\frac{3}{4}$ mean partition, and so forth. After $s = \lceil \log_2 n \rceil$, we have narrowed the area down to an interval of length 2^{-s}, and we know that the value of the vertex lies is in this interval.

2. For each denominator, there is at most one numerator in this interval[4]. Thus, going through all possible denominators, we can then sort the resulting fractions

[4]Exception: $n = 2^s$ But then we can simply look for the 1-mean partition and are done.

in a balanced tree. It suffices to take those, which are relative prime. (The value is otherwise already in the balanced tree.)

3. We then try the values of in the balanced tree, starting with the root. At most height-of-the-tree many further iterations are needed ($O(\log n)$ many).

The number of different values of nodes in a 2MPG is usually small, and certainly it would be much smaller than the number of vertices in the game. Consequently, the number of constraint systems is also small for a small number of players.

4.7.3 Linear Programming problem

We used the techniques from Section 4.7.2 to evaluate a number of randomly created three player MPGs, where the player take turns. We consider three players – player 1, player 2 and a leader and two different evaluations on the same game graph. We first see how each player fares when they try to maximise their return against a coalition of all other players, including the leader.

In the first evaluation, leader is with player 1 and they form a coalition (minimiser) against player 2 (maximiser) on the payoffs for player 2. We find the different possible mean values in this evaluation, using the algorithm from above. In the second evaluation, leader is with player 2 and they form a coalition (minimiser) against player 1 (maximiser) on the payoffs for player 1. We also note the different possible mean values in this evaluation, using the algorithm from above.

The results of these 2MPGs provide the constraints for the linear programming problems. These different values form the different thresholds that we have to consider. We now consider all possible combinations of these values for the followers and determine the vertices that comply with them. We then do the following for each combination:

- recursively remove the nodes that have no successor
- remove the nodes that are not reachable from the initial state, i.e., we determine the set of reachable vertices
- for the set of reachable vertices, we then determine the strongly connected set of components (SCCs)
- for the SCCs, we build and solve the respective linear program

Constraints on SCCs

To construct the linear programs over the SCCs formed from above, we have side constraints on the edge-ratio and vertex-ratio (to comply with the limit behaviour of nodes and edges) and we have constraint over the reward of player. These constraints over ratios of edge and vertices for leader and incentive equilibrium follow from Section 4.4.2.

For constraints over the rewards for the players, we refer to Section 4.5 for the construction of an incentive equilibrium and to Section 4.4.2 for the construction of a

leader equilibrium. For readability purpose, we reiterate the constraints here. First we have constraints on the nodes and edges that form part of the strongly connected component S:

- ratio of vertices and edges that are not part of S is 0,
- ratio of vertices and edges that are in S is ≥ 0,
- the sum of ratio of vertices is 1

The second part of constraint system is constraint over rewards:

- for every player p other than the leader, that own some vertex in S, we have constraint over her reward

$$\iota_p + \sum_{e \in E} p_e r_p(e) \geq \max_{v \in Q}(r_p(v))$$

  where p_e is the ratio by which edge e is taken, $r_p(e)$ is the edge weight for player p at edge e, $r_p(v)$ is the mean value of game for player p in set S and ι_p is an incentive given to player p and $\iota_p \geq 0$

- objective of the constraint system is to maximise leader's reward, i.e., maximising reward at leader nodes in S. This gives us objective function: maximise $\sum_{e \in E} p_e r_l(e) - \sum_{p \in P} \iota_p$

For a set Q, we may have number of $SCCs$ and there is a constraint system for every such component. In this case, we would take the one that maximises leader's reward.

For leader equilibrium, however, there is no incentive in second part of the constraint system.

4.7.4 Experimental Results

We implemented our tool [mpg] in C++ and carried out experiments on a Linux machine. Our tool is available for download and requires C++ 11 standard compiler for compilation. It makes call to the linear programming solver [MKP] for solving the underlying linear programs. Although market tools are available for solving two-player mean-payoff games but we implemented our own solver because of the following reasons. Firstly, we implemented powerful strategy improvement algorithm [Sch08] whereas the other known solvers [[CPR14],[CHJS11]] implement the original approach of Paterson and Zwick [ZP96]. Secondly, the update complexity of strategy improvement algorithm is low as compared to the standard algorithm [ZP96]. Also, the strategy improvement algorithm allows us to take weights in binary that is useful for our purpose.

Our experiments indicate that our implementation of the algorithm can solve examples of size 100 nodes and 10 players within 30 minutes. The algorithm is, of course,

much faster for the games with two or three players. Our current implementation accepts mean-payoff games with three players where weights are given in binary (0 or 1). The algorithm we implemented solve two-player mean-payoff games(2MPG). The test cases we currently use are for parity games, and we had to use some method to convert a parity game to mean-payoff bipartite game. Also, we have given payoffs on vertices and not on edges. We assign a weight of $(-n)^c$ to a vertex in the mean-payoff game if it has a parity 'c' in a parity game with 'n' vertices. we take a parity game as an input and construct a three player mean-payoff game where for every vertex a coin is tossed to decide the owner of the vertex, the reward for every player, and the initial vertex. We have generated several set of examples and generated parity games by PGSolver suite [PGS], and obtain results for leader equilibrium as well as incentive equilibrium for mean-payoff games. Our tool [mpg] currently solve games of 40-50 nodes fairly quickly. However, for large number of these examples answer for both equilibria is the same. Figures 4.10 and 4.11 show the experimental results for the following two problem classes.

- Figure 4.10 shows results for a generalisation of example from Figure 4.7 for multiple players with n nodes in the inner cycle and $n-1$ nodes in outer cycles where n is the number of players. Recall the example from Figure 4.7. We generalise this example for token ring graph parameterised by 2 variables, n and d. It has 'n' nodes on the inner cycle, each of which correspond to 'n' different players and each of these 'n' nodes is also present on another cycle of length 'd'. The weights are set such that, all players except the leader get '1/n' if they chose the inner ring and get '1/d' if they chose their respective outer ring. The leader reward is '1' in the inner ring and '1/d' in all the other rings. The data supports the pen-and-paper analysis that incentives are useful iff $n > d > n(n-1)/(2n-1)$ holds. Figure 4.10 shows the leader reward for this example and the running time of our

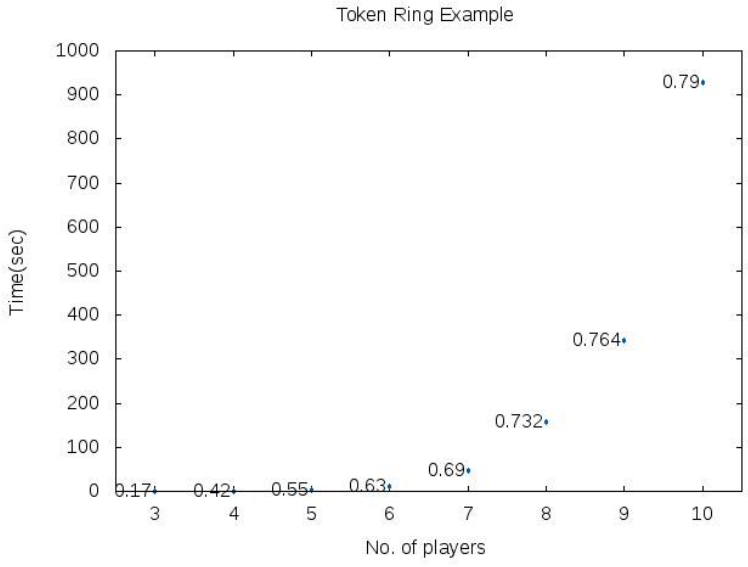

FIGURE 4.10: Token-ring example.

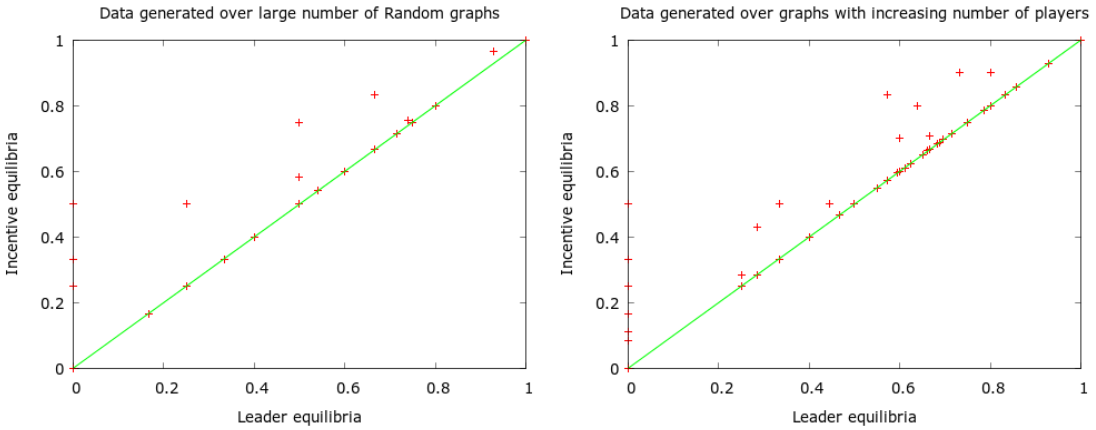

FIGURE 4.11: Results for randomly generated MMPGs.

tool to compute it. We represent on y-axis the time taken (in seconds) to compute incentive equilibrium and on x-axis we represent the total number of players in the game. The data in the figure shows the leader's reward in an incentive equilibrium for given number of players.

- Figure 4.11 draws the comparison between leader equilibrium and incentive equilibrium w.r.t. the leader's reward. The figures show the leader's reward for same game graph in different equilibrium (leader and incentive equilibrium). The left plot shows the difference between incentive equilibrium and leader equilibrium for randomly generated three player mean-payoff games. The number of nodes is changed from 10 to 15 and for each number, 10 cases have been generated, and a total of 150 cases have been generated. The right plot in the figure shows similar results on random graphs, with number of players increasing from 3 to 10. Number of nodes is fixed around 20 and a total of 150 cases have been generated.

The evaluation results confirm that the leader reward increases significantly in incentive equilibria when compared to leader equilibria.

4.8 Discussion

The two main contributions of this chapter are the introduction of leader and incentive equilibria in multi-player mean-payoff games and the concept of well behaved reward and punish strategy profiles as a technical foundation to them.

Well behaved reward and punish strategy profiles are general instruments for optimising the payoff of one player, while projecting away problems like the potential non-existence of limit average values. It is our belief that they will be useful in many related optimisation problems. The introduction of leader equilibria is a conceptual change to Nash equilibria, where an interested party overcomes the antinomy of Nash equilibria exemplified in Figure 4.8: the interested party (which we christened the leader) might improve her payoff by choosing a strategy, which is not stable for herself in the

Nash sense of not being able to improve the payoff by unilaterally deviating from her strategy. The concept extends the set of rational control objectives. It allows, for example, for mixing environments that are rationally following their own objectives with a hostile environment. For such environments, it provides worst case *rational* results, where rational refers to the way the rational players behave: we assume that they follow a strategy that they do not have an incentive to deviate from.

We extend the settings by showing that the rational leader can further improve over her outcome by paying small incentives to her followers. At first, it may not seem to be a rational move of the leader, but close insight would show how a leader might improve her reward in this way. The incentive equilibria are therefore seen as an extension to leader equilibria, where a rational leader, by giving an incentive to every other player in the game, can derive an optimal strategy profile.

The solutions one obtains can be used to create stable rules that optimise various outcomes, including social optima as well as egoistic solutions. We believe that these techniques are helpful for the leader when maximising the return for a single player and would also be instrumental in defining stable rules and optimising various outcomes. We implement these techniques in a tool and our evaluation results from Section 4.7 show that the results are significantly better for the leader in an incentive equilibrium as compared to her return in a leader equilibrium.

Chapter 5

Discounted sum games

This chapter is mainly based on the results from [GSW15]. It extends the use of leader equilibria in multi-player discounted sum games. In this chapter, we establish the existence of optimal bounded memory leader strategy profiles in multi-player discounted sum games. Note that we have given a general introduction to the discounted sum games in Chapter 1 (cf. Section 1.2).

