Topics in Repeated Discounted Games
With Imperfect Monitoring

Thesis submitted for the degree of "Doctor of Philosophy"
by
Galit Ashkenazi-Golam

Submitted to the senate of Tel-Aviv University
December 2007
This work was carried out under the supervision of

Prof. Ehud Lehrer
Acknowledgements

It is a pleasure to thank all those who supported me along the way.

First, Professor Ehud Lehrer, a teacher and a guide, with vast knowledge and true passion for science, I cherish the hours we spent together, and I am honored to be your student.
I thank Professor Silvain Sorin, for the warm hospitality of the Laboratoire d'Economètre de l'École Polytechnique. The atmosphere in the Laboratoire helped a lot of the ideas expressed in this work to form. I wish to thank Professor Olivier Gossner for long interesting conversations, Professor Nicholas Vielle, Professor Jérom Renault, Professor Tristan Tomala and Dr. Ridah Laraki for their friendly help whenever it was needed. I thank Professor Dinah Rosenberg for becoming a true loved friend.
I also wish to thank my fellow students in the of the Laboratoire d'Economètre de l'École Polytechnique, and especially Dr. Yannick Viossat for reading and commenting parts of this work, and Dr. Pierre Fleckinger and Dr. Ignace Adant for cooperating in research projects. You all made our stay in Paris unforgettable and fruitful.
Dr. Dipjyoti Majumdar allowed me to include our work in this dissertation, and I thank him for his cooperation and friendship.
I wish to thank INSEAD Business School and Professor Timothy Van-Zandt for generously welcoming me for a visit, and for giving me the opportunity to be a part of the INSEAD experience.
I thank Professor Eilon Solan, Professor David Schmeidler, Professor Dov Samet and Dr. Ido Eliazar for their help along the way.
I thank my fellow students in Tel-Aviv University – Dr. Eran Shmaia, Yaron Azriely and Roe Tepper. It's been a pleasure working together.
A special thank goes to Nurit Liberman and the secretaries of our school, for making bureaucrat barriers disappear with a smile.

Last, but most – I thank my family. I thank my parents, brother and sisters for their support, encouragement, for believing in me and loving me.
I thank Joy for long walks that helped me clear my thoughts; I thank Omer for being the light of my life, and I am still grateful for being late for a class, about eleven years ago, and meeting Ronny. Through Tel-Aviv, Paris, Fontainebleau, back to Tel-Aviv, and wherever life will take us next – I love you.
Abstract

This work revolves around discounted repeated games with imperfect monitoring. It involves both private monitoring and public monitoring with private strategies.

This thesis is composed of three main parts.

In the first part, we explore a repeated game with imperfect private monitoring and communication. We obtain a Nash-threat folk theorem under assumptions regarding the observation of the players. In our construction, the deviating player should convey a special signal “confessing” a deviation, and thus obtaining a smaller sanction, a “pardon”.

In the second part, we consider two-player games where the payoffs of the players are deterministic given the action-profile played. Own-payoffs are observed and communication is not allowed. We obtain a folk-theorem under the assumption that a payoff that is internal to the set of payoffs can be obtained via sequential equilibrium, and we show how this can be done in all games but a very degenerate set of games.

In the third part, cooperation with Dr. Dipjyoti Majumdar, we consider a repeated auction with two bidders. The values of the sold goods are private and independent, the bidders are not allowed to communicate, and at the end of each period only the identity of the winner is announced. We show how when the bidders use private strategies they can approach efficient collusion via epsilon-equilibrium if the bidders are sufficiently patient, while if the players are restricted to using only public strategies, their payoffs are bounded away from the efficient collusion, even when using epsilon-equilibrium.

The organization of the thesis is as follows:

- In chapter 1 we provide an overview and introduce the main issues dealt with in this work.
- In chapter 2 we deal with repeated discounted games with communication and introduce the “confession and pardon” result.
- Chapter 3 discusses two-player discounted game with observable deterministic own-payoffs and without communication. We obtain the folk-theorem under some assumptions.
- Chapter 4 deals with a repeated private-value auction of two bidders, where only the identity of the winner is announced. We demonstrate how efficient collusion can be supported using private strategies via epsilon-equilibrium, and how efficient collusion cannot be obtained using public strategies alone.
## Contents

1 Introduction .................................................. 5

1.1 Imperfect Monitoring ........................................ 6

1.1.1 First Part - Private Monitoring with Communication . 8

1.1.2 Second Part - Deterministic Observable Payoffs ....... 8

1.1.3 Third Part - a Repeated Auction ....................... 9

2 A Model of Private Monitoring and Communication .... 11

2.1 Introduction .................................................. 11

2.2 Detecting Deviations with Private Monitoring and Communication - Examples 15

2.2.1 Example 1: a partnership game .......................... 15

2.2.2 Example 2: A Principal-Agent Game ..................... 16

2.3 Preliminaries ................................................ 17

2.3.1 The Stage Game ........................................... 17

2.3.2 The Repeated Game ....................................... 18

2.3.3 Sequential Equilibria .................................... 18

2.3.4 Observation Assumption ................................... 19

2.3.5 The Equilibrium Path of Fudenberg and Maskin ....... 20

2.4 Communication, Deviations and Confessions - Nash Threat Folk Theorem . 21

2.5 Discussion ................................................... 25

2.5.1 Variety of Sanctions as a Tool to Enable Efficiency, Transparency and Trust ......................................................... 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5 From Terminal Payoffs to Continuation Payoffs</td>
<td>53</td>
</tr>
<tr>
<td>3.5.1 First Step: A Two-Dimensional Set of Payoffs with Observable Deviations, ( W )</td>
<td>53</td>
</tr>
<tr>
<td>3.5.2 Second Step: Using the set ( W )</td>
<td>57</td>
</tr>
<tr>
<td>3.5.3 Step 3: Communicating the Continuation Payoff</td>
<td>61</td>
</tr>
<tr>
<td>3.5.4 Off Equilibrium Path Beliefs</td>
<td>67</td>
</tr>
<tr>
<td>3.6 Obtaining an Internal Point in Other Matrices</td>
<td>68</td>
</tr>
<tr>
<td>3.6.1 A Simple Randomization with Continuation Payoffs</td>
<td>68</td>
</tr>
<tr>
<td>3.6.2 A Complicated Randomization with Continuation Payoffs</td>
<td>69</td>
</tr>
<tr>
<td>4 Repeated Two-Bidders Auction</td>
<td>73</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>73</td>
</tr>
<tr>
<td>4.2 Preliminaries</td>
<td>77</td>
</tr>
<tr>
<td>4.2.1 The Stage Game</td>
<td>77</td>
</tr>
<tr>
<td>4.2.2 The Repeated Game</td>
<td>78</td>
</tr>
<tr>
<td>4.2.3 The Equilibrium Scheme</td>
<td>79</td>
</tr>
<tr>
<td>4.3 ( \epsilon )-Consistent Equilibrium</td>
<td>82</td>
</tr>
<tr>
<td>4.4 Main Result</td>
<td>84</td>
</tr>
<tr>
<td>4.4.1 Proof of Lemma 4.1</td>
<td>84</td>
</tr>
<tr>
<td>4.4.2 The Punishment Phase</td>
<td>86</td>
</tr>
<tr>
<td>4.4.3 Proof of Lemma 4.2</td>
<td>87</td>
</tr>
<tr>
<td>4.4.4 Proof of Lemma 4.3</td>
<td>88</td>
</tr>
<tr>
<td>4.4.5 The Communication Phase</td>
<td>89</td>
</tr>
<tr>
<td>4.4.6 The Cooperation Phase</td>
<td>90</td>
</tr>
<tr>
<td>4.5 Public Strategies</td>
<td>97</td>
</tr>
<tr>
<td>4.6 Discussion: Correlated Valuations</td>
<td>101</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Game-theory aims to model strategic interactions between agents (people, firms, countries etc.). The theory of repeated games attempts to model such cases where the same interaction repeats itself. Two firms (a duopoly) that operate in the same market, an employer and an employee, and long-term partnerships are all examples of repeated games.

The “solution concept” of game-theory is the equilibrium - a situation where each agent plays a strategy that is his best-response against the strategies of his opponents. This is also what we call a “self-enforcing agreement”.

The main result of the theory of repeated games is the “folk theorem”. The folk theorem states that there is a large variety of equilibria payoffs when the game is infinitely repeated, and the players are patient. In technical language, any payoff that is feasible and individually rational can prevail as an equilibrium payoff if the players are patient enough. This result is due to Aumann and Shapley (1994) and Rubinstein (1994) in the non-discounted model, and Fudenberg and Maskin (1986, 1991) for the discounted case.

However, three strong assumptions are needed for the folk theorem to hold. The first is that the game is perfectly known to all players. The players should know who are their opponents, be aware of all their possible actions, all the opponents’ possible actions, and the resulting payoffs for each combination of actions (action profile) for all the players. This assumption is called “complete information”.

5
The second assumption is that the players have “perfect recall”, that is, they remember (and can condition their actions on) a history that is as long as they wish.

The third assumption is that the actions taken by each player are perfectly observed (in a costless manner) and by all the players. This assumption is called “perfect monitoring”. ¹

Attempts to relax those three assumptions and analyze the resulting games were the center of many recent researches.

My dissertation belongs to the branch of literature that tries to relax the perfect monitoring assumption, and figure out what can be said about games where the players do not observe each others’ actions, but rather some imperfect signals that result from those actions.


Some other results concern reputation effects (see Evans and Thomas (1997), for example).

Repeated games where the players have bounded recall were analyzed, among others, by Lehrer (1994, 1988), Barlo and Carmona (2003), and Sabourian, H. (1989).

### 1.1 Imperfect Monitoring

Repeated games with imperfect monitoring can be further divided to games with public monitoring and games with private monitoring.

Games with public monitoring are games where during each period, the signal that the players observe (after their actions are taken) is public. That is, all the players observe the same signal.

Games with private monitoring are games where each player observes his signal privately.

¹Another assumption is “perfect rationality”, but this the discussion of rationality is a bit far from the scope of this introduction.
Given the action-profile played, the signals can be independent or correlated.

When analyzing games with public monitoring, the solution concept of perfect public equilibrium is often used (see the introduction of the third chapter for further details and references). Such an equilibrium limits the players to using strategies that are conditioned only on the public history (that is, the sequence of public signals observed). This limitation simplifies the analysis since each player knows exactly all the parts of the history that his opponents are conditioning their actions on.

However, when one allows the players to condition their actions on the public history and on their private history (i.e. the actions that they took as the game unfolded, actions that are not perfectly observed by their opponents), then the analysis becomes more involved. In this case, a player does not know the entire histories of the opponents that are relevant for determining future actions, and therefore, the players will need to have beliefs about each other’s history, and beliefs regarding the opponents’ beliefs etc. In equilibrium, each player will need to best-respond based on his beliefs.

A similar problem occurs when we consider models of private monitoring: the players don’t share the same knowledge about the relevant history and hence need to form beliefs about the private histories of their opponents etc.

One way to try and overcome such a difficulty is to introduce communication into the game. When we assume that players can communicate, we allow the players to convey public messages after they observe their private signal. This will create a public history of messages that the players can condition their strategies on. However, the design of the equilibrium will need to provide the players with incentives to convey the proper messages.

This dissertation has three parts. The first part, chapter 2, considers a model of repeated games with private monitoring and communication; the second part, chapter 3, analyzes a game with deterministic observable own-payoffs without communication; and the third part, (chapter 4) concerns a repeated auction of two bidders without communication - a model of public monitoring when we allow the players to use private strategies.

\footnote{There are also models where the players communicate through a mediator as in Forges (1986)}
1.1.1 First Part - Private Monitoring with Communication

In the following chapter, I describe in details the result regarding games with private monitoring and communication. I obtain a Nash-Threat folk theorem, under an observation assumption.

The observation assumption is that for every payoff that is an extreme point of the set of payoffs there exists an action-profile such that any profitable deviation from that action-profile is detectable in at least one of the two following ways:

The first way of detecting a deviation is by a signal-profile for the opponents of the deviator that indicates a deviation. The second is through loss of information for the deviator because of the deviation. This loss of information will not enable the deviator to report observing a signal that completes the signals of his opponent to a signal-profile that is consistent with the equilibrium path.

Enforcing truthful revelation of the signals is through a variety of sanctions, and a “confession”: a deviation that is followed by a special message of the deviator that admits having deviated is lighter than a deviation that is not followed by such a message, thus, a confession induces “pardon”.

1.1.2 Second Part - Deterministic Observable Payoffs

In the third chapter I analyze two-players discounted repeated games with private monitoring and without communication. During each period, after the actions are taken, each player observes his own payoffs. The payoffs for each pair of actions are deterministic and communication is not allowed. I prove that the efficient frontier can be obtained via sequential equilibrium in games where there exists an internal payoff that can be used as a continuation payoff. In addition, I show how to obtain such an internal payoff in all but a very degenerate set of games.
1.1.3 Third Part - a Repeated Auction

Chapter four was written is cooperation with Dr. Dipjyoti Majumdar. We analyze two-players private-value repeated auction, where only the winner’s identity is announced. At the beginning of each period, each bidder learns his own value for the good that is to be sold. After the auction takes place, no other information is available to the players but the identity of the winner. We show that there exist $\epsilon$-equilibria approximating first-best collusive outcomes when the bidders are sufficiently patient and are allowed to condition their strategies on their private information. In addition, we show that when the players are restricted to using only public strategies, then their equilibrium payoff is bounded away from efficiency in any $\epsilon$-equilibria.
Chapter 2

A Model of Private Monitoring and Communication

2.1 Introduction

Long-lasting economical relationships between a small number of agents are often characterized by frequent communication and private monitoring. Such interaction enables flexibility to choose from a variety of continuations in case problems occur. One example of such relationships are principle-agents games. Another example are games of partnership. The theory of infinitely-repeated private-monitoring games with communication attempts to model such interactions.

Consider, as a general example, principal-agent relationships under subjective evaluation (which is equivalent to private monitoring). In case where a deviation is not necessarily detected, there could be room for shirking. To discourage shirking, the principal might resort to a harsh punishment (such as firing) upon detection. Firing, however, would eliminate the possibility of future gains from the continuation of the relationships. Thus, harsh punishments are a non-credible threat in a situation where there is a positive continuation value to the relationship for the principal, and, in addition, the agent cannot know whether the shirking was detected or not. We offer a construction that restores contract-enforcing and,
in turn, efficiency. This can be facilitated by the incorporation of a range of sanctions to be applied once a deviation is detected. With regard to the literature of optimal contracting, this suggests that termination of the contract is not sufficient for efficiency, a result in the spirit of McLeod (2003). From within the range of sanctions, lighter punishments will be employed if shirking was followed by a “confession”. Hence, a post-deviation discussion between the employer and the employee, followed by a potential amicable arrangement is a necessary part of our construction.

A second example one could imagine is of two partners who are working together on a project from a distance (programming a software, designing a campaign, writing a paper etc.). Partner A decides to take a few days off. This action is un-observable by Partner B. Partner B calls and demands to have answers regarding e-mails he sent. Partner A can either “confess” not reading them or he can try to guess their content and try and relate to them. If Partner A attempts to “fake” reading the e-mails he risks exposing his ignorance which in turn reveals his shirking.

Yet a third example is that of a diner in a restaurant. Occasionally, the restaurant manager figures out that the diner’s favorite dish falls short of standards. On such occasions, the manager faces a dilemma. On the one hand he could keep silent, hoping that the diner would not notice the difference, thus risking disappointment and dissatisfaction, which may lead the diner to refrain from returning to the restaurant for a long period of time (a “harsh” punishment). On the other hand, he could “confess”, maybe loose the current meal of the diner (a “soft” punishment), and retain the customer’s trust.

In the principal-agent example, a deviation from an agreement results in a positive probability for the deviation being detected. In the partnership example, the deviation results in a loss of information, the content of the e-mails, and when communication between the agents takes place, this loss, which indicates a deviation, may be exposed.

We are interested in long-term interactions that involve communication, where the agents try to establish cooperation and trust under possibly profitable deviations. If both players are interested in preserving cooperation, and if there is only one sanction available, then the
private monitoring could make the implementation of such a sanction un-credible. If the deviator cannot know whether his deviation is observed, then the opponent would choose to overlook the deviation in order to sustain the cooperation. However, if there are a few levels of sanctions, then a more truthful communication can be established. Assume that if a deviator “confesses” a deviation, a lighter sanction is employed, and in case a deviation is detected and no confession is made, a heavier sanction is executed. In such a case both players are motivated to communicate truthfully about their observations and past actions. This is in line with the notion that confessing a “mistake” might reduce the long-lasting damage to the trust in a relationship.

Trust is essential when players cannot fully monitor each other, and it is easier to establish when the number of players is relatively small and they communicate with each other frequently.

Theoretical results that apply to the model of two-player repeated discounted games with private monitoring are relatively scarce. There are some results concerning prisoner’s dilemma and a recent working paper by Obara (2006) which assumes an observation mechanism different from ours (Obara requires full-support of the signals of each player, i.e., every signal is observed by each player with a positive probability after any common action played, so his informational assumptions are different, maybe complementary, to ours).

The main result of this chapter is based on an assumption that refers only to payoffs, called extreme payoffs, which are extreme points of the set of all possible payoffs (the feasible set). We assume that every deviation from a common action, whose payoff is extreme, is detectable in one of two ways. The first is by observing the deviation. This might occur when a deviation induces a positive probability for at least one profile of private signals (of the opponents) assigned a zero probability under the distribution corresponding to the agreed upon joint action.

The second way to detect a deviation is the indirect one: the deviation may result in a loss of information that the deviator would otherwise receive. On equilibrium paths, the players are supposed to publicly report their private signal after each period. A report that
 CHAPTER 2. A MODEL OF PRIVATE MONITORING AND COMMUNICATION  14

is not consistent with the equilibrium path, signifies that at least one player has deviated. While deviating a player might lose information necessary to report in a manner that is consistent with others’ report. Thus, a possible loss of information upon deviation makes the deviation detectable.

A signal of a deviator will be called sufficiently informative if it has two properties. First, it lets her know that her opponents’ signals are consistent with the equilibrium path; and second, it allows her to complete these signals with a signal of her own, so that all together they are consistent with the equilibrium path. A sufficiently informative signal enables a player to get away with a deviation. We assume that, having deviated from an action profile whose payoff is extreme, a deviator observes a sufficiently informative signal with a probability strictly less than 1.

A detection of a deviation is followed by a punishment. The following punishment mechanism enforces truth-telling when the players are communicating. There are three types of punishment phases, differing according to their durations: a short-term, a medium-term, and a long-term. During any punishment phase, a one-period Nash-Equilibrium is played. A short-term punishment phase will take place when first, all the players, but the deviator, announce a combination of signals which means that a deviation took place and, second, the deviator confesses a deviation. A medium-term punishment phase takes place when only the deviator confesses, and the combination of signals of the other players is consistent with the agreed upon action. And a long-term punishment phase takes place when the deviator does not confess, while the reports are inconsistent with the equilibrium path.

According to the construction, when a player “confesses” a deviation, the harshest punishment she might get is the medium-term punishment. When a player does not confess, she risks the long-term punishment. The duration of the various punishment phases are designed to provide a deviator a motivation to “confess” her deviation even if the deviation has a small probability of being detected. The opponents, in turn, know that whenever they observe a signal-profile indicating a deviation, it will be followed by a confession. Such a confession results in a short-term punishment if the opponents’ reports will indicate a deviation and
in a medium-term punishment otherwise. Since, a short term punishment is preferred to
a medium-term punishment, the players have incentives to truthfully report their signals.
After the punishment, the players restart the equilibrium path.

A punishment phase is triggered by both the deviator’s confession and her opponents’
reports. Thus, for a player observing the signal which indicates a deviation, overlooking
the deviation is unprofitable since it is accompanied by a confession of the deviator. For a
deviator who receives a signal that is not sufficiently informative, trying to avoid punishment
by not confessing is un-profitable due to the positive probability of the deviation being
detected.