This chapter is organised as follows. We give the main contributions of this chapter in Section 5.2.2. In Section 5.3, we give a formal introduction to multi-player discounted sum games and give the notations used throughout this chapter. In Section 5.4, we show that leader equilibria are superior to Nash equilibria in discounted sum games. We also establish the usefulness of memory by giving relevant examples. We show that more memory would help and that there are cases where leader might benefit from having infinite memory. We give the construction of optimal leader strategy profiles with a discussion on reward and punish strategy strategies in Section 5.5. We establish the NP-completeness of the related decision problem. We discuss why pure strategies are a natural choice to study reward and punish strategies in Section 5.7 and give conclusions in Section 5.8.

5.1 Abstract

In this chapter, we establish the existence of optimal bounded memory strategy profiles in multi-player discounted sum games. We introduce a non-deterministic approach to compute optimal strategy profiles with bounded memory. Our approach can be used to obtain optimal rewards in a setting where a powerful player selects the strategies of all players for Nash and leader equilibria, where in leader equilibria the Nash condition is waived for the strategy of this powerful player. The resulting strategy profiles are optimal for this player among all strategy profiles that respect the given memory bound, and the related decision problem is NP-complete. We also provide simple examples, which show that having more memory will improve the optimal strategy profile, and that sufficient memory to obtain optimal strategy profiles cannot be inferred from the structure of the game.

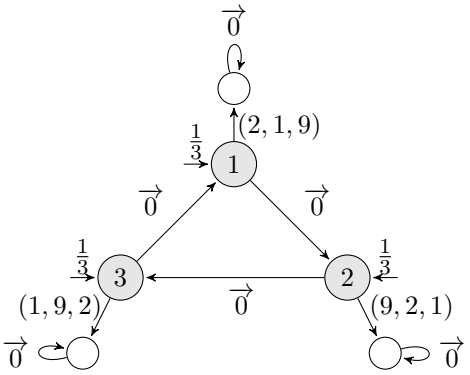

FIGURE 5.1: A discounted sum game with no memoryless Nash or leader equilibrium.

5.2 Introduction

Discounted sum games [Sha53, Sen94] are the stochastic games with quantitative objectives that have been introduced by Shapley [Sha53]. They are played on a finite directed graph without sinks, where each vertex is owned by one of the players. Intuitively, they are played by placing a token on the graph, which is moved forward by the players. We consider an initial probability distribution over all vertices to select the start vertex. As an example, we refer to Figure 5.1, where vertex 1, vertex 2 or vertex 3 each can be taken as a start (or: initial) vertex with probability $\frac{1}{3}$.

Initially, the token is placed on a start vertex. Whenever the token is on a vertex, the player who owns this vertex will select an outgoing edge and move the token along this edge. This way, the players construct an infinite play. Quantitative games [BDS13] are good models for studying non-terminating programs with multiple components that interact in non-cooperation mode. In quantitative games, players have goals defined by the payoffs on the edges (sometimes on the vertices). For these payoffs, the players have quantitative targets, such as maximising their individual limit average or the discounted sum of their individual rewards, where the value of a play is computed under a discount factor. Solutions to these games are the strategy profiles that consists of strategies–recipes how to play–for each player. However, in a realistic situation, these solutions need to be implementable, and thus players have to cope with limited resources such as limited memory. Strategy profiles should also satisfy same basic consistency constraints. The reason for this is that the players are assumed to be rational. The lowest level of rationality for a player is to take a look at her strategy profile, and to check if she would gain by changing her own strategy. Strategy profiles where all strategies pass this test are *stable* in terms of Nash equilibria [Leh90, Nas50, OR94]. Thus, in a Nash equilibrium, no player benefits from changing her strategy unilaterally.

The second eminent class of equilibria goes back to von Stackelberg and is referred to as Stackelberg equilibria or *leader equilibria* [vS34]. In economic game theory, leader equilibria refer to a setting, where a powerful player can move first, or announce her move first, rather than moving at the same time as the remaining players. This 'right of

the first move' provides her with some advantage over the other players. Broadly speaking, the Nash requirements of having no incentive to deviate only affects the remaining players, but not the leader herself. The *leader* can assign the strategies to all players, including herself. While we still require the strategy profile to be stable in that the *other* players do not have an incentive to deviate, the leader herself may be in a position to improve over her current strategy by deviating unilaterally. Thus, every Nash equilibrium is a leader strategy profile, but not every leader strategy profile is Nash. We call strategy profiles that are optimal for the leader *leader equilibria (LE)*. The more relaxed condition of a leader strategy profile implies that leader equilibrium can be selected from a larger base (cf. Figure 5.3). The leader's payoff can therefore improve as compared to Nash equilibria. In this chapter, we study leader equilibria and Nash equilibria for the leader in discounted sum games (DSGs) that use bounded memory.

5.2.1 Related Work

The theory of stochastic games was introduced by Shapley in [Sha53]. He showed that every two player discounted zero-sum game has a value and that optimal positional strategies exist for both the players. This idea is further extended in [Fin64] to establish the existence of stationary equilibria in stochastic multi-player games. Bewley and Kohlberg [BK78] have shown that, in two player zero-sum undiscounted stochastic games where both the set of action and the state spaces are finite, stationary optimal strategies exist for both the players. Gimbert and Zielonka [GZ04] have studied infinite two player antagonistic games with more general reward functions. They have given sufficient conditions that ensure both the players to have positional (memoryless) optimal strategies. Letchford et al. [LMC$^+$12] have considered computing optimal Stackelberg strategies in stochastic games. They studied this in context with correlation equilibria and discuss the value of correlation and commitment in stochastic games. The importance of discounting has been discussed in detail in [dAHM03]. They studied it in system theory and established that discounting has a natural place in non-terminating systems, in probabilistic systems as well as in multi-component systems. They established discounted version of non-terminating system properties that corresponds to ω-regular properties. Discounting the payoff values is an important criterion in events where near future is considered more important than far-away future. Besides, it is considered an important factor in Markov decision processes, economic applications, and, in game theory.

Their complexity is same as mean-payoff games and lies in complexity class NP$\cap$Co-NP [EM79], in UP$\cap$Co-UP [ZP96]. Computing an optimal value in these games can be done in pseudo-polynomial time [ZP96, BCD$^+$11], smoothed polynomial time [BEF$^+$11], PPAD [EY10], and randomised subexponential [BV07] time. Berg and Kitti [BK13] have studied subgame perfect pure strategy equilibria in discounted sum games. They analyse subgame perfect equilibria in games with perfect information. Brihaye et al. [BDS13] have studied the existence of simple Nash equilibria in non-terminating games with various mixed reward functions. The strategies used in this chapter are inspired by

the strategies introduced in [Fri71]. Gupta and Schewe [GS14] have studied the optimal leader strategy profiles in context with multi-player mean payoff games.

5.2.2 Contributions

The main contributions of this chapter are as follows. We show in this chapter that, in multi-player discounted sum games, the leader can benefit from more memory (Lemma 5.3), and that there are actually cases, where infinite memory is needed for leader equilibria (Theorem 5.4 and Theorem 5.5) and Nash equilibria (Theorem 5.6).

We do not hold strategies that require infinite memory to be realistic, and therefore discuss the construction of strategies that use only bounded memory. We first show that memoryless leader equilibria do not always exist, a simple corollary from the existence of games without memoryless subgame perfect equilibrium [KFSV09]. The example from Figure 5.1, inspired by [KFSV09], has no memoryless Nash equilibria.

The example shows a discounted sum game with discount factor $\lambda = \frac{1}{2}$ and with no memoryless Nash (nor leader) equilibrium. There are three players and the start vertex is picked uniformly at random out of the vertices 1, 2 and 3. These vertices are controlled by players 1, 2 and 3 respectively. Each edge is labelled with a reward vector (r_1, r_2, r_3) where r_i is the reward player i gets for traversing that edge.

Therefore, when the leader is not among the three players who own the three central vertices, there is no memoryless leader equilibrium for this game. There even exists a game with a fixed starting position where no pure Nash equilibria exist [GO14].

This problem, however, seems artificial when reviewing traditional classes of Nash equilibria. They often use the traditional form of 'reward and punish' strategy profiles [Fri71, BDS13, GS14]. Strategy profiles define a play, the play that ensues when all players follow the strategies assigned to them. Reward and punish strategy profiles broadly consist of this play, and an agreement that the first player who deviates is punished: all other players collude henceforth, following the new goal to harm the deviator.

Upon deviation, reward and punish strategy profiles therefore turn into two-player games, and thus enjoy the usual memoryless determinacy. The memory needed for this is tiny: one only needs to store who has deviated. We therefore argue that the resource bounds should refer to the construction of the main play, i.e., main path before deviation.

We give a simple non-deterministic polynomial time approach for assigning reward and punish strategies that meet or exceed a given payoff bound for the leader and uses memory only within a given bound. In Section 5.6, we show that the decision problem whether a pure strategy with bounded memory that gives a reward greater than or equal to some threshold value exists is NP-complete.

5.3 Preliminaries

A multi-player discounted sum game (MDSG) is a game played on the finite directed weighted graph $\mathcal{G}$ defined as a tuple $\langle P, V, \{V_p \mid p \in P\}, \Delta, A, T, \{t_p : V \times A \to \mathbb{Q} \mid p \in P\}\rangle$, where P is a finite set of players, V is a finite set of vertices, $\Delta : V \to [0,1]$ is a probability distribution over V, which for each $v \in V$ specifies the probability of selecting v as the start vertex. $\{V_p \mid p \in P\}$ is a partition of the vertices V into the sets V_p of vertices owned by player p, A is a finite set of actions, $T : V \times A \to V$ is a set of transitions that maps vertices and actions to vertices, and $\{t_p \mid p \in P\}$ is a family of reward functions defined as $t_p : V \times A \to \mathbb{Q}$ for all $p \in P$ that assigns, for each respective player p, a reward for each action a that is taken from a vertex v (or, likewise, for the transition taken). The game is played by moving a token along the edges of the graph, starting from the start vertex as given by the probability distribution Δ. We use this initial probability distribution to select a start vertex. Each vertex v belongs to exactly one player p. At vertex v, the player who owns v selects the next action a. The token is then moved forward to the vertex as given by the transition $T(v,a)$. This results in an infinite path, called a *play*. We denote the reward for player p at any transition $T(v,a)$ by $t_p(v,a)$. An MDSG is called a zero-sum game if, for all vertices $v \in V$ and for all actions $a \in A$, $\sum_{p \in P} t_p(v,a) = 0$ holds. The payoff at every transition is discounted by a discount factor λ, where $0 < \lambda < 1$. In DSGs, the payoff (or: reward) for player p at the i^{th} transition is given by $t_p(v_i, a_i) \cdot \lambda^i$. For an infinite play $\pi = v_0, a_0, v_1, \ldots$, we denote the reward for player p by $r_p(\pi) = \sum_{i=0}^{\infty} t_p(v_i, a_i) \cdot \lambda^i$.

The way that the respective player p chooses the successor vertex is defined by a *strategy* σ_p. We consider *pure strategies*, which are functions $\sigma_p : (VA)^* V_p \to A$ from initial sequences of plays to actions. We focus on two types of pure strategies, memoryless and bounded memory strategies. A pure memoryless strategy (or: a positional strategy) is a strategy, in which the choice of the next vertex depends only on the current position, whereas a pure bounded memory strategy is a strategy, where the choice of next vertex depends on finite memory. For a bounded memory M (where M is simply a finite set of fixed size, the memory bound with a dedicated initial value m_0), we define two functions: the memory update function, and the memory usage function that provides us with the action that is to be selected. The memory usage function is a mapping $\mathcal{U} : M \times V \to A$ that maps a memory state and a vertex to an action. In the classic memory model, the memory update function $\mathcal{M} : M \times (V \times A) \to M$ defines how the memory is updated; it maps a memory state, a vertex, and an action to a new memory state. Thus, the memory works as a Moore machine without output, where M is the memory and $\mathcal{M}$ is the transition function.

As discussed in the introduction, the example from Figure 5.1 shows that this memory model does not always lead to an equilibrium, at least not for arbitrary M. We therefore define a memory model for reasoning with bounded resources (cf. Corollary 5.11). We refer to this model as *compliance memory*, as it only refers to the histories, where all

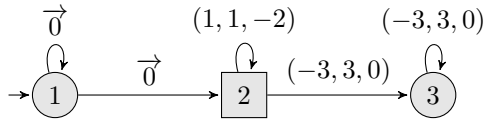

FIGURE 5.2: A discounted sum game with discount factor $\frac{1}{2}$.

players have complied to their strategies. This justifies a *partial* memory update function $\mathcal{M} : M \times (V \times A) \to M$, where $\mathcal{M}(m; v, a)$ is defined if, and only if, $a = \mathcal{U}(m, v)$. When the action a differs from the action defined by the memory usage function, the system remembers only who caused the deviation, and then switches into a different mode, where it uses a memoryless strategy (cf. Theorem 5.9 and Corollary 5.10).