The sequel of the paper chapter is organized as follows: Section 2 presents examples of
games for which our model and assumptions apply naturally. In Section 3 introduces the
formal model. Section 4 contains the main result - a Nash-threat folk theorem. This result
is discussed in Section 5. A literature survey is conducted in Section 6. Section 7 integrates
our results with those of Kandori and Matsushima (1998) and Compte (1998), in order to
obtain a richer set of sequential equilibria payoffs.

2.2 Detecting Deviations with Private Monitoring and
Communication - Examples

2.2.1 Example 1: a partnership game

Consider the following partnership game, which is a version of the Prisoner’s dilemma. In
this game there are two partners, each one of them can “work” (w) or “shirk” (s)\(^1\). The
expected payoffs from the actions are given in table 1:

<table>
<thead>
<tr>
<th></th>
<th>work</th>
<th>shirk</th>
</tr>
</thead>
<tbody>
<tr>
<td>work</td>
<td>1,1</td>
<td>-1,2</td>
</tr>
<tr>
<td>shirk</td>
<td>2,-1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

\(^1\)This example appears in Fudenberg et al(1994)
Since the profitable deviations are only from “work” to “shirk”, we shall concentrate on deterring only such deviations.

One example of an observation mechanism that will enable the player to detect deviations is a variation of the standard-trivial model, introduced by Lehrer (1990). In the standard-trivial model the signals of the players are either the action taken by the opponent or a null signal. If there is always a positive probability that the opponent observes a signal indicating the action, then deviations can be detected. We don’t need to make any assumption about the correlation of the signals, so they can be perfectly correlated (which gives a public-monitoring game), independent, or any other form.

Note that in the independent signals case (as well as in many other cases), when a player observes a signal which indicates a deviation, he knows that the deviator does not know whether the deviation has been detected. We shall use the one-period equilibrium payoff \((0, 0)\) as a punishment, so the punishment will be costly for the punisher as well. In this case, we will need to motivate a player observing a deviation to report it, even though the punishment is costly, and even though he knows that the deviator does not know that the deviation has been detected. Motivating such a truthful communication will be done using the pardon and the variety of sanctions available in this game.

### 2.2.2 Example 2: A Principal-Agent Game

A second example of an observation mechanism that satisfies our assumption can be found in principal-agent games.

The actions of the principal are perfectly observed; the agent exerts effort, and (only) the principal observes a stochastic result. The imperfectness of the monitoring gives rise to moral-hazard situations, and the one-period equilibrium is often strictly Pareto-dominated by the efficient outcome.

In some games, when the agent chooses not to exert enough effort (not to study a subject thoroughly enough, not to be present in certain situations or places) the agent risks losing information. In the following communication between the principal and the agent, if there
is a chance that this loss of information will be exposed, then we will say that the deviation (of not exerting enough effort) is detectable.

2.3 Preliminaries

The model outlined below features $n$-player repeated games with stochastic private signals. After each period, the players observe private signals, and after observing their private signals, the players simultaneously send public messages.

2.3.1 The Stage Game

In the stage game, players move simultaneously and each player $i \in N$ chooses an action $a_i$ from a finite set of actions $A_i$. After actions are played, each player $i$ observes a signal, $y_i$ which is not observed by the opponents. Let $n$ be the number of players, $|N|$. Let $Y_i$ be the finite set of possible private signals for player $i$. A signal profile is an $n$-tuple $y = (y^1, ..., y^n) \in Y = \Pi_{i \in N}Y_i$. Each action profile $a = (a_1, ..., a_n) \in A \equiv \times_{i \in N}A_i$ induces a probability distribution over signal profiles. Let $p(\cdot|a)$ be the common distribution of the private signals, conditioned on the common action $a$. Let $(a_{-i}, a'_i) \in \times_{j \neq i}A_j \times A_i$ be the action profile where all the players but $i$ follow the action profile $a$, and player $i$ is playing $a'_i$. Let $y_{-i} \in Y_{-i} = \times_{j \neq i}Y_j$ be the signal profile of the opponents of player $i$. Let $p_{-i}(\cdot|a)$ be the common distribution of the signals of the opponents of player $i$ when the common action taken is $a$. Let $q_i(y_{-i}|a, y_i)$ be the probability that the opponents of player $i$ received the signal profile $y_{-i}$ when the action profile was $a$ and the signal player $i$ received was $y_i$.

Each player $i$’s mean payoff $g_i(a)$ depends on the action profile played. The realized payoff can be dependent on the signals, and in general is not known to the player.

We allow players to *communicate* with each other. After choosing actions and observing their private signals, the players simultaneously and publicly announce messages. Player $i$ announces a message taken from the finite set $W_i$. Thus, a profile of messages is $(w_1, ..., w_n)$, where $w_i \in w_i$ for $i = 1, ..., n$. 
2.3.2 The Repeated Game

At each date $t = 1, 2, \ldots$ the stage game is played and private signals are observed. At the end of period $t$, the private history of player $i$ consists of player $i$’s past actions, past private signals, and the public messages: $h_i^t = (a_i(1), y_i(1), w(1), \ldots, a_i(t), y_i(t), w(t))$. We denote by $h_i(0)$ the null private history of player $i$. We denote by $h$ the sequence of actions, signals and messages taken so far (the history of the game), and by $\mathcal{H}$ the set of all possible histories. A pure strategy $(\sigma_i, \tau_i)$ for player $i$ is a pair of sequences of maps $\{\sigma_i^t\}_{t=1}^\infty$, $\{\tau_i^t\}_{t=1}^\infty$, where $\sigma_i^t$ maps each history that ends with the public messages, to an action in $A_i$ to be taken in the next period, and $\tau_i^t$ maps each history that ends with a private signal, to the public message the player should announce. The payoffs for the players are a result of a randomization that depends on the actions taken and the signals observed. This model encompasses both cases where the payoffs are directly observed (that is, they are a part of the private signal) and cases where the payoffs are not observed.

Without loss of generality, we restrict our discussion to behavior strategies\textsuperscript{2}. Formally, a behavior strategy is $(\sigma_i^t, \tau_i^t)$,

\[
\begin{align*}
\sigma_i^t &: \times_{t' = 1, \ldots, t} A_i \times_{t' = 1, \ldots, t} Y_i \times_{t' = 1, \ldots, t} (W(1), \ldots, W(n)) \rightarrow \Delta A_i \\
\tau_i^t &: \times_{t' = 1, \ldots, t} A_i \times_{t' = 1, \ldots, t} Y_i \times_{t' = 1, \ldots, t} (W(1), \ldots, W(n)) \rightarrow W_i
\end{align*}
\]

Each strategy profile $(\sigma, \tau) = \times_{i \in N}(\sigma_i, \tau_i)$ generates a probability distribution over future streams of actions, payoffs and messages, which in turn induces a distribution over future payoffs. Players are assumed to discount future payoffs with a common discount factor $\delta$.

Player $i$’s average discounted expected payoff from $\sigma$ is

\[
v_i(\sigma, \tau) = (1 - \delta)E[\sum_{t \geq 1} \delta^{t-1} g_i(a(t))].
\]

2.3.3 Sequential Equilibria

A strategy profile $(\sigma, \tau)$ is a Nash equilibrium if and only if for any player $i$ and for any strategy $(\sigma_i', \tau_i')$, $v_i(\sigma, \tau) \geq v_i((\sigma_i', \tau_i')(\sigma_{-i}, \tau_{-i}))$.

\textsuperscript{2}Kuhn (1953) showed that for games with perfect recall considering behavior strategies is sufficient.
An assessment is a trio \((\sigma, \tau, \mu)\), where \(\sigma\) is a profile of behavioral strategies, \(\tau\) is the decision rule regarding the messages to convey and \(\mu\) is a function that assigns to every information set a probability measure on the set of histories in the information set. We shall refer to \(\mu(h_i, h)\) as the beliefs of the players, the probabilities a player assigns to the history \(h \in \mathcal{H}\) conditional on the private history \(h_i\) being observed.

The assessment \((\sigma, \tau, \mu)\) is sequentially rational if for every player \(i \in N\) and for every information set of player \(i\), the strategy of player \(i\) is a best response to the strategies of the other players, given the information set and the belief at this information set.

An assessment is consistent if there is a sequence \(((\sigma^n, \tau^n, \mu^n))_{n=1}^{\infty}\) of assessments that converges to \((\sigma, \tau, \mu)\) in Euclidean space and has the properties that each strategy profile \((\sigma^n, \tau^n)\) is completely mixed (meaning that it assigns positive probability to every action at every information set) and that each belief system \(\mu^n\) is derived from \((\sigma^n, \tau^n)\) using Bayes’ rule.

An assessment is a sequential equilibrium if it is sequentially rational and consistent.

### 2.3.4 Observation Assumption

We shall call a private signal observed by a deviating player sufficiently informative if it enables the deviator to choose a private signal that will complete the signals of her opponents to a signal profile that is consistent with the agreed upon action. The signal will need to indicate that the opponents’ common signal does not indicate a deviation and to allow choosing a private signal that will complete any signal-profiles that the opponents might have (assuming the opponents did not deviate and given the private signal) to a common signal that is consistent with the equilibrium path.

We assume that every profitable deviation from a pure action profile whose payoff is an extreme point of the set of feasible payoffs induces a probability strictly less than one for the deviator to observe a sufficiently informative signal.

Formally, let \(EX \subset A\) be the set of action profiles whose payoff is an extreme point of the set of feasible payoffs.
Definition 2.1: A message of player $i$, announcing that he observed a signal $y_i$ following a period when player $i$ played $a_i'$ and observed $y_i'$ is consistent with action-profile $a$ if for every signal profile of player $i$'s opponents, $y_{-i} \in Y_{-i}$, such that $p(y_{-i}, y_i'|a_{-i}, a_i') > 0$ it holds that $p(y_{-i}, y_i|a) > 0$.

A message is consistent with an action profile, if given the information available to player $i$ when he needs to convey that message (an information that includes the action player $i$ played, the action profile the players were supposed to play, and the signal player $i$ observed) if player $i$ chooses to convey that message, then this message will complete any report of a message that the opponent might observe to a signal profile that is consistent with the action profile.