The input alphabet $V \times A$ is a product of the last vertex, the action selected, and the vertex reached on a transition. A strategy profile σ defines an expected reward, denoted $\mathbb{E}_p(\sigma)$ for each player p. In this chapter, we shall focus on the reward of positional and bounded memory strategy profiles. For a positional strategy profile σ, the payoff from every vertex is well defined. By abuse of notation, we use $\mathbb{E}_p(\sigma, v) = t_p\big(v, \sigma(v)\big) + \lambda \mathbb{E}_p\Big(\sigma, T\big(v, \sigma(v)\big)\Big)$ to denote the payoff for player p when starting in a vertex v. Note that this implies $\mathbb{E}_p(\sigma) = \sum_{v \in V} \Delta(v) \mathbb{E}_p(\sigma, v)$. Note that we have formally introduced equilibrium concepts in Chapter 2. However, we now use expected reward $\mathbb{E}_p(\sigma)$ for each player p. We only define here concepts that have not been introduced before.

In two-player DSGs, the set of vertices in $\mathcal{G}$ is partitioned into two sets where each vertex belongs to exactly one of the players and the player who owns the vertex decides the next move. For a MDSG $\mathcal{G} = \langle P, V, \{V_p \mid p \in P\}, \Delta, A, T, \{t_p : V \times A \to \mathbb{Q} \mid p \in P\}\rangle$, we define the two-player zero-sum DSG $\mathcal{G} = \langle P, V, \{V_p, V_o\}, \Delta, A, T, \{t_p, t_o\}\rangle$ played between player p and an opponent o, where the nodes of p and o partition V into two sets ($V_o = V \setminus V_p$) and their goals are antagonistic ($t_o(v, a) \mapsto -t_p(v, a)$). Note that not all MDSGs with two players in game are two-player games in this sense (two-player games need to be antagonistic zero-sum games). We denote the expected outcome for player p in a two-player game that starts at any vertex v by $r_p(v)$. A game is called memoryless determined if all players have optimal memoryless strategies. Two-player DSGs are memoryless determined [ZP96]: both players have an optimal positional strategy.

Theorem 5.1. *[ZP96] Two-player DSGs are memoryless determined.*

5.4 Leader and Nash equilibria

In this section, we show that leader equilibria are superior to Nash equilibria in simple zero-sum DSGs. For this, consider the three-player game from Figure 5.2. One of the players, player 2, acts as the *leader*. The game is played on a simple graph with three vertices, named 1, 2, and 3, owned by the respective player with the same name. Note that we denote the vertices owned by leader (resp. other players) by square (resp. circle)

vertices. We used the same notation throughout this chapter. In all remaining examples, we select an initial vertex with probability 1, and therefore mark the initial vertex with an incoming arrow. The game graph with the payoff vectors of each transition is shown in Figure 5.2, and we use a discount factor of $\lambda = \frac{1}{2}$. The payoff vectors represent the payoff of player 1, the leader, and player 3, in this order. Initially, player 1 can choose to play to vertex 2 or she can choose to remain in vertex 1. She plays to vertex 2 only if the leader, in her strategy profile, chooses to remain in vertex 2 for a while. At vertex 2, the leader has different options.

She can choose to play to vertex 3 (this is the option where she maximises her reward), she can choose to remain in 2 for a while, before continuing to vertex 3, or she can stay in vertex 2 forever. It is easy to notice that, when in vertex 2, the leader will immediately continue to vertex 3 in all Nash equilibria. Consequently, player 1 would never play to vertex 2 from vertex 1: staying in vertex 1 forever will yield a payoff of 0, while moving to vertex 2 in round i would, for $\lambda = \frac{1}{2}$, result in a payoff of $-\frac{3}{2^i}$. Thus, the only play that can result from a Nash equilibrium is the play 1^ω, where the overall reward for all participating players is 0. However, in a leader equilibrium the leader stays twice in vertex 2 and then progresses to vertex 3. In this case, the leader can assign player 1 the strategy to immediately progress to vertex 2, resulting in the play $1, 2, 2, 2, 3^\omega$. This will provide an overall payoff of 0 for player 1, 1.5 for the leader, and -1.5 for player 3.

Theorem 5.2. *Compared to Nash equilibria, leader equilibria may result in higher, but will never provide smaller rewards for the leader.*

Proof. While the example has proven the 'higher' part, note for the 'not smaller' part that all Nash equilibria are leader equilibria, such that a leader equilibrium cannot be inferior to a Nash equilibrium. They can, of course, be equal when a leader equilibrium is Nash. This is, for example, the case when leader owns no vertex. Thus, leader equilibria gives more leeway to the leader for the selection of optimal strategy profiles and forms a larger base of strategy profiles to choose from, as shown in Figure 5.3.

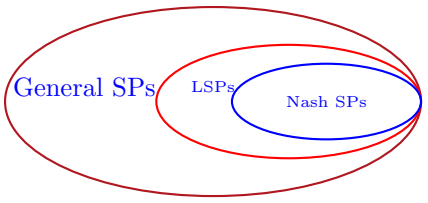

FIGURE 5.3: General strategy profiles ⊇ Leader strategy profiles ⊇ Nash strategy profiles.

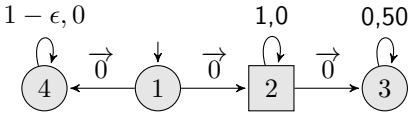

FIGURE 5.4: Increasing the memory helps.

Note that the game from Figure 5.2 can be used to argue that having memory helps, and having more memory helps more. Among the positional strategies of the leader, staying in vertex 2 forever (with an overall payoff of 1 for player 1 and the leader, and -2 for player 3, respectively) is superior to continuing immediately to vertex 3 (because in the latter case player 1 will stay in vertex 1, see above). So, while still superior to the only Nash equilibrium, it is inferior to the strategy described above, which uses a tiny amount of memory. To observe that, in general, more memory helps more, consider the situation where one lets λ grow towards one. It is easy to see that, the closer λ gets to one, the longer leader would stay in vertex 2 in leader equilibrium for the respective discount factor. The optimal memory bounded strategy for the leader therefore improves with the memory we allow for.

Lemma 5.3. *The optimal reward for the leader in a Nash or leader equilibrium improves with the increase of the available memory.*

It now becomes tempting to assume that we could use this observation to identify a situation where an optimal leader strategy profile is reached. That is, given a fixed discount factor, is there a $k \in \mathbb{N}$ such that an optimal leader strategy profile for memory k is considered optimal for infinite memory? The answer to this question is negative.

Theorem 5.4. *For any fixed discount factor λ, there is no memory bound k such that an optimal leader strategy profile with memory bound k is an optimal leader strategy profile.*

Proof. For this, we refer to the example from Figure 5.4, where leader acts as player 2. Here, we argue that having a finite memory at the vertices is sufficient for a leader equilibrium, but the effect of increasing the memory is different than in our first example. Irrespective of the discount factor it is apparent that the leader needs to promise sufficiently many, say s, loops in vertex 2 so that $\sum_{i=0}^{s-1} \lambda^i \geq \frac{1-\varepsilon}{1-\lambda}$. Consequently, the number of repetitions grows to infinity, for all $\lambda \in]0, 1[$, and with ε falling to 0. If the memory is smaller than minimal such s, then the leader would receive an overall reward of 0, either because she promises to stay for more than the memory bound many steps (and thus for ever) in vertex 2, or by not promising to do so and hence tempting the first player to move to vertex 4. If, on the other hand, the memory size is at least s, then the leader has enough memory to play the optimal pure strategy to move to vertex 3 after s loops in vertex 2.

Finally, so far for a fixed discount factor and a fixed game graph with weights, bounded memory was sufficient to guarantee optimal reward to the leader. We now show that infinite memory is sometimes needed in a leader equilibrium.

Theorem 5.5. *Optimal leader strategy profile may require infinite amount of memory even for a fixed two-player game with a fixed discount factor.*

Proof. We show this for a two-player game with three vertices depicted in Figure 5.5, where the leader is player 2. Vertices 1 and 3 belong to player 1 and vertex 2 belongs

to the leader. The rewards are depicted in the order (player 1, player 2) and we set $\lambda = 2/3$. Notice that player 1 will move to vertex 3 unless the leader can guarantee him a reward $\geq -1/\lambda = -3/2$ from vertex 2, because only then his total reward would be $\geq 1 - \lambda/\lambda = 0$. On the other hand, in the optimal leader strategy, the leader will try to give him exactly that much, because only then her payoff would be equal to $\lambda/\lambda = 1$. Proposition 1 in [CFW13] shows that the leader can achieve this value with a pure strategy, but only if she has an infinite amount of memory.

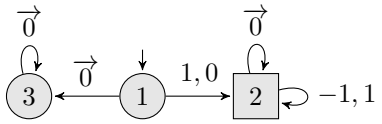

FIGURE 5.5: Leader benefits from infinite memory.

We can show the same for Nash equilibrium, but with 3-players. Also, we show that the optimal payoff of a player cannot be approximated by considering strategies with bounded memory only.

Theorem 5.6. *An optimal Nash equilibrium may require infinite amount of memory even for a fixed 3-player game with a fixed discount factor. Moreover, leader's optimal payoff can be arbitrary far away from her optimal payoff for bounded memory strategies.*

Proof. To show this, we refer to the Figure 5.6. We have three players here – player 1, player 2 and leader. The vertex 1, vertex 2 and vertex 3 are owned by player 1, player 2 and leader respectively. Rewards are given on the edges and are shown in the order (player 1, player 2, leader). We set the value of discount factor to be $\lambda = 2/3$. Starting from the initial vertex (vertex 1), player 1 can either go to the terminal state that has a reward of 0 for all the three players, or can move to the vertex 2. Similarly, at vertex 2, player 2 can either go to the terminal state or move to the leader vertex.

For an optimal strategy profile, leader has to promise to both player 1 and player 2 a reward of at least 3/2 at vertex 3, as otherwise at least one of them would prefer to terminate the game at their respective vertices. On the other hand, no matter what leader does, their rewards at vertex 3 sum up to 3, because the sum of their payoffs on the edges from vertex 3 is constant and equal to 1. Therefore, the leader has to promise to both player 1 and player 2 a reward of exactly 3/2. Proposition 1 in [CFW13] shows that the leader can achieve this value with a pure strategy, but only if she has an infinite amount of memory. The overall rewards of player 1 and player 2 from such a play $1 \cdot 2 \cdot 3^\omega$ would be 0. Note that this strategy profile would be a Nash equilibrium where leader's payoff is 2.

Finally, if the leader has only bounded amount of memory then one of the other players has to receive less than 0 from a play $1 \cdot 2 \cdot 3^\omega$ and would prefer to terminate the game before it reaches vertex 3. This implies that the optimal payoff of the leader for bounded strategies is 0, while for general strategies it is 2. The difference between these two can be made arbitrarily large by scaling the payoffs on the edges in this game.

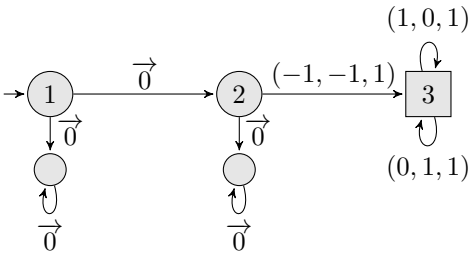

FIGURE 5.6: Leader benefits from infinite memory in Nash equilibria.

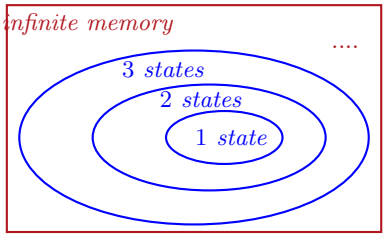

FIGURE 5.7: More memory states ⇒ more strategies.

Thus, an optimal strategy profile for a given player can be formed from a memoryless strategy, finite memory strategy or from infinite memory. More memory would, therefore, give more leeway to leader to select an optimal strategy profile (cf. Figure 5.7).

5.5 Reward and punish strategy profiles in discounted sum games

In this section, we show that for a play π, we could establish if there exists a leader (or Nash) strategy profile σ with $\pi = \pi_\sigma$, and, moreover, its extension to such a strategy profile is simple. For this, we first introduce reward and punish strategy profiles.

In reward and punish strategy profiles [GS14], the leader assigns a strategy to each player and each of them co-operates to produce a play π while playing in accordance with the assigned strategies. As soon as one player deviates, the remaining players team up with the leader and co-operate against the deviating player i. That is, they will henceforth follow the goal to minimise the payoff of player i, and act jointly as the antagonist of i in the underlying two-player DSG. Thus, in the resultant two-player game, while the objective of player i is still the same, the objective of all other players (including leader), is changed and has become to minimise the payoff of player i. Assuming that positional optimal strategies in this two-player DSG are fixed, π thus defines a reward and punish strategy profile, which we denote by $\mathsf{rps}(\pi)$. We now argue that

1. every leader resp. Nash strategy profile σ can be transformed into a leader resp. Nash strategy profile σ' with $\pi_\sigma = \pi_{\sigma'}$, and thus with similar rewards for all players, and

2. give necessary and sufficient conditions for a play π to be defined by some leader resp. Nash strategy profile.

We first discuss the necessary conditions for a path to be the outcome of a Nash (resp. leader) equilibrium, and then show that it is sufficient for a path to be the outcome of a Nash (resp. leader) reward and punish strategy profile.

Lemma 5.7. *If $\pi = v_0, a_0, v_1, \dots$ is the outcome of a Nash (resp. leader) equilibrium, then, for all $j \in \mathbb{N}$ and all players p (resp. all players $p \neq l$), $r_p(v_j) \leq \sum_{i=0}^{\infty} t_p(v_{j+i}, a_{j+i}) \cdot \lambda^i$ holds.*

Proof. We assume for contradiction that the condition is violated. We therefore select a $j \in \mathbb{N}$, and a player p (for leader equilibria a player $p \neq l$) such that $r_p(v_j) > \sum_{i=0}^{\infty} t_p(v_{j+i}, a_{j+i}) \cdot \lambda^i$. We then change the strategy of player p to follow her strategy from the two player discounted sum game from position j onwards. The resulting play $\pi' = v'_0, a'_0, v'_1, \dots$ with $v'_i = v_i$ for all $i \leq j$ and $a'_i = a_i$ for all $i < j$ satisfies

$$r_p(\pi') = \sum_{i=0}^{\infty} t_p(v'_i, a'_i) \cdot \lambda^i = \sum_{i=0}^{j-1} t_p(v'_i, a'_i) \cdot \lambda^i + \lambda^j \sum_{i=0}^{\infty} t_p(v'_{j+i}, a'_{j+i}) \cdot \lambda^i$$
$$\geq \sum_{i=0}^{j-1} t_p(v_i, a_i) \cdot \lambda^i + \lambda^j r_p(v_j) > \sum_{i=0}^{j-1} t_p(v_i, a_i) \cdot \lambda^i + \lambda^j \sum_{i=0}^{\infty} t_p(v_{j+i}, a_{j+i}) \cdot \lambda^i$$
$$= \sum_{i=0}^{\infty} t_p(v_i, a_i) \cdot \lambda^i = r_p(\pi).$$

□

Lemma 5.8. *If $\pi = v_0, a_0, v_1, \dots$ satisfies $r_p(v_j) \leq \sum_{i=0}^{\infty} t_p(v_{j+i}, a_{j+i}) \cdot \lambda^i$ for all $j \in \mathbb{N}$ and all players p (resp. all players $p \neq l$), then $\mathsf{rps}(\pi)$ is a Nash (resp. leader) equilibrium.*

Proof. We assume for contradiction that a player p (for leader equilibria a player $p \neq l$) has an incentive to deviate, and that the first position where player p selects a different action is $j \in \mathbb{N}$. Let $\pi' = v_0, a'_0, v'_1, \dots$, where $v'_i = v_i$ for all $i \leq j$ and $a'_i = a_i$ for all $i < j$, be the resulting play. We have,

$$r_p(\pi') = \sum_{i=0}^{\infty} t_p(v'_i, a'_i) \cdot \lambda^i = \sum_{i=0}^{j-1} t_p(v'_i, a'_i) \cdot \lambda^i + \lambda^j \sum_{i=0}^{\infty} t_p(v'_{j+i}, a'_{j+i}) \cdot \lambda^i$$
$$\leq \sum_{i=0}^{j-1} t_p(v_i, a_i) \cdot \lambda^i + \lambda^j r_p(v_j) \leq \sum_{i=0}^{j-1} t_p(v_i, a_i) \cdot \lambda^i + \lambda^j \sum_{i=0}^{\infty} t_p(v_{j+i}, a_{j+i}) \cdot \lambda^i$$
$$= \sum_{i=0}^{\infty} t_p(v_i, a_i) \cdot \lambda^i = r_p(\pi).$$

□

The first '$\leq$' is implied by the definition of rps, as the remaining players will play antagonistic to p, such that p cannot yield a better result than $r_p(v_j)$ starting from v_j. Together with the observation that pure Nash equilibria always exist [BDS13]—leader equilibria can be formed by all players (playing as if they played their respective two-player discounted sum game)—these lemmas provide the following theorem.

Theorem 5.9. *Pure Nash and leader strategy profiles always exist in MDSGs, and for finding optimal ones, it suffices to consider reward and punish strategies.*

This is particularly interesting when we focus on the implementable strategy profiles. A strategy is *implementable*, if it is realisable with finite memory. We are particularly interested in finite memory strategies with a given small bound b on the memory used. Note that, for reward and punish strategy profiles, we do not have to record the reaction upon deviation, as it is implicitly described by the punishment part. Thus, we do not want to reason about the trivial part in the strategy, and therefore do not count the tiny bit of memory required for the punishment part. This part does not need much memory: it suffices to memorise which player is responsible for the deviation and at which vertex. When we allow for finite memory M, this effectively defines a larger game, on which a memoryless strategy is used. For a game $\mathcal{G} = \langle P, V, \{V_p \mid p \in P\}, \Delta, A, T, \{t_p : V \times A \to \mathbb{Q} \mid p \in P\}\rangle$ and finite memory M with initial memory $m_0 \in M$, we can simply define $\mathcal{G}^M = \langle P, V', \{V'_p \mid p \in P\}, \Delta', A, T', \{t'_p : V' \times A \to \mathbb{Q} \mid p \in P\}\rangle$ with $V' = V \times M$, $V'_p = V_p \times M$, $T' : V' \times A \to V'$ is a set of transitions that maps vertices and actions to vertices, $\Delta'(v, m) = \Delta(v)$ if $m = m_0$ and $\Delta'(v, m) = 0$ otherwise, and $t'_p : ((v, m), a) \mapsto t_p(v, a)$.

Corollary 5.10. *Pure memoryless, and, for a given memory bound b, pure bounded memory Nash and leader strategy profiles always exist in MDSGs, and for finding the optimal ones, it suffices to consider reward and punish strategies.*

Corollary 5.11. *For optimal reward and punish strategy profiles, it suffices to consider the memory needed before deviation, i.e., compliance memory and additional k memory states for the k followers, rather than considering an arbitrary memory M.*

5.6 Constraints for finite pure reward and punish strategy profiles

We first state that optimal strategies exist for all memory bounds. This is a simple implication of Theorem 5.9 and the finite space of candidate strategy profiles.

Lemma 5.12. *For all MDSGs and for all memory bounds, optimal strategy profiles exist among the Nash and leader equilibria.*

We infer a necessary and sufficient constraint system for the strategy profiles in Nash and leader equilibria in MDSGs. Theorem 5.9 implies that, whenever a player deviates at some vertex v, then the remainder of the game resembles a two-player game that starts at v. The player who owns vertex v therefore has an incentive to deviate if, and only if, her payoff from now onwards would be less than the payoff she receives in this underlying two-player game. This provides us with a first necessary constraint, namely

- at any history h that ends in a vertex v owned by player $p \in P$, $\mathbb{E}_p(\sigma, h) \geq r_p(v)$.

For positional (or: memoryless) reward and punish strategies σ, the subtrees in all histories h that end in v coincide, such that one can write $\mathbb{E}_p(\sigma, v)$ instead of $\mathbb{E}_p(\sigma, h)$.

For pure strategies, we require for every vertex v that

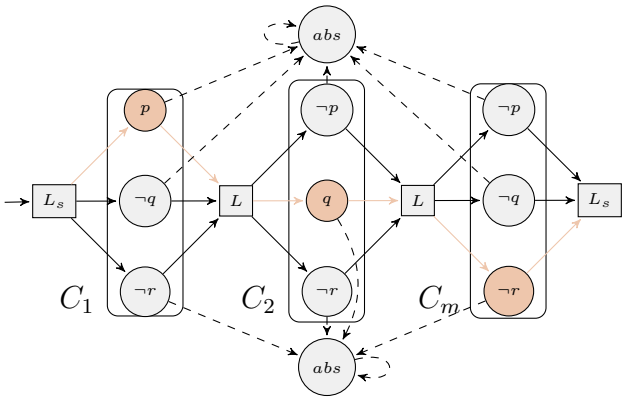

FIGURE 5.8: $C_1, C_2 \ldots\ldots C_m$ are $'m'$ conjuncts each with $'n'$ variables and there are intermediate leader $'L'$ nodes. A path through the satisfying assignment is shown here.

- for all players $p \in P$, $\mathbb{E}_p(\sigma, v) = t_p\big(v, \sigma(v)\big) + \lambda \mathbb{E}_p\Big(\sigma, T\big(v, \sigma(v)\big)\Big)$.

The action $\sigma(v)$ from these constraints refers to the action selected at vertex v by player p in strategy profile σ. Once these actions are fixed, we therefore have a simple linear equation system of full degree, that can easily be solved. To determine if the resulting system is in equilibrium we can simply check if the first set of constraints hold for all players (Nash equilibrium) or for all players but the leader (leader equilibrium). To validate that there is a pure strategy profile of a predefined quality can therefore be checked in non-deterministic polynomial time.

Lemma 5.13. *We can check, if there is a positional strategy profile that meets or exceeds a given threshold t for the leader reward and is a leader or Nash equilibrium, in non-deterministic polynomial time.*

For strategy profiles with bounded memory, we can simply use the extended memory game instead. We can also prove NP-hardness of this problem using standard reduction from 3-SAT as in [UW11, Umm08]. By putting these two together we obtain the following theorem.

Theorem 5.14. *To check, if there is a pure positional or bounded memory strategy profile with fixed memory bound b that meets or exceeds a given threshold t for the leader and is a leader or Nash equilibrium, is NP-complete.*

Proof. In order to establish NP-completeness, we reduce the satisfiability of a 3SAT formula φ over n atomic propositions with m conjuncts to solving a multi-player discounted sum game with $2n + 1$ players and $4m + 5n + 2$ vertices that uses only payoffs -1 and 0. Note that we have not considered discount factor in the proof. We gave a standard reduction, which is similar to the reduction for mean payoff games [GS14], safe for the weights.

We consider the reduction for the example of the 3SAT formula $(p \vee \neg q \vee \neg r) \wedge (\neg p \vee q \vee \neg r) \wedge (\neg p \vee \neg q \vee \neg r)$. The $2n + 1$ players consists of $2n$ players for the $2n$ literals

corresponding to the n variables, and the *leader*, who intuitively tries to validate the formula. The vertices are labelled by their owner.

The payoff for a transition that goes from a vertex owned by a literal player l to a vertex different to the absorbing state 'abs' has a payoff of -1 for the player $\neg l$, and of 0 for every other player. The self-loop at 'abs' has a payoff of -1 for the leader, and of 0 for every other player. The remaining transitions have payoffs of 0 for every player.

If φ is satisfiable, the leader can use a satisfying assignment to determine a cycle through the game graph that does not pass by two vertices owned by opposing literal players p and $\neg p$. All players that make a decision in the unfolding infinite path have a reward of 0, which is the optimal reward obtainable in any play, as there are no positive rewards on any edge. In this case, the leader reward is 0.

Let us assume that φ is unsatisfiable, and the play defined by the leader in a leader strategy profile does not end in the absorbing state. Then there is a first literal l on the play, whose negation $\neg l$ occurs later. The player who owns l will receive a negative return when complying, and hence deviate by moving to the absorbing state. This way, the player receives a reward of 0. Hence, every play in a leader equilibrium for unsatisfiable assignments must end in the absorbing state, which implies that the leader receives a negative reward.