Definition 2.2: A signal $y_i$ of player $i$, following a deviation $a_i'$ from the common action profile $a$ is sufficiently informative if there exists a signal $y_i'$ that is consistent with action-profile $a$.

We make the following assumption:

Assumption 2.1: Every profitable deviation from a pure action profile $a \in EX$ induces a probability strictly less than one for the deviator to observe a sufficiently informative signal.

Assumption 2.1 is equivalent to assuming that every profitable deviation from a pure action profile $a \in EX$ is detectable with positive probability.

2.3.5 The Equilibrium Path of Fudenberg and Maskin

As a part of the equilibrium construction detailed below, we use a part the equilibrium structure of Fudenberg and Maskin. Fudenberg and Maskin show (lemma 2) that for any $\epsilon > 0$ there exists $\delta' < 1$ such that for all $\delta \geq \delta'$ and every feasible payoff $v$ with $v_i \geq \epsilon$ for all $i$ there is a deterministic sequence of pure strategies whose discounted average payoffs are $v$, and whose continuation payoffs at each time $t$ are within $\epsilon$ from $v$.

Roughly, the idea is to divide the game into blocks, such that the discounted payoff of each block is within an $\epsilon$ from $v_i$ and such that their there exists a sequence if those blocks
whose overall discounted payoff is exactly \( v \).

### 2.4 Communication, Deviations and Confessions - Nash Threat Folk Theorem

In this section we shall prove the Nash-threat folk theorem.

The main idea of our equilibrium construction is to provide a player with an incentive to “confess” a deviation if she indeed deviated and to “report” a deviation if she observed it.

Denote by \( V \) the set of feasible payoffs, by \( V^* \) the set sequential equilibria payoffs and by \( V^{**} \) the set of feasible payoffs strictly Pareto dominating one-period equilibrium of the repeated game.

**Theorem 2.1:** When the players are allowed to communicate, and for a discount factor close enough to 1 \( V^{**} \subseteq V^* \).

**proof:**

Let \( v = (v_1, v_2, ..., v_n) \in V^{**} \). We shall follow the pure-action equilibrium path, as constructed by Fudenberg and Maskin. In their equilibrium path the continuation payoff is always within an \( \varepsilon \) distance of \( v \). The exact \( \varepsilon \) that we shall use is dependent on the payoffs and the information structure, as will be shown in the following.

The equilibria strategy we construct is to follow the equilibrium path as in Fudenberg and Maskin, and after each period convey a public message that informs the opponents of the private signal that was observed, until a deviation from this path has occurred. If a player deviated and observes a message that is not sufficiently informative, he conveys a special message, “confessing” his deviation, a message that informs the opponents that he

\[3\text{In fact, Let } W \text{ be the convex-hull of all the payoffs of pure-actions such that every profitable deviation induces a probability strictly less than one for the deviator to observe a sufficiently informative signal. Then the following proof can be used to show that any payoff inside } W \text{ that is strictly Pareto dominating a one-period Nash equilibrium payoff can be supported as sequential equilibrium payoff for patient enough players.}\]
deviated in the last period.

To create the proper incentives to convey these messages (that could trigger a punishment phase), three different punishments are constructed: a short-term punishment in case both the deviator confessed his deviation and his opponents conveyed a private messages profile that indicates that a deviation took place; a medium-term punishment in case only the deviator conveyed his confessing message; and an “eternal” punishment in case the signal profile reported is inconsistent with the equilibrium path, but no player confessed a deviation.⁴

The three possible punishments will be three durations of punishment phases when all players play the one-period equilibrium, followed by re-starting the equilibrium path. The length (number of periods) of the short-term punishment will be \( L_1 \), of the medium-term \( L_2 \) and the long-term punishment will last forever. Without loss of generality, we shall assume that the dominated one-period equilibrium payoffs are 0 for all players. The lengths of punishments will be the same for the different deviations of the different players, and therefore there is no need to specify who was the deviator in case that there was no confession. The punishments are the same.

In this construction, the (pure) equilibrium strategy is strictly more profitable than any deviation. Therefore, as long as a player observes signals that are consistent with the equilibrium path, his beliefs (the belief part of the assessment) will be that the opponent is indeed conforming to this path. Following a deviation, again, it is strictly more profitable to follow the confession-and-pardon instructions than to convey any other messages, and once the confession-and-pardon takes place, it is common knowledge that the players should play the one-period Nash equilibrium, which satisfies the conditions of sequential equilibrium.

Let \( \overline{G} \) be the maximal one-period payoff over all players. Let \( 1 - p \) be the maximal probability, over all players and all profitable deviations from all common actions whose

---

⁴The definition of sequential equilibrium demands that we shall also deal with cases of simultaneous deviations of two players or more. It is easy to see that it can be solved in the same fashion with the assumption of more messages, for example, a message which means:”I deviated and I observe the signal \( x \)”. The length of the punishment will be determined, again, by how much all messages are consistent.
payoffs are in $EX$, that the deviator will observe a signal that is sufficiently informative (that he will “get away” with the deviation). We shall induce punishment whenever the signal observed by the deviator is not sufficiently informative.

Now, with probability at least $p$ the deviator will have a signal that is not sufficiently informative. When the signal is not sufficiently informative, then every choice of the deviator of a message to convey leads with a positive probability to detecting the deviation (because the signal profile reported will be inconsistent with the equilibrium path). Let that (positive) probability be $r$. Let $r'$ be the minimum of the $r$’s, over all players, and all their combinations of deviation and not sufficiently informative signals.

We shall describe the three punishment phases - short-term in case both the deviator’s opponents report a signal combination that is inconsistent with the agreed upon common action and a signal of confession is observed, medium-term in case there was only a confession and long-term, in case the signal profile announced is inconsistent with the equilibrium path instructions, but no confession was announced.

From the deviator’s point of view, when his signal is not sufficiently informative, confessing will be followed, at worst, with a medium-term punishment, (a “pardon”). Sufficiently large difference between the long-term and the medium-term punishments will induce the deviator to confess. From the deviator’s opponents point of view, if before sharing the private signal, the opponent does not know that his private signal will help detecting a deviation - then there is no harm in announcing it. If he does know, then the following argument holds: if his signal is a part of a signal profile which indicates a deviation, then the deviator cannot have a signal that is sufficiently informative, hence he will confess. So the choice is between reporting and continuing to the short-term punishment and not reporting, which will result in the medium-term punishment. Any difference between the short-term and the medium-term punishment will suffice to induce reporting a deviation.

For the description above to be an equilibrium, it is sufficient that for all the players:

When the deviator observes a signal that is not sufficiently informative, confessing is more profitable than not confessing:
(1) \( v_i^M > (1 - r')(v_i + \varepsilon) \)

Staying in the equilibrium path is more profitable than deviating (when deviating, with probability at most \(1 - p\) there is no punishment, and with probability at least \(p\) there is at least the small punishment):

(2) \( v_i - \varepsilon > G(1 - \delta) + \delta(1 - p)(v_i + \varepsilon) + \beta p v^S_i \)

Reporting is more profitable than not reporting:

(3) \( v^S_i > v_i^M \)

where,

(4) \( v^S_i = \frac{\delta}{2} L_1 v_i \)

(5) \( v_i^M = \frac{\delta}{2} L_2 v_i \)

We need to show that for \(\delta\) close enough to 1, there are values for \(v^S_i\) and \(v_i^M\) which solve (1), (2) and (3) for all the players.

First, we note that since \(r' > 0\), we can find \(v^S_i\) and \(v_i^M\) such that

\[
(1 - r')v_i + \frac{1}{2} r'v_i < v^S_i < (1 - r')v_i + \frac{3}{4} r'v_i
\]

\[
(1 - r')v_i + \frac{1}{4} r'v_i < v_i^M < (1 - r')v_i + \frac{1}{2} r'v_i
\]

We get the following inequality:

(*) \( v_i^M > (1 - r')v_i \)

In addition, we have:

(**) \( v^S_i < (1 - r')v_i + \frac{3}{4} r'v_i < v_i \)

Inequality (*) is inequality (1) for \(\varepsilon = 0\); inequality (**) is inequality (2) for \(\varepsilon = 0\) and \(\delta = 1\); and inequality (3) is also satisfied. Since the inequalities are satisfied strictly and since they are continuous in \(\delta\) and \(\varepsilon\), then for \(\delta\) close enough to 1 and \(\varepsilon\) close enough to 0 they will be satisfied as well. We might need to further increase \(\delta\) in order to have enough flexibility for choosing proper \(L_1\) and \(L_2\) such that:

\[
(1 - r')v_i + \frac{1}{2} r'v_i < \delta L_1 v_i < (1 - r')v_i + \frac{3}{4} r'v_i
\]
\[(1 - r')v_i + \frac{1}{2}r'v_i < \delta^L_2 v_i < (1 - r')v_i + \frac{1}{2}r'v_i\]

Note that all inequalities are strict, so that a deviation from the equilibrium instructions is strictly less profitable than following them. This is in order to make more clear that the players’s beliefs should be that the opponent is following the equilibrium path. For example, if we choose \(L_1 = L_2\), then a player observing a deviation is indifferent between reporting it and not reporting it. In this case, it makes more sense NOT to report, because if the deviator does not confess, the punishment is avoided. In other words, if one chooses \(L_1 = L_2\) then the ”reporting” strategy is weekly dominated by not reporting.

2.5 Discussion

2.5.1 Variety of Sanctions as a Tool to Enable Efficiency, Transparency and Trust

When monitoring is private, the transparency is lost, hence there is a need to establish some form of “trust” between the players. Communication enables us to restore this transparency. However, when punishments are too harsh, achieving truthful communication when shirking takes place becomes problematic. Restoring transparency and trust is due to the variety of sanctions that is available to the players. The possibility to “Pardon” and to reduce the punishment in case of truthful revelation of past “mistakes” is crucial.

This notion is in line with some basic intuitions. First, that flexibility is an important tool in long-lasting relationships; and second, it is plausible that truthful communication regarding past “mistakes” enhances the “trust” in the relationship, i.e., a deviation that is “confessed” is less destructive for future cooperation than a deviation that is not accompanied by a confession.
2.5.2 More Than Two Players and the Identity of the Deviator

If the game that is played is a game of more than two players, for example, when a principal is facing a team of agents, then often it is a challenge to establish the identity of the deviator. Since the punishment in our construction is a punishment for all the players - this problem is simply solved. If a deviation was detected, and none of the players “confessed” then all the players are being simultaneously punished with the harsh punishment.