The example is depicted in the Figure 5.8. There are total m conjuncts and each conjunct has n literal variables. Thus, for n propositions, there are $2n$ literal variables. We refer to leader nodes as $'L'$, leader's starting node as 'L_s' and there is one absorbing state 'abs'. 'L_s' is taken as start node with probability 1. The two depicted copies of the vertices 'abs' and 'L_s' each refer to one vertex. As inclusion in NP has been shown in Lemma 5.13 for positional strategies and we can simply use the extended memory game instead, we infer NP-completeness.

□

5.7 Equilibria with extended observations

We first argue why we have focus on the pure strategies only in the previous sections, although mixed strategies are a more general choice. In principle, all arguments from the previous sections also extend to the randomised strategies and strategy profiles, such that one might argue to use the randomised model. The reason why we refrained from doing so is that reward and punish strategies rely on the observability of deviation.

For pure strategies, a deviation by a player can be observed immediately: s/he simply plays a different action than the action defined by the strategy profile assigned by the leader. Let us now consider a reward and punish strategy for the simple game depicted in Figure 5.9. In this example, player 1 owns vertices 1, 2, 3 and 4. Leader owns vertices denoted by l_1 and l_2. Rewards are given on the edges and they are in the order (leader, player 1). When extending the concepts from the previous sections to mixed strategies, the optimal leader strategy profile would be to ask the player 1 to play to vertex l_1 with

a 10% chance, and to l_2 with a 90% chance. When player 1 follows his strategy, the leader pledges to take an edge from l_1 to the vertex 2. While, if player 1 deviates at vertex 1, leader would harm him by taking an edge from l_1 to the vertex 3.

The expected reward for the leader would be 8.9λ, while the expected reward of player 1 would be 0. Player 1 does not benefit from deviation, as, upon deviation, the leader would start to harm him. In particular, she plays to the vertex 3 from l_1.

The catch in this concept is that, with normal observational power (where the players can only observe vertices and actions), the leader (and other players in a multi-player game) would only be able to observe *which* action has been taken, but not *why*. The leader (and other players) cannot distinguish whether the player 1 has moved to l_1 because he conducted a fair experiment with a 10% chance to move to l_1, whose outcome was to move there, or because he simply moved there (with a probability 1) under deviation from the assigned strategy to improve his payoff.

To be able to distinguish compliance from deviation in mixed strategies, we would therefore need a stronger observation model, where the randomised decision (in our example, the decision to play to l_1 with a 10% chance) or the random experiment itself can be observed. Under such an extended observation model, deviation can be observed and we briefly discuss why the results from the previous sections extend to mixed strategies when we assume this observational power.

Also, these temporal dependencies are *not* common in the definition of Nash equilibria. This is also unsurprising when given their origin in the normal-form games[Nas50], where only a single move is played and the concept of history and temporal order of cause and effect does not apply. For us, the concept of observability of deviation by a player outweighs the generality of randomised strategies.

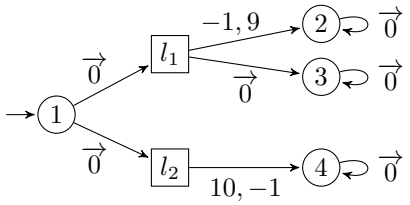

FIGURE 5.9: Unobservability of deviation in mixed strategy with discount factor λ.

The above argument driven by the unobservability of deviation in mixed strategies made us focus only on the pure strategies. However, an alternative to this restriction is to lift the restriction of our observational power: instead of observing the outcome of a decision, we observe the decision itself. Note that this would imply an uncountable set of possible actions, as encoded in the different selected probability distribution over actions, which are possible in every vertex. To justify making this observable, one might think of externalising how to resolve the probabilities, say, by a highly trusted third party. Also, note that allowing for mixed strategies does not remove the usefulness of memory. In the example from Figure 5.10 (player 1 owns vertex 1, leader owns vertex l and rewards are

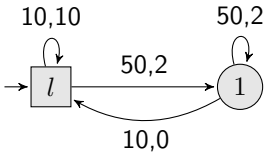

FIGURE 5.10: Leader benefits from memory in mixed strategies.

in the order leader, player 1), when in vertex 1, the leader can only assign an equilibrium strategy to player 1, which is not worse for player 1 than staying in vertex 1. Initially (that is, on the empty history), however, she does not have to take the interest of player 1 into account and can progress to vertex 1 with probability 1. With this motivation in mind, we define *mixed strategies*, which are functions $\sigma_p : (VA)^*V_p \to \mathsf{dist}(A)$ from initial sequences of plays that end in some vertex of player p to a distribution over the actions in A. This implies re-writing the expected reward for player p as follows. We use $\mathbb{E}_p(\sigma, v) = \sum_{a \in A} \sigma(v)(a) \cdot \Big(t_p(v,a) + \lambda \mathbb{E}_p(\sigma, T(v,a))\Big)$ to denote the payoff for player p when starting at vertex v. We again have $\mathbb{E}_p(\sigma) = \sum_{v \in V} \Delta(v) \mathbb{E}_p(\sigma, v)$.

Corollary 5.10 establishes that it suffices to focus only on the reward and punish strategy profiles. This implies a simple constraint system for extended memory games: no player (except for the leader in leader equilibria) may reach a position, where a player would benefit from changing her strategy in a reward and punish strategy profile (that is assigned by the leader). Thus, at every vertex v of the extended memory game (with memory m), it must hold that $\mathbb{E}_p(v, m) \geq \mathbb{E}_{p,2}(v)$. Here, $\mathbb{E}_p(v, m)$ is the expected reward for player p at vertex v in extended memory game and $\mathbb{E}_{p,2}(v)$ is the expected reward for player p at vertex v in two-player game that would result if player p chooses to deviate at vertex v.

We can again use a non-deterministic approach to solve the related decision problem. We can start by guessing a probability distribution at each vertex v_i on all its outgoing actions, and guess, for each action, a target memory value. Once these distributions are fixed, we can again solve the resulting linear equation system, and simply check that it satisfies the constraints from above and meets the required threshold value. Unlike the pure case, where the existence of an optimal solution is implied by the existence of a finite set of possible strategies, we have to provide an argument for the existence of an optimal strategy profile with given memory bound in this setting. According to the constraint system from above, the leader assigns probabilities to the actions and selects the memory updates. If the resulting system complies with the first set of constraints, then it is a Nash (resp. leader) equilibrium. Technically, the converse (only if) does not hold, as these constraints only need to be satisfied by the reachable vertices. We could, however, require the same for unreachable vertices without excluding relevant solutions.

Theorem 5.15. *For multi-player DSGs with perfect observation and predefined memory an optimal leader strategy profile exists.*

Proof. First, we know that a strategy profile that satisfies the constraints exists (cf., Section 5.7). Further, to see that an *optimal* strategy profile exists, we look at the

reward obtained at the different probabilistic transitions. That is, we consider the reward obtained on the different probabilities assigned on different transitions. We define the payoff vector as a direct function on the probability assigned on the transitions and the strategy profile as the set $\mathcal{D}$ (for decisions) of probability vectors over actions, or a finite dimensional closed subset of $[0,1]^n$ for some $n \in \mathbb{N}$. This set of probability distributions over the possible actions gives the expected payoff for all players at all positions of the extended memory game graph (game graph with memory of pre-defined size m) and is defined by the memory copies at all vertices. The resultant payoff for all players at all vertices of the extended game graph is, thus, again a subset of a finite dimensional product of closed and bounded intervals, referred to as $\mathcal{P}$ (for payoff). The intervals are bounded because, if p defines the maximal absolute value of any of the individual payoffs in the discounted sum game, then every payoff must be in the interval $[-\frac{p}{1-\lambda}, \frac{p}{1-\lambda}]$. Given a strategy profile, represented by a $\overrightarrow{d} \in \mathcal{D}$, we can compute the payoffs, represented by a vector $\overrightarrow{p} \in \mathcal{P}$. We represent this by a valuation function $\mathsf{val} : \mathcal{D} \to \mathcal{P}$, that maps each probability vector to a payoff vector. The valuation function is continuous: if the decision vector $\mathcal{D}$ changes only marginally, then the payoff vector $\mathcal{P}$ changes only marginally, too. Thus, if we fix an $\varepsilon > 0$ then we can first choose a natural number l, such that $\sum_{i=l}^{\infty} \lambda^i p < \varepsilon$, and then choose a $\delta \in]0,1[$ such that the change between two consecutive probabilities that is given by $l\big(1-(1-\delta)^l\big) < \frac{\varepsilon}{pl}$ is only marginal. Then, if the absolute sum of changes of all probabilities is below δ, we can estimate the difference by $\sum_{i=0}^{\infty} 2\lambda^i p\big(1-(1-\delta)^i\big)$. For the estimation of this difference, assume that we start with the probability vector $\overrightarrow{d}_m$, which is the point-wise minimum of $\overrightarrow{d}$ and $\overrightarrow{d}'$. Then the difference can be estimated by choosing the joint actions with the probability described in $\overrightarrow{d}_m$, and simply marking the positions with the missing probability (the difference between the sum of the probabilities reflected in $\overrightarrow{d}_m$ and 1 at every position in the extended game) as deviation. This difference is bounded by δ.

The likelihood of being in a state where *no* difference has occurred so far is, after i rounds, $\geq (1-\delta)^i$. The likelihood that a difference has occurred so far can therefore be estimated by $\big(1-(1-\delta)^i\big)$. Using this estimation, we can estimate the difference,

$$\sum_{i=0}^{l-1} 2\lambda^i p\big(1-(1-\delta)^i\big) + \sum_{i=l}^{\infty} 2\lambda^i p\big(1-(1-\delta)^i\big) < \sum_{i=0}^{l-1} 2p\big(1-(1-\delta)^l\big) + 2\varepsilon < 4\varepsilon,$$

where the first inequality uses the definition of l, $\lambda^i \leq 1$, and $1-(1-\delta)^i < 1-(1-\delta)^l$, while the second estimation uses the definition of δ. Thus, $\forall \varepsilon > 0\ \exists \delta > 0$ such that $\|\overrightarrow{d} - \overrightarrow{d}'\| < \delta$ implies $\|\mathsf{val}(\overrightarrow{d}) - \mathsf{val}(\overrightarrow{d}')\| < 4\varepsilon$. The subset $\mathcal{C} \subseteq \mathcal{P}$ of the set of payoffs that comply with the constraint system is obviously still closed, as it is still a product of finitely many closed intervals. (Only the lower bound of these intervals may have changed.) As val is continuous, the preimage $\mathcal{D}'$ of the closed and bounded set $\mathcal{C}$ is closed and bounded. When val is restricted to $\mathcal{D}'$, then the maximum w.r.t. the value of the leader in the initial state exists. That is, the supremum is taken for some value. □

5.8 Discussion

We have established the importance of memory in forming an optimal strategy profile with maximal reward for the leader. We have discussed a memory model for reasoning with bounded resources and show that it suffices to consider memory needed before deviation, i.e., memory for the main path only. We further show that we need only k memory states for k followers. We have given the non-deterministic approach for the construction of optimal strategy profiles (which are Nash or leader equilibria and meet a given threshold for the overall payoff of the leader) and the NP-hardness proof. Possible future work could be to implement our non-deterministic approach for solving these games in SMT solvers like Yices [dMDS07, yic] and see how well they perform on small examples.

We have further observed that the techniques from incentive equilibria in multi-player non-terminating games (as discussed in Chapter 4) can also be applied to discounted sum games. The gain that the leader would receive from incentive equilibria is natural in these games as well. Incentives can help to improve the leader return in discounted sum games just as they do in mean-payoff games. For a strategy profile, the leader might like to wait initially (paying an incentive of 0) until the payoff is discounted enough to allow her to incentivise with maximal power: if we denote by $\|A\|$ the maximal absolute difference between any two payoff values of $\mathcal{G}$, then the leader can wait till she can incentivise with the value $\|A\|$.

We now refer to an example from the Figure 5.11. It shows a discounted sum game with two players – player 1 (or: follower) and the leader. We assume the discount factor to be $\frac{1}{2}$. Vertex 1 and vertex 2 are owned by player 1. Vertex 3 and vertex 4 have only one self loop such that it does not matter who owns them. The rewards to players are given on the edges in the order (player 1, leader). Initially, player 1 moves the token to vertex 2. At vertex 2, player 1 has two choices – either to move the token to vertex 3 or to move to vertex 4. Note that the leader gets the maximal reward if player 1 at vertex 2 moves to vertex 4. The leader can therefore incentivise player 1 for the strategy profile $\langle 1 \cdot 2 \cdot 4^{\omega} \rangle$.

We note that the leader can distribute the incentive amount to be given to her follower in different possible ways. For example, she can incentivise her follower on every edge with an incentive of $\frac{1}{4}$. The follower reward from this incentive strategy profile would be 0.125 while his reward from a Nash or leader strategy profile is 0. Note that a rational leader would not distribute the incentive this way as the follower's reward is more than 0 here. Another way to incentivise the follower for the strategy profile $\langle 1 \cdot 2 \cdot 4^{\omega} \rangle$ is as follows. The leader chooses to give an incentive of 0 on the edge $(1, 2)$ and an incentive of 1 on the edge $(2, 4)$ and gives no incentive thereafter. Although the follower return is 0 here and the leader return is 0.5, we note that a better solution exists, wherein the leader can incentivise her follower with the maximal bribe possible.

To incentivise with the maximal possible reward, the leader would pay nothing on the edges $(1, 2)$ and $(2, 4)$ and also not for the first iteration at vertex 4. However, from

the second iteration onwards, the leader can incentivise her follower fully by giving an incentive of 2 at every step. The follower and the leader return from this incentive strategy profile are 0 and 0.5 respectively. The order in which incentives are paid therefore matters in discounted sum games.

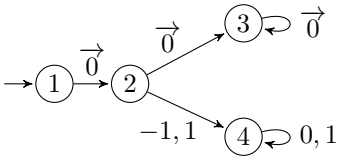

FIGURE 5.11: Use of incentives in discounted sum game.

Chapter 6

Summary and conclusions

We summarise this thesis in this chapter and give conclusions.

6.1 Summary

We have studied the construction of optimal strategy profiles in various game settings where we aim at maximising the leader's reward. We have introduced leader equilibria in multi-player mean-payoff games and in multi-player discounted sum games. We have introduced incentive equilibria as an extension of leader equilibria in bi-matrix games and in multi-player mean-payoff games.

Our simple examples (cf. Figure 4.6 and Figure 5.2) show that, even in multi-player non-terminating games, no Nash equilibrium be superior to leader equilibria. Thus, the leader's reward from any leader equilibrium can only be greater or equal to her reward from any Nash equilibrium. The concept of incentive equilibria can be applied to different types of games, including bi-matrix and non-terminating multi-player games. In Chapter 1, we have introduced the general concepts used in game theory like equilibrium notions, their complexity, etc. We have also given an informal introduction to various game types that we have considered in this thesis.

We have started with the discussion of incentive and other related equilibria in bi-matrix games in Chapter 3. We have considered various follower models assuming both considerate follower and a non-considerate or an adversarial follower. The leader would behave ex-aequo, i.e., the leader would break ties in favour of the follower and against the follower in the respective model. We have made some notable observations regarding the impact of choosing a hostile over a friendly follower model. If the follower chooses to behave non-friendly, then it is his loss only and not the leader's.

That is, in a non-friendly follower model, the cost that the leader has to pay is either arbitrarily small or nothing. The follower would lose either an arbitrary increase in his incentive amount or suffers seriously otherwise. If secure incentive equilibria do not exist, then the leader only needs to increase the incentive amount by $\epsilon > 0$ value. This is to make the incentive equilibrium secure for her. Whereas if secure incentive equilibria exist, then there is no change in the leader's payoff. However, the follower

suffers severely here from his own non-friendly choice. The worst case for the follower is when secure incentive equilibria exist that are not friendly. The leader would then break the ties against the follower making him suffer potentially (cf. example from Table 3.15).

We have shown that the construction of incentive equilibria, friendly incentive equilibria, leader equilibria, and friendly leader equilibria is tractable. We have discussed a tractable technique to determine if a secure incentive equilibrium exists (and to compute its value when they exist). Computing ϵ-incentive equilibria for the general case when secure incentive equilibria do not exist is also tractable.

We have implemented an algorithm to compute friendly incentive equilibria and friendly leader equilibria. The algorithm computes the leader's reward in friendly incentive and friendly leader equilibria. Our results are based on different sets of benchmarks considering both integer payoffs and continuous payoffs. The results have confirmed that incentive equilibrium is superior to leader equilibrium that is superior to Nash equilibrium. This observation holds for all cases w.r.t. the leader reward. However, for the follower's reward, the observation holds on average. As there exist examples where the follower might lose in incentive equilibria as compared to leader or Nash equilibria.

We moved on to multi-player non-terminating games and established the existence of leader equilibria and incentive equilibria in multi-player mean-payoff games in Chapter 4. Unsurprisingly, the complexity of finding leader equilibria equals incentive equilibria in these games and is NP-complete. We have given a non-deterministic approach to compute the optimal strategy profiles. That is, to form an optimal leader resp. incentive strategy profile, we have guessed the set of reachable vertices and the strongly connected sets of vertices and the constraint system defined by them. We then solve the number of linear programmes formed with an objective of maximising the leader's reward. The constraint system that gave maximal reward to the leader form an optimal solution. Note that the linear programmes were polynomial in game size and thus has a polynomial time solution. The solution is therefore verifiable in polynomial time.

We have discussed the parametrized complexity for the cases where the number of players is bounded. We show for these cases that the problem lies in the same complexity class as for solving two-player mean-payoff games by giving a polynomial reduction to solving two-player games. For a game graph with n vertices and k players, there are at most n^k many thresholds to consider. These different thresholds refer to the different values from underlying two-player games that may start at any vertex.

We therefore needed an oracle to solve two-player games. We have used an efficient algorithm from [Sch08] to determine the mean-values at the vertices and thus to solve the two-player mean-payoff games. The algorithm is used first for the quantitative analysis of mean-payoff games and then to solve them. Our results (cf. Figures 4.10 and 4.11) have confirmed that the leader's reward is greater in incentive equilibria when compared to leader equilibria.

We then established the existence of optimal bounded memory leader strategy profiles in discounted sum games in Chapter 5. We have observed that positional leader equilibria

exist and we also infer the existence of optimal positional and pure memory bounded strategies. We have argued that (and why) the detectability of deviation makes the restriction to pure strategies a natural choice. We show that the related decision problem (Is there a Nash resp. leader equilibrium that provides a payoff that meets or exceeds a given threshold?) is NP-complete. We have also discussed the extension to mixed strategies and the extension of the observation model that is needed to make such strategies reasonable. Further, we have established the usefulness of memory in forming optimal leader strategy profiles. Unsurprisingly, more memory would always help. Our simple example from Figure 5.4 show that there is no upper bound that can be inferred from the structure of the game such that one could know that memory of this size would suffice if unbounded memory would. However, we have observed that in some cases infinite memory would also be needed (cf., Theorem 5.4, Theorem 5.5 and Theorem 5.6).

6.2 Conclusions and Future work

In this thesis, we have introduced incentive equilibria that form a natural extension of Stackelberg or leader equilibria. In a leader equilibrium, the leader is able to commit to a strategy profile beforehand and her follower only needs to observe the leader's actions and then moves accordingly. The model therefore requires basic assumption that the leader needs to communicate her strategy effectively to the follower and that there is mutual trust between players. While the communication is always from the leader's side, note that the trust is usually required mutually from both sides. That is, not only the leader but the follower should also be trustworthy. The follower needs to trust that the leader would commit to a strategy profile and the follower should also be trustworthy in that he follows the strategy profile that is assigned by the leader.

The leader who can communicate her strategy is further capable of offering more to her follower. She can therefore also announce an incentive to be given to her follower for an assigned strategy profile. The follower who can trust the leader on a strategy profile can also trust the leader for the incentives that are announced. That is, it is not the incentives that raises the question of trust but it is the commitment to a strategy profile. Giving incentives to follow a strategy profile makes a binding commitment between players. The leader offers a better return to her follower and in return only expect her follower to be friendly and trustworthy. The leader is obliged to make follower friendly choices for a friendly follower. Thus, for all those strategy profiles where the leader gets an equal return, the leader would select one that also gives maximal reward to the follower.

On the other hand, if the follower is not trustworthy and acts as an adversary, then the prime concern for the leader becomes to secure her payoff. The leader is no longer able to make follower friendly choices and would assign a strategy profile that is secure for herself. The amount that the leader has to pay to assign a secure equilibrium is negligible or sometimes nothing. However, it might affect the payoff of the hostile

follower adversely. The leader would now break the ties that are not in the favour of the follower. It is therefore an interesting observation that it is really the follower who benefits significantly for being friendly towards the leader.

Although bi-matrix games form a natural model to study these equilibria, note that the use of leader equilibria and incentive equilibria is not limited to these games only. These concepts can be well applied to multi-player games that are of infinite duration. We show in the case of multi-player mean-payoff games that how incentivising the followers result in a strategy profile that is beneficial for the leader and is also more stable. The leader here needs to incentivise every non-deviating follower who complies with the assigned strategy profile. It also accounts for higher stability in that the resulting strategy profile is now subgame perfect relative to the leader. This is because the leader is allowed to benefit from deviation so subgame perfect criterion does not hold for her. Thus, the leader might improve over her payoff by giving small incentives to her followers while at the same time also leading to a more stable strategy profile.

Incentive equilibria are therefore a generalisation of leader equilibria where we aim at maximising the payoff of one player. A rational leader might use these techniques in order to assign an optimal and stable strategy profile with maximal reward for her. It allows to mix well a rational environment – several rational components following their own objectives – with a rational controller. The results are also rational in that at the least followers are assigned a strategy profile that they do not benefit from deviation.

The techniques that we have proposed and discussed in this thesis are hence instrumental in maximising the payoff of a central player. It also wards away any possible side effects like effecting the utility of other players or effecting the stability of a strategy profile. The return to every other player in an incentive or leader equilibrium is at least as good as their return from a Nash equilibrium. For multi-player games, perfectly incentivised strategy profiles are also subgame perfect and therefore more stable. Security also comes very cheap here as compared to secure Nash or leader equilibria. This is because in incentive equilibria the leader can simply increase the incentive by a marginal amount in order to get a secure strategy profile. We have therefore established the use of incentives in varied game settings.

We have however considered only pure strategies in non-terminating multi-player games as we rely on the use of reward and punish strategies. Although mixed strategies are more general but it becomes difficult to distinguish that whether a follower is deviating or complying with the assigned strategy. We therefore need a higher observational model to make this observable and for this one might need to think how to resolve the probabilities (for example, externalising this to a trusted party). Our example from the Figure 5.9 shows this unobservability of deviation in mixed strategies and shows that it is not immediately clear to the leader that whether the follower is actually following a strategy profile or whether the follower is deviating. Nevertheless our observation that the leader benefits from memory in an optimal leader strategy profile holds for

mixed strategies as well (cf. example from Figure 5.10). One would need a more precise observation model as discussed briefly in Chapter 5.

There is a lot of scope for future research where we can consider mild variants of the proposed game settings. For example, we can introduce formal contracting in the game where the leader would formally enter into contract with other players and ask them to follow the contract. While incentivising for a strategy profile is straightforward, the introduction of formal contracts would bring other aspects to consider. We may explore the side conditions associated with formal contracting and how the players behave in such cases. In the proposed model, incentivising is only unidirectional so it would be interesting to see if more than one players are in the position of offering side payments and how they achieve an equilibrium. The game model could be studied in stages where players can offer side-payments to each other and their payoff matrices are altered along with time. Other interesting aspects that invite future research is regarding the complexity of our techniques and algorithms. Is there a way to improve upon the complexity of the techniques we study? Can our techniques do better? Is there a more efficient way in which the leader can assign strategies? It would also be interesting to see how the leader derives benefit from optimal strategy profiles in qualitative game settings.

Appendix A

Appendix

A.1 Leader equilibria

In this appendix, we give in detail the constraint system needed to compute leader equilibria and friendly leader equilibria. Existence of leader equilibria is well known in the literature and their computation is also known. Conitzer and Sandholm [CS06] have given how to compute optimal strategies to commit to in Stackelberg models. We adjust the constraint system given here to the one that is given in Section 3.8. The adjusted constraint system is used to find out a secure leader equilibria, in case it exists.

Similar to the incentive equilibria, leader equilibria are also defined only among the simple leader strategy profiles. We define them by a set of linear inequations.