2.5.3 The Connection to Fudenberg and Maskin (1991)

We use the equilibrium-path description of Fudenberg and Maskin, however, we obtain a weaker result - they obtain all the feasible individually rational payoffs as perfect equilibrium payoffs while we obtain only those Pareto dominating one-period equilibrium payoffs. The reason is that we use a different punishment system because of the imperfect monitoring. Their punishments are to minimax the deviator for a number of periods. Since in our construction we rely on the players reporting the deviations of their opponents, we cannot trivially use their method since in general the player reporting a deviation can profit from minimaxing the alleged deviator, which would trigger false reports.

2.5.4 Mutual Minimaxing and Other Threats

It is natural to consider other mutual threat-points instead of the one-period Nash equilibrium. However, attempting to implement the confession-and-pardon construction in order to deter deviations during the punishment phase will fail. It will fail because it requires three lengths of punishment. “Re-starting” a punishment phase in case a deviation during the punishment phase is detected will make deviations at the beginning of a long-term punishment profitable - it may replace the long punishment with a shorter one. Adding periods of punishments to the existing punishment phase will fail as well, since after large enough number of deviations, the weight of the additional punishment will become insignificant, and a deviation will become profitable.
Hence, stronger observation assumptions are required for the mutual-threat point that is not a one-period Nash equilibrium. We can instruct the players to add one period to the punishment in which a deviation is detected. That will make a deviation un-profitable if the detection probability is large enough compared to the possible profit. It will, on the other hand, lead the players to try and “avoid” detecting deviations.

For example, if there are two signals for player 1 that are consistent with the equilibrium path, but such that signal $a$ “detects” certain deviations but signal $b$ “detects” only a subset of those deviations, then player 1 will choose to report observing the “less informative” signal, signal $b$.

Exploring the possibility of enforcing a detection of deviation from the punishment phase calls for an extensive discussion regarding the “informativeness” of signals given the set of possible deviations and the observation mechanism. We feel that this is an interesting direction for further research.

2.6 Combined Theorem - Constant and Moving Support

In the papers of Kandori and Matsushima(1998) and Compte(1998), the authors prove several folk theorem for games with full support of the signals (all signal-profiles are observed with a positive probability after every action profile) and communication, when the number of players is at least 3. These results can be combined with ours in several ways, to enlarge the set of payoffs that can be supported as sequential equilibria payoffs. We first present those papers’ results and main ideas, then two examples to demonstrate the synergy between their methods and ours, and then the combined theorem.
2.6.1 The Results of Kandori and Matsushima (1998), and Compte (1998)

Both the paper of Kandori and Matsushima (1998) and the paper of Compte (1998) prove folk theorem for games with private monitoring, when communication is allowed and with full support of the private signals profiles. Both papers use dynamic programming techniques and the assumption of at least three players. The papers use delay of the communication (meaningful communication is carried on only every $k$ periods) to achieve efficiency.

Here are sufficient conditions under which there exists a folk theorem:

**First assumption:** Every deviation of a player $i$ from the common action minimaxing player $j$, $j \neq i$, is either not profitable or statistically detectable by player $j$’s opponents.

Officially: Let $\mu^i$ be the minimax profile for player $i$, when $\mu^j_i$ is the (possibly mixed) strategy of player $j$ when player $i$ is to be minimaxed.

(A1) - For all $i$ and $j \neq i$, if there is a mixed strategy $\alpha_j \in \Delta A_j$ such that $p_{-j} (\cdot | \mu^i) = p_{-j} (\cdot | \mu^i_{-j}, \alpha_j)$ then $g_j (\mu^i) \geq g_j (\mu^i_{-j}, \alpha_j)$.

**Second assumption:** All mixed strategy deviations of every player $i$, are statistically detected by the $i, j$ opponents, for every $j \neq i$. Define, for each pair $i \neq j$ and each action profile $a \in A$, $Q_{ij}(a) = \{ p_{-ij}(a_{-i}, a'_i) | a'_i \in A_i \setminus a_i \}$. This is a collection of distributions of $ij$-opponents’ signals, generated by player $i$’s deviations from the profile $a$.

(A2) - For each player $i \neq j$ and each $a \in EX$,

$p_{-ij}(a) \notin \text{conv} (Q_{ij}(a) \cup Q_{ji}(a))$.

**Third assumption:** For every two players $i$ and $j \neq i$, the opponents of $i$ and $j$ can statistically discriminate player $i$’s (possibly mixed) deviations from player $j$’s. The deviations of the different players create different distributions of the signals of their opponents.

(A3) - For each pair $i \neq j$ and each $a \in EX(A),$

\text{If there is more than one strategy profile that minimaxes player $i$, then there has to be at least one that follows the axioms.}
Let $v^*_i$ be the minimax value of player $i$ and define the feasible and individually rational payoff set by

$$W = \{ v \in co(g(A)) | v \geq v^* \}.$$  

Assume perfect support of the private signals profiles.

The main theorem is:

**Theorem (Kandori and Matsushima):** Suppose that there are more than two players $(n > 2)$ and the information structure satisfies condition (A1), (A2) and (A3). Also suppose that the dimension of $W$ is equal to the number of players. Then, any interior point in $W$ can be achieved as a sequential equilibrium average payoff profile of the repeated game with communication, if the discount factor $\delta$ is close enough to 1.

### 2.6.2 Using Confessions and Reports Method to Support Dynamic Programming Methods

Consider the following game:

<table>
<thead>
<tr>
<th></th>
<th>l</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>1,0,0</td>
<td>0,1,0</td>
</tr>
<tr>
<td></td>
<td>0,0,0</td>
<td>0,0,1</td>
</tr>
<tr>
<td>R</td>
<td>0,0,0</td>
<td>0,1,0</td>
</tr>
<tr>
<td></td>
<td>1,0,0</td>
<td>1,0,0</td>
</tr>
</tbody>
</table>

Assume that the signals to the three players are according to assumptions (A1), (A2) and (A3).

The one-period equilibrium is when each player randomizes with probability half for each action, and the payoff is $(1/4, 1/4, 1/4)$.

Now consider the following addition to the above game:

<table>
<thead>
<tr>
<th></th>
<th>l</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>1,0,0</td>
<td>0,1,0</td>
</tr>
<tr>
<td></td>
<td>0,0,0</td>
<td>0,0,1</td>
</tr>
<tr>
<td></td>
<td>5,-7,7</td>
<td>5,-7,-7</td>
</tr>
<tr>
<td>R</td>
<td>0,0,1</td>
<td>0,0,0</td>
</tr>
<tr>
<td></td>
<td>0,1,0</td>
<td>1,0,0</td>
</tr>
<tr>
<td></td>
<td>0,0,0</td>
<td>0,0,0</td>
</tr>
</tbody>
</table>
Note that now there is an additional equilibrium \((bb, l, R)\) with the payoff \((0, 0, 0)\).

Assume that we add now another private signal for player 2. Assume that when player 3 plays \(L\) and player 1 plays \(bb\), this additional private signal is observed by player 2 and that the signal is observed with probability 0 when player 1 does not play the additional action, \(bb\). Since the convex-hull of the original game is Pareto dominating the one-period equilibrium, we can still have the entire set of feasible individually rational payoffs as sequential equilibria payoffs. The set of feasible individually rational payoffs are all in the convex-hull of the original game (without the additional action). We can get all the payoffs of the original game through the method of Kandori and Matsushima, and in case player 1 deviates to \(bb\) when player 3 plays \(L\), we can use our method of confession and reports - player 3 will convey a signal whose meaning is that a deviation took place, and player 1 will confess (the one period equilibrium that will be used as a punishment can be \((0, 0, 0)\)). Under that construction, when player 3 plays \(R\), there is no reason for player 1 to play his additional action, \(bb\).

2.6.3 Supporting Confession and Reports Method with Dynamic Programming Methods

Consider the following version of the prisoners’ dilemma. The signals can take the values 1 or 0:

<table>
<thead>
<tr>
<th></th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2,2</td>
<td>1-L,2+H</td>
</tr>
<tr>
<td>D</td>
<td>2+H,1-L</td>
<td>1,1</td>
</tr>
</tbody>
</table>

Kandori and Matsushimam (1998) proved folk theorem for this specific game, under the following monitoring-technology conditions:

- The signals of the players are independent given any pure action profile.
The marginal distributions of the private signal of the players are symmetric, and
\[ p_1(1|D, d) > p_1(1|D, c) \] and \[ p_1(1|D, c) > p_1(1|C, c) \].

Now consider the game with additional actions to the players:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3,1-0.5L</td>
<td>0,-1</td>
<td>2,2</td>
<td>1-L,2+H</td>
</tr>
<tr>
<td>D</td>
<td>0,-1</td>
<td>-1,-1</td>
<td>-1,0</td>
<td>-1,0</td>
</tr>
</tbody>
</table>

The minimax payoff is (0, 0)

![Figure 1. - Prisoner’s Dilemma with additional actions](image)

In order to support the entire efficient frontier as sequential equilibria payoffs, the common action (C, a) should be supported (see figure 1). The payoff (3, 1 - 0.5L) does not dominate the one-period equilibrium payoff (1, 1), however, if we assume that a deviation of player 2 from a to c or d when player 1 plays C induces a positive probability player 2 to observe a signal that is not sufficiently informative, we can still support this common action.

To see how, first note that now the entire convex-hull of \{(2, 2), (1 - L, 2 + H), (2 + H, 1 - L), (1, 1)\} is individually rational. Let this convex-hull be U. Looking carefully at the construction of Kandori and Matsushima, one can verify that it still holds for the entire U.