Theorem A.1. *For a bi-matrix game $\mathcal{G}(A, B)$, $(\langle \sigma, j \rangle)$ is a simple leader strategy profile if, and only if, $\sigma B^{\rightarrow}_{j} \geq \sigma B^{\rightarrow}_{i}$ holds for all pure strategies i=1, . . . , n of the follower.*

Proof. If $\langle \sigma, j \rangle$ is a simple leader strategy profile, then in particular changing the strategy to a different pure strategy i cannot be beneficial for the follower. Consequently, it holds for all $i = 1, \ldots, n$ that $\sigma B^{\rightarrow}_{j} \geq \sigma B^{\rightarrow}_{i}$. If it holds for all $i = 1, \ldots, n$ that $\sigma B^{\rightarrow}_{j} \geq \sigma B^{\rightarrow}_{i}$, then we note that, for any follower strategy δ, the payoff under σ is an affine combination $\mathsf{fpayoff}(B; \langle \sigma, \delta \rangle) = \sum_{i \in S} \delta(i) \cdot \sigma B^{\rightarrow}_{i}$ of the payoffs for the individual pure strategies. Using $\sigma B^{\rightarrow}_{j} \geq \sigma B^{\rightarrow}_{i}$, we get $\mathsf{fpayoff}(B; \langle \sigma, \delta \rangle) = \sum_{i \in S} \delta(i) \cdot \sigma B^{\rightarrow}_{i} \leq \sum_{i \in S} \delta(i) \cdot \sigma B^{\rightarrow}_{j} = \sigma B^{\rightarrow}_{j} = \mathsf{fpayoff}(B; \langle \sigma, j \rangle, \beta)$.

A.1.1 Computing simple leader equilibria

We give here the constraint system for the computation of leader equilibria. We use the definitions from Section 3.4 to define bi-matrix game and the leader strategy profiles. We next define a constraint system $\mathcal{L}_j^{\mathcal{G}(A,B)}$ for each pure follower strategy j. For a simple leader strategy profile $\langle \sigma, j \rangle$, strategy σ is described as the vector σ by $\sigma^T = (p_1, \ldots, p_m)$. The constraint system $\mathcal{L}_j^{\mathcal{G}(A,B)}$ consists of $m + n$ constraints, where $m + 1$ constraints describe that σ is a strategy,

- $\sum_{i=1}^{m} p_i = 1$ (the sum of the weights is 1) and

- the m non-negativity requirements $p_i \geq 0$ for $1 \leq i \leq m$,

and $n-1$ constraints reflect the conditions on a leader equilibrium. That is, $\mathcal{L}_j^{\mathcal{G}(A,B)}$ contains $n-1$ constraints of the form

- $\sum_{k=1}^{m}(b_{kj} - b_{ki})p_k \geq 0$,

one for each $i \neq j$ with $1 \leq i \leq n$.

This gives us the following corollary.

Corollary A.2. *The solutions to $\mathcal{L}_j^{\mathcal{G}(A,B)}$ describe the set of leader strategies σ, such that $\langle \sigma, j \rangle$ is a leader equilibrium for $\mathcal{G}(A,B)$.*

Note that the constraints do *not* depend on the payoff matrix of the leader. The formalisation of the objective function, however, depends on the payoff matrix of the leader. We denote with $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ the linear programming problem that consists of the constraint system $\mathcal{L}_j^{\mathcal{G}(A,B)}$ and the objective function $\sum_{k=1}^{m} a_{kj}p_k \mapsto \max$, where $\sum_{k=1}^{m} a_{kj}p_k = \mathsf{lpayoff}(A, \langle \sigma, j \rangle)$ is the payoff the leader obtains for such a simple leader strategy profile $\langle \sigma, j \rangle$ with $\sigma = (p_1, \ldots, p_m)^T$.

Corollary A.3. *The solutions to $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ are the leader strategies σ, such that $\langle \sigma, j \rangle$ is a simple leader strategy profile for $\mathcal{G}(A,B)$, that is optimal among the leader strategy profiles with the pure strategy j for the follower.*

Corollary A.4. *To find a leader equilibrium for a game $\mathcal{G}(A,B)$, it suffices to solve the linear programming problems $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ for all $1 \leq j \leq n$, to select a j with maximal solution among them, and to use a solution σ_j of $\mathcal{LP}_j^{\mathcal{G}(A,B)}$. For the leader strategy σ described by this solution, $\langle \sigma, j \rangle$ is a leader equilibrium.*

Proof. Obviously, $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ has *some* solution iff $\mathcal{L}_j^{\mathcal{G}(A,B)}$ is satisfiable, and thus, by Corollary A.2, if there is a leader strategy σ such that $\langle \sigma, j \rangle$ is a simple leader strategy profile. In this case, a solution to $\mathcal{LP}_j^{\mathcal{G}(A,B)}$ reflects an optimal simple leader strategy profiles, the simple leader strategy profiles of the form $\langle \sigma, j \rangle$; in particular, such an optimum exists. Thus, the returned result is the optimal solution among all simple leader strategy profiles.

As linear programming problems can be solved in polynomial time [Kha79, Kar84], finding leader equilibria is tractable.

A.1.2 Friendly leader equilibria

Here, the leader follows a secondary objective of being benign to her follower. We notice that it is cheap and simple to determine the value $v_{\max}^l$ a leader can at most acquire in a leader equilibrium. We next determine the highest payoff $v_{\max}^f$ for her follower in a leader equilibrium with leader payoff $v_{\max}^l$, i.e., to construct *friendly* leader equilibria. We note that friendliness does not come to the cost of simplicity.

Theorem A.5. *For every bi-matrix game $\mathcal{G}(A, B)$ and every leader equilibrium $\langle \sigma, \delta \rangle$ with payoff v for the follower, it holds for all pure strategies $j \in \mathsf{support}(\delta)$ that $\langle \sigma, j \rangle$ is a leader equilibrium with payoff v for the follower.*

Proof. Let $S = \mathsf{support}(\delta)$ be the support of δ and let $v^l_{\max} = \mathsf{lpayoff}(A; \langle \sigma, \delta \rangle)$ be the leader payoff for $\langle \sigma, \delta \rangle$. It holds for all $j \in S$ that $\langle \sigma, j \rangle$ is a simple leader strategy profile with the same payoff $\mathsf{fpayoff}(B; \langle \sigma, j \rangle) = \mathsf{fpayoff}(B; \langle \sigma, \delta \rangle)$ for the follower. To establish that $\langle \sigma, j \rangle$ is also a leader equilibrium, we first note that the leader payoff cannot be higher than in a leader equilibrium, such that $\mathsf{lpayoff}(A; \langle \sigma, j \rangle) \leq v^l_{\max}$ holds for all $j \in S$. Assuming for contradiction that there is an $j \in S$ with $\mathsf{lpayoff}(A; \langle \sigma, j \rangle) < v^l_{\max}$ would, together with the previous observation that $\mathsf{lpayoff}(A; \langle \sigma, j \rangle) \leq v^l_{\max}$ holds for all $j \in S$, imply that the affine combination $\sum_{j \in S} \delta(j) \cdot \mathsf{fpayoff}(B; \langle \sigma, j \rangle) < v^l_{\max}$ of these values defined by δ is strictly smaller than $v^l_{\max}$, and would therefore lead to a contradiction.

Similar to ordinary leader equilibria, we therefore focus on pure follower strategies when seeking friendly leader equilibria.

Recall that each constraint system $\mathcal{L}_j^{\mathcal{G}(A,B)}$ describes the set of leader strategies σ, such that $\langle \sigma, j \rangle$ is a simple leader strategy profile. In order to be a leader equilibrium, it also has to satisfy the optimality constraint

$$\sum_{k=1}^{m} a_{kj} p_k \geq v^l_{\max}.$$

We refer to the extended constraint system by $\mathcal{LE}_j^{\mathcal{G}(A,B)}$. By Corollary A.2, the set of solutions to this constraint system is non-empty iff there is a leader equilibrium of the form $\langle \sigma, j \rangle$.

Corollary A.6. *The solutions to $\mathcal{LE}_j^{\mathcal{G}(A,B)}$ describe the set of leader strategies σ, such that $\langle \sigma, j \rangle$ is a leader equilibrium for $\mathcal{G}(A, B)$.*

We now extend the constraint system $\mathcal{LE}_j^{\mathcal{G}(A,B)}$ to an extended linear programming problem $\mathcal{LEP}_j^{\mathcal{G}(A,B)}$ by adding an objective to maximise the follower return, i.e.,

$$\sum_{k=1}^{m} b_{kj} p_k \mapsto \max.$$

Corollary A.7. *The solutions to $\mathcal{LEP}_j^{\mathcal{G}(A,B)}$ describe the set of leader strategies σ, such that $\langle \sigma, j \rangle$ is a leader equilibrium for $\mathcal{G}(A, B)$ and, if such a solution exists, such that the return for the follower is maximal among them.*

Corollary A.8. *To find a friendly leader equilibrium for a game $\mathcal{G}(A, B)$, it suffices to solve the linear programming problems $\mathcal{LEP}_j^{\mathcal{G}(A,B)}$ for all $1 \leq j \leq n$ and to select a i with maximal solution and a solution of $\mathcal{LEP}_j^{\mathcal{G}(A,B)}$ among them. For the leader strategy σ described by this solution, $\langle \sigma, i \rangle$ is a friendly leader equilibrium.*

Proof. $\mathcal{LEP}_j^{\mathcal{G}(A,B)}$ has *some* solution iff $\mathcal{LE}_j^{\mathcal{G}(A,B)}$ is satisfiable, and thus, by Corollary A.6, if there is a leader strategy σ such that $\langle \sigma, j \rangle$ is a leader equilibrium.

In this case, a solution to $\mathcal{LEP}_j^{\mathcal{G}(A,B)}$ reflects a leader equilibrium with the maximal follower payoff among the leader equilibria of the form $\langle \sigma, j \rangle$; in particular, such an optimum exists. Thus, the returned result is the optimal solution among all simple leader equilibria.

Assuming that a better non-simple friendly leader equilibrium exists implies by Theorem A.5 that a better simple friendly leader equilibrium exists as well, and thus leads to a contradiction.

Note that the existence of some simple optimal leader equilibrium is implied by Corollary A.4.

As linear programming problems can be solved in polynomial time [Kha79, Kar84], finding friendly leader equilibria is tractable.

Corollary A.9. *A simple friendly leader equilibrium can be constructed in polynomial time.*

Bibliography

[ASA10] Ashton Anderson, Yoav Shoham, and Alon Altman. Internal implementation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 191–198. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[Aum74] Robert J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, March 1974.

[BCD+11] Lubos Brim, Jakub Chaloupka, Laurent Doyen, Raffaella Gentilini, and Jean-François Raskin. Faster algorithms for mean-payoff games. *Formal Methods in System Design*, 38(2):97–118, 2011.

[BDS13] Thomas Brihaye, Julie De Pril, and Sven Schewe. Multiplayer cost games with simple Nash equilibria. In *Proceedings of the Symposium on Logical Foundations of Computer Science (LFCS 2013), January 6–8, San Diego, California, USA*, volume 7734 of *Lecture Notes in Computer Science*, pages 59–73. Springer-Verlag, 2013.

[BEF+11] Endre Boros, Khaled M. Elbassioni, Mahmoud Fouz, Vladimir Gurvich, Kazuhisa Makino, and Bodo Manthey. Stochastic mean payoff games: Smoothed analysis and approximation schemes. In *Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP 2011), July 4-8, Zürich, Switzerland*, volume 6755 of *Lecture Notes in Computer Science*, pages 147–158. Springer, 2011.

[BK78] Truman Bewley and Elon Kohlberg. On stochastic games with stationary optimal strategies. *Mathematics of Operations Research*, 3(2):104–125, 1978.

[BK13] Kimmo Berg and Mitri Kitti. Computing equilibria in Discounted 2×2 supergames. *Comput. Econ.*, 41(1):71–88, January 2013.

[BV07] Henrik Björklund and Sergei Vorobyov. A combinatorial strongly subexponential strategy improvement algorithm for mean payoff games. *Discrete Applied Mathematics*, 155(2):210–229, 2007.

[CFW13] Krishnendu Chatterjee, Vojtech Forejt, and Dominik Wojtczak. Multi-objective discounted reward verification in graphs and mdps. In *LPAR*, volume 8312 of *Lecture Notes in Computer Science*, pages 228–242, 2013.