Second, note that the payoff (3, 1 - 0.5L) is Pareto dominating a two-dimensional non-empty subset of U. We can now replace the three lengths of punishments with three different continuation payoffs. There is a \( \gamma > 0.5 \) such that the payoff (2 + 0.5H, 1 - \( \gamma L \)) is in the interior of U. We shall pick our three possible continuation payoffs, which are analog to the three lengths of punishment on the line connecting (2 + 0.5H, 1 - \( \gamma L \)) and (3, 1 - 0.5L). The
continuation payoff \((2 + 0.5H, 1 - \gamma L)\) will be played in case player 1 announces observing the signal that is observed when player 1 deviates to playing \(a\) and player 2 is not confessing (long-term punishment) and two other points, closer to \((3, 1 - 0.5L)\) on this line can be chosen to supply the incentives for player 1 to confess and for player 2 to report a deviation (short-term and medium-term punishments analogs) if the players are patient enough. The logic of the proof is the same.

In this case, we support the confession-and-report construction not by the one-period equilibrium punishments, but rather by a set of payoffs that is itself achieved via dynamic programming construction. Note that this set has to be of dimension \(n\).

### 2.6.4 The General Construction

In general, it is easy to see that one can use the following algorithm to find out the set of payoffs that can be supported as sequential equilibria payoffs when communication is allowed (denote it \(E\), by combining dynamic-programming and confessions-and-reports methods:

1. Let \(E\) be the convex-hull of the one-period Nash Equilibrium payoffs. Those are naturally included in the set of sequential equilibrium payoffs.

2. Add to \(E\) the convex-hull of the payoffs of all sub-matrices for which all the following conditions hold:

   a. Follow the conditions of Kandori and Matsushima (and therefore the convex-hull of their payoffs can be supported using Kandori and Matsushima’s methods, unless a player deviates to an action that is outside the support of the sub-matrix.

   b. Any deviation from the sub-matrix to an action that is outside the support of the sub-matrix is either unprofitable or detectable (detectable in the sense of inducing a probability of less than one for the deviating player to obtain a signal that is sufficiently informative).

   c. Any payoff in the sub-matrix strictly Pareto dominates either one-period Nash equilibrium payoff or three payoffs in the existing \(E\) such that one Pareto dominates the second which in turn Pareto dominates the third. This will allow the players to have three dif-
ferent punishments and hence enables them to use confession-and-pardon methods to deter deviations to actions outside the support of the sub-matrix.

3. Go back to 2, until no more such sub-matrices can be found.

The proof follows the same logic as the two examples above.

The result of this algorithm is a set of payoffs that can be supported via sequential equilibrium. Further work is needed in order to obtain conditions for a proper folk theorem, in combined methods, but we merely wanted to point the possibility of such a combination.
Chapter 3

Observable Deterministic Payoffs

3.1 Introduction

Consider the following game, where Alice and Bob choose separately whether to go to the opera, the boxing place or to stay at home. Bob prefers boxing over the opera, and prefers the opera over staying at home. On the other hand, he enjoys the opera more when Alice is there, but she absolutely destroys all joy he has from boxing. As for Alice, she does not care much for boxing, neither does she care for Bob. She likes the opera, but she prefers to stay at home if Bob is at the opera. We can sum up this story in the following payoff-matrix:

<table>
<thead>
<tr>
<th>Alice \ Bob</th>
<th>Opera</th>
<th>Boxing</th>
<th>Home</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opera</td>
<td>-1,4</td>
<td>3,3</td>
<td>3,0</td>
</tr>
<tr>
<td>Boxing</td>
<td>0,2</td>
<td>-1,-1</td>
<td>0,0</td>
</tr>
<tr>
<td>Home</td>
<td>0,2</td>
<td>0,3</td>
<td>0,0</td>
</tr>
</tbody>
</table>

The traditional folk theorem tells us that when monitoring is perfect, any payoff Pareto dominating the minimax point (0,2) can be achieved via perfect equilibrium in a repeated framework, when players are patient enough. But in this story, why should the monitoring be perfect? For example, why would Alice, when going to the opera, know whether Bob
stayed at home or went to the Boxing? All she observes is that he is not at the opera. What is, in that case, the set of sequential equilibrium payoffs?

Consider any duopoly model with a known demand and secret price-cuts. Each firm knows its own profits, but does not observe the opponent’s actions. If there is a number of products sold, the firms can offer menus, bundle products etc. Is the information given by the total demand and own’s profits alone enough to support a large set sequential equilibrium payoffs in a repeated framework?

The model we study in this chapter is a model where following each period, each player observes only his own payoff. In addition, given the pair of actions taken by the players, the payoffs are deterministic. This model was introduced by Lehrer (1992), who analyzed it for the case where players do not discount future payoffs, and also studied by Tomala (1999), who obtained results that concern the set of Nash-Equilibrium payoffs (as opposed to sequential equilibrium payoffs) of the discounted repeated game\(^1\).

We refer to the convex-hull of the payoffs of the one-shot game as the set of feasible payoffs, and the set of payoffs that Pareto-dominate the maximal payoffs the players can each defend when being minimaxed in pure strategies as individually rational payoffs\(^2\). We show that, all individually rational payoffs on the strictly efficient frontier can be obtained as sequential equilibria payoffs, in games where one can have as a continuation payoff at least one payoff that is internal to the set of feasible payoffs. We also show how such a continuation payoff can be obtained in all but a set of degenerate games.

Since any profitable deviation from the strictly efficient frontier implies a lower payoff for the opponent, it is immediately detectable. Hence, the challenge in this model is the construction of the punishment. If we merely instruct the punishing player to play the strategy which minimaxes the opponent, he may profitably deviate without being detected.

\(^1\)Tomala assumed that each player also observes the payoffs of the other players, that is, that the entire payoff-vector that was obtained is publicly known following each period.

\(^2\)The equilibrium construction heavily relies on the pureness of the minimaxing strategies, and we don’t see any obvious way to avoid it.
In order to resolve the problem of possible profitable deviations of the punishing player, we instruct the punishing player to do exactly that, meaning to play with some (positive) probability the profitable deviations from the minimaxing. The punishing player will randomize between the minimaxing action and some profitable deviations from it. Of course, playing a profitable deviation induces a higher payoff to the punishing player, so we will need to balance it with appropriate continuation payoffs.

Hence, different actions taken by the punishing player during the punishment should be followed by different continuation payoffs. Naturally, the punishing player cannot be trusted with the task of determining his own continuation payoffs, so the player being punished should be the one to do it. In order to be able to do that, the player being punished should observe different distributions of signals for the different actions of the punishing player. The action that is the best response for being minimaxed may not be informative enough to allow him to design the appropriate continuation payoffs for his opponent (the punisher).

Thus the punished player might be asked to play an action that is not his original best response against the pure minimaxing. We wouldn’t want to enforce any further punishments for a player who is currently being punished\(^3\), so we make sure that all the actions the punished player is being instructed to use during the punishment are indeed best responses to the (now possibly mixed) action taken by the punisher. In section 3.3 we show how such a punishment phase can be constructed for an example, and in section 3.4 we show it for the general case.

In order to obtain different continuation payoffs we shall need a two-dimensional set of continuation payoffs. In section 3.5 we show how to obtain such a set in all games but a set of degenerate games.

Having constructed the punishment phase and the continuation payoffs, we describe in section 3.5.3 how the players can communicate the continuation payoffs. In section 3.6 we

\(^3\)Because of the discounting, and because we wish to obtain a sequential equilibrium, punishing a player for not cooperating with his own punishment is hard to implement. Eventually the punishment that we will "pile up" for not cooperating with former punishments will be so far in the future game, that the discounting will make it insignificant, and the construction will unravel.
summarize and obtain the main result: all payoffs on the strictly efficient frontier can be achieved via sequential equilibria, if such two-dimensional set exists.

3.2 The Model and The Results

3.2.1 The Stage Game

We consider a repeated discounted two-players game. Let \( \{1, 2\} \) be the set of players, and \( A_1 \) and \( A_2 \) be the sets of actions available at each period to player 1 and player 2 respectively. Let \( |A_1| \) and \( |A_2| \) be the number of actions in those sets, and \( P(A_i) \) be a distribution over the set \( A_i \). Let \( u_i(a_1, a_2) \) be the payoff for player \( i \), \( i \in \{1, 2\} \) when player 1 plays action \( a_1 \) and player 2 action \( a_2 \). The payoffs are deterministic given the actions taken by the players. Let \( V \) be the convex-hull of all feasible payoffs. Recall that the payoffs that are on the strictly efficient frontier are payoffs \( (v_1, v_2) \) that are feasible, and such that for every payoff \( (v_1', v_2') \), if \( v_1' > v_1 \) and \( v_2' \geq v_2 \), then \( (v_1', v_2') \not\in V \), and if \( v_1' \geq v_1 \) and \( v_2' > v_2' \), then \( (v_1', v_2) \not\in V \).

Normalize the payoffs of the game so that zero will be the minimax payoff of the players, when the minimaxing strategy is pure. We refer to a pair of payoffs as strictly individually rational if both payoffs are positive.

3.2.2 The Repeated Game

At each period \( t = 1, 2, ... \) the stage game is played. At the beginning of each period each player \( i = 1, 2 \) takes action \( a^t_i \) and at the end of the period he observes his own payoffs \( u_i(a^t_1, a^t_2) \). At the end of period \( t \), the private history of player \( i \) consists of player \( i \)’s past actions and past payoffs.

\[
h^t_i = a^t_1, u_i(a^t_1, a^t_2), a^t_2, u_i(a^t_1, a^t_2), ..., a^t_i, u_i(a^t_1, a^t_2).
\]

We denote by \( h^0_i \) the null private history of player \( i \). We denote by \( h \) the sequence of actions and payoffs taken so far, and by \( \mathcal{H} \) the set of all possible histories.
A strategy $\sigma_i$ of player $i$ is a sequence of maps $\{\sigma^t_i\}_{t=1}^{\infty}$, where $\sigma^t_i$ maps each private history that ends with a private signal (which is, in our case, a payoff) to a distribution over actions in $A_i$ to be taken in the next period.

Formally, a strategy is $\sigma^t_i$:

$$\sigma^t_i : \times^t_{t'=1} (a^t_i \times u^t_i(a^t_1, a^t_2)) \rightarrow P(A_i)$$

Each strategy profile $\sigma = (\sigma_1, \sigma_2)$ generates a probability distribution over future streams of actions, which, in turn, induces a distribution over future payoffs. Players are assumed to discount future payoffs with a common discount factor $\delta$.