[CHJ05a] Krishnendu Chatterjee, Thomas A. Henzinger, and Marcin Jurdzinski. Games with Secure Equilibria,. In FrankS. de Boer, MarcelloM. Bonsangue, Susanne Graf, and Willem-Paul de Roever, editors, *Formal Methods for Components and Objects*, volume 3657 of *Lecture Notes in Computer Science*, pages 141–161. Springer Berlin Heidelberg, 2005.

[CHJ05b] Krishnendu Chatterjee, Thomas A. Henzinger, and Marcin Jurdzinski. Mean-payoff parity games. In *Proceedings of the 20th IEEE Symposium on Logic in Computer Science (LICS 2005), 26-29 June 2005, Chicago, IL, USA*, pages 178–187, 2005.

[CHJ06] Krishnendu Chatterjee, Thomas A. Henzinger, and Marcin Jurdzinski. Games with secure equilibria. *Theoretical Computer Science*, 365(12):67 – 82, 2006. Formal Methods for Components and Objects.

[CHJS11] Krishnendu Chatterjee, Thomas A Henzinger, Barbara Jobstmann, and Rohit Singh. Quasy: Quantitative synthesis tool. In *Tools and Algorithms for the Construction and Analysis of Systems*, pages 267–271. Springer, 2011.

[Con93] Anne Condon. On algorithms for simple stochastic games. In *Advances in Computational Complexity Theory, volume 13 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 51–73. American Mathematical Society, 1993.

[CPR14] C. Comin, R. Posenato, and R. Rizzi. A tractable generalization of simple temporal networks and its relation to mean payoff games. In *2014 21st International Symposium on Temporal Representation and Reasoning*, pages 7–16, Sept 2014.

[CS06] Vincent Conitzer and Tuomas Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, EC '06, pages 82–90, New York, NY, USA, 2006. ACM.

[dAFH+04] Luca de Alfaro, Marco Faella, Thomas A. Henzinger, Rupak Majumdar, and Marille Stoelinga. Model checking discounted temporal properties. In Kurt Jensen and Andreas Podelski, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 2988 of *Lecture Notes in Computer Science*, pages 77–92. Springer Berlin Heidelberg, 2004.

[dAHM03] Luca de Alfaro, Tom Henzinger, and Rupak Majumdar. Discounting the future in systems theory. In *Proc. of the 30th International Colloquium on Automata, Languages, and Programming (ICALP)*, 2003.

[DGP09] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *Commun. ACM*, 52(2):89–97, 2009.

[dMDS07] Leonardo Mendonça de Moura, Bruno Dutertre, and Natarajan Shankar. A tutorial on satisfiability modulo theories. In *Proceedings of the 19th International Conference oon Computer Aided Verification (CAV 2007), July 3-7, Berlin, Germany*, volume 4590 of *Lecture Notes in Computer Science*, pages 20–36. Springer, 2007.

[DPFK+14] Julie De Pril, Jnos Flesch, Jeroen Kuipers, Gijs Schoenmakers, and Koos Vrieze. Existence of secure equilibrium in multi-player games with perfect information. In Erzsbet Csuhaj-Varj, Martin Dietzfelbinger, and Zoltn sik, editors, *Mathematical Foundations of Computer Science 2014*, volume 8635 of *Lecture Notes in Computer Science*, pages 213–225. Springer Berlin Heidelberg, 2014.

[EH86] Harri Ehtamo and Raimo P. Hmlinen. On Affine Incentives for Dynamic Decision Problems. In *Dynamic Games and Applications in Economics*, volume 265 of *Lecture Notes in Economics and Mathematical Systems*, pages 47–63. Springer Berlin Heidelberg, 1986.

[EH89] Harri Ehtamo and Raimo P. Hmlinen. Incentive strategies and equilibria for dynamic games with delayed information. *Journal of Optimization Theory and Applications*, 63(3):355–369, 1989.

[EM79] A. Ehrenfeucht and J. Mycielski. Positional strategies for mean payoff games. *International Journal of Game Theory*, 8(2):109–113, 1979.

[EY10] Kousha Etessami and Mihalis Yannakakis. On the complexity of Nash equilibria and other fixed points. *SIAM J. Comput.*, 39(6):2531–2597, 2010.

[Fin64] A. M. Fink. Equilibrium in a stochastic n-person game. 28:89–93, 1964.

[Fri71] James W. Friedman. A Non-cooperative Equilibrium for Supergames. *The Review of Economic Studies*, 38(1):1–12, 1971.

[Fri77] James W. Friedman. *Oligopoly and the theory of games.* Advanced textbooks in economics. North-Holland Pub. Co., 1977.

[GO14] Vladimir Gurvich and Vladimir Oudalov. On Nash-solvability in pure stationary strategies of the deterministic n-person games with perfect information and mean or total effective cost. *Discrete Applied Mathematics*, pages 131–143, 2014.

[GS14] Anshul Gupta and Sven Schewe. Quantitative verification in rational environments. In *Proceedings of the 21st International Symposium on Temporal Representation and Reasoning (TIME 2014), 8–10 September, Verona, Italy*, pages 123–131. IEEE Computer Society Press, 2014.

[GS15] Anshul Gupta and Sven Schewe. It pays to pay in bi-matrix games: a rational explanation for bribery. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), May 4-8, Istanbul, Turkey*, pages 1361–1369. ACM, 2015.

[GST+16] Anshul Gupta, Sven Schewe, Ashutosh Trivedi, Maram Sai Krishna Deepak, and Bharath Kumar Padarthi. Incentive stackelberg mean-payoff games. In *Proceedings of the 14th International Conference on Software Engineering and Formal Methods*, 2016.

[GSW15] Anshul Gupta, Sven Schewe, and Dominik Wojtczak. Making the best of limited memory in multi-player discounted sum games. In Javier Esparza and Enrico Tronci, editors, Proceedings Sixth International Symposium on *Games, Automata, Logics and Formal Verification,* Genoa, Italy, 21-22nd September 2015, volume 193 of *Electronic Proceedings in Theoretical Computer Science*, pages 16–30. Open Publishing Association, 2015.

[GZ04] Hugo Gimbert and Wieslaw Zielonka. When can you play positionally? In Jirí Fiala, Václav Koubek, and Jan Kratochvíl, editors, *MFCS*, volume 3153 of *Lecture Notes in Computer Science*, pages 686–697. Springer, 2004.

[Hen13] ThomasA. Henzinger. Quantitative reactive modeling and verification. *Computer Science - Research and Development*, 28(4):331–344, 2013.

[Jur98] Marcin Jurdziński. Deciding the winner in parity games is in UP ∩ co-UP. *Information Processing Letters*, 68(3):119–124, November 1998.

[Kar84] N. Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing (STOC 1984), 30 April – 2 May, Washington, DC, USA*, pages 302–311. ACM Press, 1984.

[KFSV09] J. Kuipers, J. Flesch, G. Schoenmakers, and K. Vrieze. Pure subgame-perfect equilibria in free transition games. *European Journal of Operational Research*, 199(2):442 – 447, 2009.

[Kha79] L. G. Khachian. A polynomial algorithm in linear programming. *Doklady Akademii Nauk SSSR*, 244:1093–1096, 1979.

[Leh90] E. Lehrer. Nash equilibria of n-player repeated games with semi-standard information. *International Journal of Game Theory*, 19(2):191–217, 1990.

[LH64] Carlton E Lemke and Joseph T Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for Industrial & Applied Mathematics*, 12(2):413–423, 1964.

[LMC+12] Joshua Letchford, Liam MacDermed, Vincent Conitzer, Ronald Parr, and Charles L. Isbell. Computing stackelberg strategies in stochastic games. *SIGecom Exch.*, 11(2):36–40, December 2012.

[MKP] Berkelaar M., Eikland K., and Notebaert P. lp_solve, a Mixed Integer Linear Programming (MILP) solver. Website.

[MMT13] Richard D. Mckelvey, Andrew M. Mclennan, and Theodore L. Turocy. Gambit: Software Tools for Game Theory, http://www.gambit-project.org. Technical report, Version 13.1.0., 2013.

[MOJ05] Simon Wilkie Matthew O. Jackson. Endogenous games and mechanisms: Side payments among players. *The Review of Economic Studies*, 72(2):543–566, 2005.

[mpg] Mmpg solver: `http://www.cse.iitb.ac.in/~trivedi/mmpgsolver`.

[MS96] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996.

[MT04] Dov Monderer and Moshe Tennenholtz. K-implementation. *Journal of Artificial Intelligence Research*, 21:37–62, 2004.

[Nas50] John Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[NRTV07] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic Game Theory.* Cambridge University Press Cambridge, 2007.

[OR94] Martin J. Osborne and Ariel Rubinstein. *A course in game theory.* The MIT Press, Cambridge, USA, 1994. electronic edition.

[PGS] Pg solver: `https://github.com/tcsprojects/pgsolver/`.

[PR89] Amir Pnueli and Roni Rosner. On the synthesis of a reactive module. In *Proceedings of the 16th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, pages 179–190. ACM, 1989.

[Rab93] Matthew Rabin. Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83(5):1281–1302, December 1993.

[Rou10] Tim Roughgarden. Algorithmic game theory. pages 78–86, 2010.

[RW89] Peter JG Ramadge and W Murray Wonham. The control of discrete event systems. *Proceedings of the IEEE*, 77(1):81–98, 1989.

[Sch08] Sven Schewe. An optimal strategy improvement algorithm for solving parity and payoff games. In *Proceedings of the 17th Annual Conference of the European Association for Computer Science Logic (CSL 2008), 15–19 September, Bertinoro, Italy*, volume 5213 of *Lecture Notes in Computer Science*, pages 368–383. Springer-Verlag, 2008.

[Sch09] Sven Schewe. From parity and payoff games to linear programming. In *Proceedings of the 34th International Symposium on Mathematical Foundations of Computer Science (MFCS 2009), 24–28 August, Novy Smokovec, Slovakia*, volume 5734 of *Lecture Notes in Computer Science*, pages 675–686. Springer-Verlag, 2009.

[Sen94] Linn I. Sennott. Zero-sum stochastic games with unbounded costs: Discounted and average cost cases. *Math. Meth. of OR*, 39(2):209–225, 1994.

[Sha53] L. S. Shapley. Stochastic Games. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 39, pages 1095–1100, Washington, DC, October 1953. National Academy of Sciences.

[Sta85] Oded Stark. On private charity and altruism. *Public Choice*, 46(3):325–332, 1985.

[Sta89] Oded Stark. Altruism and the quality of life. *The American Economic Review*, pages 86–90, 1989.

[TR97] Frank Thuijsman and Thirukkannamangai E.S. Raghavan. Perfect information stochastic games and related classes. *International Journal of Game Theory*, 26(3):403–408, 1997.

[Tuc50] A.W. Tucker. *A two-person dilemma*. Stanford University, 1950.

[Umm05] Michael Ummels. Rational Behaviour and Strategy Construction in Infinite Multiplayer Games. Diploma thesis, RWTH Aachen, 2005.

[Umm06] Michael Ummels. Rational behaviour and strategy construction in infinite multiplayer games. In *FSTTCS*, pages 212–223, 2006.

[Umm08] Michael Ummels. The complexity of Nash equilibria in infinite multiplayer games. In *Proceedings of the 11th International Conference on Foundations of Software Science and Computational Structures (FoSSaCS 2008), March 29 - April 6, Budapest, Hungary*, volume 4962 of *Lecture Notes in Computer Science*, pages 20–34. Springer, 2008.

[UW11] Michael Ummels and Dominik Wojtczak. The complexity of Nash equilibria in limit-average games. In *Proceedings of the 22Nd International Conference on Concurrency Theory*, CONCUR'11, pages 482–496, Berlin, Heidelberg, 2011. Springer-Verlag.

[vS34] H. von Stackelberg. *Marktform und Gleichgewicht*. J. Springer, 1934.

[vSZ04] Bernhard von Stengel and Shmuel Zamir. Leadership with commitment to mixed strategies. Technical report, 2004.

[vSZ10] Bernhard von Stengel and Shmuel Zamir. Leadership games with convex strategy sets. *Games and Economic Behavior*, 69(2):446–457, 2010.

[yic] Yices website: `http://yices.csl.sri.com/`.

[ZP96] Uri Zwick and Mike S. Paterson. The complexity of mean payoff games on graphs. *Theoretical Computer Science*, 158(1–2):343–359, 1996.