Player $i$’s discounted expected payoff using the strategy $\sigma$ is:

$$v_i(\sigma) = (1 - \delta) E\left[\sum_{t \geq 1} \delta^{t-1} u_i(a^t_1, a^t_2)\right]$$

### 3.2.3 Sequential Equilibrium

A strategy profile $\sigma$ is a Nash Equilibrium if and only if, for any player $i \in \{1, 2\}$, and for any strategy $\sigma'_i$, $v_i(\sigma_i, \sigma_{\{1,2\}\setminus i}) \geq v_i(\sigma'_i, \sigma_{\{1,2\}\setminus i})$.

An assessment is a pair $(\sigma, \mu)$ where $\sigma$ is a profile of behavioral strategies and $\mu$ is a function that assigns to every information set a probability measure over the set of histories in the information set. We shall refer to $\mu(h_i, h)$ as the beliefs of the players, that is, the probabilities a player assigns to the history $h \in \mathcal{H}$ conditional on the private history $h_i$ being observed.

The assessment $(\sigma, \mu)$ is sequentially rational if for every player $i \in \{1, 2\}$ and for every information set of player $i$, the strategy of player $i$ is a best response to the strategy of the other player, given the information set.

An assessment is consistent if there is a sequence $((\sigma^n, \mu^n)_{n=1}^{\infty})$ of assessments that converges to $(\sigma, \mu)$ in Euclidian space, and has the properties that each strategy profile $(\sigma^n)$ is completely mixed (meaning that it assigns a positive probability to every action at every information set), and that each belief system $\mu^n$ is derived from $\sigma^n$ using Bayes rule.

An assessment is a sequential equilibrium if it is sequentially rational and consistent.
3.2.4 The Result

We define $\Gamma$ to be the set of repeated games with deterministic observable own-payoffs, where a payoff that is an internal point of the set of feasible payoffs of the game can be obtained as a sequential equilibrium payoff, and using strategies such that the continuation payoff of the players is always within an $\epsilon$-distance from the target payoff, an $\epsilon$ that approaches zero as the discount factor $\delta$ approaches 1.

Formally, let $\Gamma$ be the set of repeated games with deterministic observable own-payoffs, such that there exists at least one payoff $(v'_1, v'_2)$ and for every $\epsilon > 0$ there exists a discount factor $\delta'$ and a sequential equilibrium $(\sigma_\epsilon, \tau_\epsilon)$ for which the following three conditions hold:

1. $(\sigma_\epsilon, \tau_\epsilon)$ is a sequential equilibrium for any discount factor $\delta > \delta'$.

2. The payoff of this equilibrium, $v(\sigma_\epsilon, \tau_\epsilon) = (v'_1, v'_2)$ is not on the efficient frontier of the set of payoffs.

3. The continuation payoffs when the players play $(\sigma_\epsilon, \tau_\epsilon)$ are within an $\epsilon$ distance from $v(\sigma_\epsilon, \tau_\epsilon)$ when the discount factor is $\delta > \delta'$.

The set $\Gamma$ includes, for example, all games where at least one of the following three conditions hold:

1. games where there is at least one one-period Nash equilibrium whose payoff is internal to the set of feasible payoffs.

2. games where the strictly efficient frontier is piecewise linear but not linear.

3. games where there exists a pair of distributions, one over the actions of player 1, $p = (p_1, ..., p_{|A_1|})$, and one over the actions of player 2, $q = (q_1, ..., q_{|A_2|})$, that satisfies all the following conditions:

   a. $u(p, q)$ is internal to the set of feasible payoffs.

   b. all actions that are played with positive probabilities have the same expected payoff when played against the opponent’s distribution, that is,

   $\forall a_1, a'_1, \ p(a_1) > 0, p(a'_1) > 0 \Rightarrow u_1(a_1, q) = u_1(a'_1, q)$

   $\forall a_2, a'_2, \ q(a_2) > 0, q(a'_2) > 0 \Rightarrow u_2(p, a_2) = u_2(p, a'_2)$
c. playing an action that is not in the support of $p$ or $q$, and that has a higher payoff than those that are in the support (when played against the distribution of the opponent) induces a positive probability for both players to observe simultaneously a payoff that cannot be observed if only actions in the support was played. Formally:

$$\forall a'_1, p(a'_1) = 0, u_1(a'_1, q) > u_1(p, q) \Rightarrow \exists a_2, q(a_2) > 0, \forall a_1, p(a_1) > 0, u_1(a'_1, a_2) \neq u_1(a_1, a_2), u_2(a'_1, a_2) \neq u_2(a_1, a_2)$$

$$\forall a'_2, q(a'_2) = 0, u_2(p, a'_2) > u_2(p, q) \Rightarrow \exists a_1, p(a_1) > 0, \forall a_2, q(a_2) > 0, u_1(a_1, a'_2) \neq u_1(a_1, a_2), u_2(a_1, a'_2) \neq u_2(a_1, a_2)$$

Condition b means that the players are indifferent between all the actions in the support of the distributions, and condition c means that any profitable deviation is detectable with a positive probability, and this detection is common knowledge.

Note that condition 1 (an internal Nash equilibrium) is a private case of condition 3, since there are no profitable deviations from a one-period Nash equilibrium.

In section 3.7 we show how an internal point can be obtained as sequential equilibrium with the required continuation payoffs in some games where none of the conditions 1-3 hold.

**Theorem 3.1:** In any game $G \in \Gamma$, any payoff on the strictly efficient frontier that is strictly individually rational can be obtained as a sequential equilibrium if the players are sufficiently patient.

### 3.2.5 An Overview of the Equilibrium Construction

The equilibrium construction for payoff $(v^e_1, v^e_2)$ on the strictly efficient frontier and such that $v^e_1 > 0$, $v^e_2 > 0$:

1. play a sequence of pure strategies that have payoffs of the strictly efficient frontier, a sequence whose total discounted payoff is $(v^e_1, v^e_2)$ and such that the continuation payoffs are always within an $\varepsilon$-distance from $(v^e_1, v^e_2)$.

---

4 The details of such a construction can be found in Fudenberg and Maskin (1991)
2. If a player deviates, play according to the minimaxing instructions that are described in section 3.3 for $N$ periods. As explained in the following, during the punishment phase, the punished player updates the continuation payoff of his opponent. If the punishing player deviates during the punishment and his deviation is detected, then his continuation payoff is reduced. The player being punished is playing only best-responses, so he has no profitable deviations.

3. play the communication periods (one to four periods, depending on the matrix, the details are found in section 3.5.3). If a detection of the deviation is reported, then the players go back to the punishment phase (phase 2 in this list).

4. play the continuation payoffs strategies, obtaining payoffs within the set $W$, as explained in section 3.5.2. In case there is a profitable deviation in this phase, then this is common-knowledge, and the players move back to the punishment phase.

### 3.3 The Punishment Phase with Terminal Payoffs - an Example

In this section we begin to deal with the main difficulty of the chapter: constructing the punishment strategies for the players.

We first consider a case where the punishing player can have different terminal payoffs following the punishment. That is, we assume that when a profitable deviation takes place (an event that is common knowledge) the players begin to play the following game:

- the set of players is $\{1, 2\}$ as in the original game.
- the set of actions available to the player are the same as in the original game $A_1$ and $A_2$.
- the players discount their payoffs using the same common discount factor $\delta$ (again, as in the original game).
- this game is finite, and it ends after \( M \) periods (the value of \( M \) will be established later), each player receives a terminal payoff. The terminal payoff for the player who is being punished is constant, and therefore has no effect on his actions, while the terminal payoff of the punishing player is determined by the player being punished. Formally, let \( Y_1 \) and \( Y_2 \) be the terminal payoffs of player 1 and player 2, respectively, then player \( i \) evaluates the stream of payoffs of this finitely repeated game this way by\(^5\):

\[
\sum_{t=0}^{\infty} (1 - \delta) \delta^t u_i (a^t_1, a^t_2) + \delta^M Y_i.
\]

After the players receive their terminal payoffs the game ends. The terminal payoff of the punishing player are determined by the punished player, so that to obtain the different terminal payoffs for different actions, the punished player should be able to differentiate between those actions. In the sections that will follow, we show how such terminal payoffs are obtained as continuation payoffs in the infinitely repeated game, and how they are communicated.

**Lemma 3.1:** In any game, there exists a strategy for player 2 that minimaxes player 1, such that the following three conditions are fulfilled:

a. player 1 plays only actions that are best-responses (and the payoff he obtains in all those best-responses is his minimax payoff, zero).

b. player 1 can construct different terminal payoffs for player 2, given his private history during the punishment, terminal payoffs that will make player 2 indifferent between all the actions he is instructed to play during the punishment.

c. Given the terminal payoffs, there is no profitable deviation for player 2 that is not detected with a positive probability by player 1.

The symmetric result, for player 1 minimaxing player 2, holds, naturally, as well.

\(^5\)The weight of the terminal payoff is not multiplied by \((1 - \delta)\) for the sake of making the transition from terminal payoffs to continuation payoffs easier. However, we make no assumptions yet regarding the set of available terminal payoffs, and the proof doesn’t change much if one defines \( Y' = Y(1 - \delta) \).
The proof of the lemma is constructive. We detail the algorithm in the next section, and we demonstrate the main ideas using Example 3.1.

Example 3.1:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.0</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>B</td>
<td>-1.1</td>
<td>3.2</td>
<td>2.1</td>
<td>0.1</td>
<td>-1.0</td>
<td>3.2</td>
</tr>
<tr>
<td>C</td>
<td>-3.1</td>
<td>-3.0</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>-3.1</td>
</tr>
<tr>
<td>D</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

This example can be solved in four phases:

3.3.1 Phase 1 - The Original Minimaxing

Assume that player 2 needs to punish player 1. The (pure) strategy that player 2 can use to minimax player 1 is a. If we instruct player 2 to play this strategy during the punishment, then player 1 will want to play a best response - action A, action D, or any randomization over A and D. We wish to avoid enforcing yet another punishment for player 1 in case he deviates during his own punishment, so we will always instruct him to play a best response. However, if player 1 indeed plays a best response, then player 2 can deviate to playing b or c, for example, without being detected. If player 1 plays with some probability D, then a deviation of player 2 to the action d is detected with that probability.

We will instruct player 1 to play action A with some probability $0 < p(A) < 1$, and action D with probability $p(D) = 1 - p(A)$.

These instructions make only a deviation of player 2 to action d detectable.

There are still four profitable deviations left: b, c, e and f.
3.3.2 Phase 2: Adding a Second Column

Consider asking player 2, instead of playing action $a$, to play with some probability action $a$ and with some probability action $b$. That is, consider the following perturbation of action $a$: $(1 - \varepsilon)a + \varepsilon b$. For a small enough $\varepsilon$ this is still an action profile which minimaxes player 1: in all rows where the payoff of player 1 against action $a$ is zero, rows $A$ and $D$, it is still zero when played against $b$ (otherwise, $b$ would have been a deviation that is detectable through best-responses), and for all other rows, where the payoff of player 1 is strictly less than zero - a small enough perturbation will keep the expected payoff lower than zero.

However, if the instructions of player 2 are to randomize during each period between actions $a$ and $b$, then player 2 will need to be indifferent between those actions. Action $b$ has higher payoff, so in order to make player 2 indifferent it will need to have lower terminal payoff. But player 1 is the one who should determine the terminal payoff of player 2, yet he cannot differentiate between action $a$ and action $b$. Therefore, player 2 will prefer to play action $b$. In this case, we will increase the probability of playing $b$ (and decrease respectively the probability of playing $a$), that is, we increase $\varepsilon$. We increase it until either the probability of action $a$ will go down to zero, or until another row’s expected payoff (for player 1) will go up to zero (whatever happens first). In our case, the payoff of row $B$ becomes zero when $\varepsilon = \frac{1}{4}$.

Now, when the payoff of row $B$ went up to zero, it became a best-response. When player 1 will play the new mixture of best-responses: $A$, $B$ and $D$, he can differentiate between action $a$ and action $b$, because they induce different payoffs when played against $B$.

Let $p(B)$ be the probability with which player 1 plays action $B$ during the punishment phase, and let $u(a)$ and $u(b)$ denote the expected payoff for player 2 when he plays actions $a$ and $b$ respectively against the profile that player 1 is playing. In order to make player 2 indifferent at period $t$ of the punishment between playing $a$ and $b$, player 1 will decrease the terminal payoff of player 2, whenever he plays $B$ and observes 3, by:

$$\frac{u(b) - u(a)}{p(B)J^t}$$
CHAPTER 3. OBSERVABLE DETERMINISTIC PAYOFFS

Having established these instructions, the question is, are there still profitable unobservable deviations for player 2?

Player 2 is instructed to play actions \( a \) and \( b \); a deviation to \( c \) is observable with a positive probability (it induces a different payoff against action \( B \)); a deviation to action \( d \) is observable when player 1 plays \( D \); it remains to check actions \( e \) and \( f \).

Action \( e \) induces the same payoffs to player 1 as action \( a \) against all actions player 1 is instructed to play, \( A, B \) and \( D \). Those payoffs, being also the observations of player 1 induce the same terminal payoffs as action \( a \). In addition, it has higher expected payoff against the actions of player 1. This combination makes action \( e \) a profitable deviation that is not detected through the current actions player 1 is instructed to play.

Thus we add action \( e \) to the support of the punishment, during phase 3:

3.3.3 Phase 3: Adding a Third Column

Player 1 cannot design terminal payoffs for player 2 such that the action \( e \) will not be a profitable deviation. This is because his observations when player 2 plays \( a \) and when he plays \( e \) are identical. But this is also the reason why we can take away a small part of the probability of playing \( a \), say \( \zeta \), and play with that probability action \( e \), without changing neither the fact that player 1 is being minimaxed, nor the set of best-responses of player 1.

Still, player 2 will prefer to play \( e \) over \( a \).

Again, we can increase the probability of playing \( e, \zeta \), while decreasing the probability of playing action \( a \) until, in this case simultaneously, both the probability of action \( a \) decreases to zero, and the payoff of row \( C \) goes up to zero, when \( \zeta = \frac{3}{4} \).

Player 1 will adjust the terminal payoffs of player 2, such that whenever player 1 plays \( C \) and observes 1, the terminal payoff of player 2 will change by:

\[
\frac{u(b) - u(e)}{p(C)\delta M - t}
\]

This makes player 2 indifferent between \( b \) and \( e \), while player 1 is being minimaxed and is using only best responses.
3.3.4 Phase 4: Another Column

Before we added column $e$ and row $C$ to the support of the punishment, column $b$ and column $f$ had the same current payoff and the same terminal payoffs, so it was not strictly profitable to deviate from $b$ to $f$. Now that row $C$ was added, column $f$ became more profitable than $b$. In addition, playing $f$ is a deviation that cannot be detected using the actions that player 1 is currently instructed to play. In this example, we can simply replace columns $f$ and $b$, that is, we can play $f$ with the probability that $b$ was supposed to be played.

3.3.5 The Strategies Construction

The strategies that were constructed by the algorithm for the minimaxing of player 1 are, then:

- player 2 should play with probability $\frac{1}{4}$ action $f$ and with probability $\frac{3}{4}$ action $e$.
- player 1 should randomize between all his actions, $A, B, C$ and $D$, using any full-support distribution over them. He should update the terminal payoffs of player 2 and add to it $\frac{u(f) - u(e)}{p(B)\delta^M}$ or $\frac{u(f) - u(b)}{p(C)\delta^M}$ whenever he played $B$ or $C$ (respectively) and observed 3 or $-3$ (resp.).

3.4 The Punishment with Terminal Payoffs

In this section, we detail the algorithm of the construction of the minimaxing strategies, that will be played during every period of the punishment phase. We still assume that when the $M$ periods of the punishment end, the punishing player, player 2, receives his terminal payoff, a terminal payoff that is determined by player 1.

A general description of the algorithm goes as follows:

- begin with instructing player 2 to play a pure action that minimaxes player 1, and instructing player 1 to mix over all his best-responses.
if there is a profitable undetectable deviation of player 2 then continue by performing at each phase procedure A if possible, and if not possible then perform procedure B. Do it until no profitable deviation exist:

Procedure A: design terminal payoffs for player 2, based on the observations of player 1, that will make player 2 indifferent between all the actions that he is instructed to play, and that will make the currently profitable deviation not profitable anymore.

Procedure B: increase the probability of the last column added, which is currently the most profitable one (while changing the probabilities of the other columns and while keeping all best-responses as best-responses), until either the probability of some column goes down to zero, or until some row’s payoff goes up to zero and this row is added to the support of the rows played during the punishment phase. This way, we can add rows to the support and/or take out columns from the support until the proper terminal payoff for player 2 can be constructed.

When the algorithm is complete, the strategies during the punishment phase will be such that all the conditions of lemma 3.1 are fulfilled.

3.4.1 The Algorithm

Let $C$ be the set of actions that player 2 is instructed to play during each period of the punishment with a positive probability, and let $p(C)$ be the distribution over those actions. Let $R$ be the set of actions that player 1 is instructed to play during each period of the punishment with a positive probability, and let $p(R)$ be the distribution over those actions.

a. begin the algorithm with $C \equiv$ the pure minimaxing action (if there are more than one, then choose any of them), and with $R \equiv$ the set of all best responses to $C$, with any full-support $p(R)$.

b. If terminal payoffs for player 2 can be constructed in the way that is described in section 3.4.2, then build them. Those are terminal payoffs such that player 2 will be indifferent between all actions in $C$ and weakly prefer them over the actions in $A_2 \setminus C$. 
c. If such terminal payoffs cannot be constructed, then iterate changing the sets \( R \) and \( C \) as explained in section 3.4.3, until appropriate terminal payoffs are established.

### 3.4.2 When are Different Terminal Payoffs Possible

The terminal payoff of player 2 can be updated whenever player 1 plays an action and observes a signal (which is his payoff in our model). Each pair of an action of player 1 and a possible observation for player 1 when playing that action is a potential updating situation.

Let us enumerate all the different updating situations by 1,2,3... . Let \( \Delta_k \) be the change of the terminal payoff for player 2 when player 1 plays and observes the pair numbered by \( k \). That is, when updating situation \( k \) occurs at period \( t \) of the punishment, the terminal payoff of player 2, then \( \frac{\Delta_k}{\delta^{M-t}} \) is added to the terminal payoff. Let \( S(R) \) be the set of pairs that can be observed using actions in \( R \), and let \( p(k) \) be the probability of observing \( k \), when the column where \( k \) can be observed is played, which is the probability of player 1 playing the row where \( k \) can be observed.

Let \( I_{k,j} \) be the function that takes the value of 1 if updating situation \( k \) can occur when column \( j \) is played, and takes the value 0 otherwise. Let \( C_1 \) to be one of the columns in \( C \) (\( C \) is never empty. Initially it contains the original minimax, and as soon will be clear, it cannot get emptied during the algorithm). Let \( u(j) \) be the (one-period) payoff-expectancy for player 2 when he plays column \( j \) against the best-responds mixture that player 1 is instructed to play. For player 2 to be indifferent between all the actions in \( C \), it has to be that for each column \( j \in C \) and for each period \( t \) of the punishment, the following equality holds:

\[
(*) \sum_{k \in S(R)} p(k) I_{k,j} \frac{\Delta_k}{\delta^{M-t}} = u(C_1) - u(j).
\]

Let \( c' \) be the column that is currently the profitable undetectable deviation of player 2. We can eliminate the profitability of this deviation by changing the terminal payoffs if we can solve simultaneously the equalities (*) and the following inequality:

\[
\sum_{k \in S(R)} p(k) I_{k,c'} \frac{\Delta_k}{\delta^{M-t}} \leq u(C_1) - u(c')
\]
We will be able to find $\Delta_k$’s that solve it in equality (together with equalities (*)) if the columns of the matrix $I_{k,j}$, $k \in S(R), j \in C \cup c'$ are linearly independent. In this case, we will add $c'$ to $C$, and then we will check whether other profitable deviations still exist.

To better understand the nature of this matrix, consider what this matrix is for example 3.1 after phase 3, when $C = \{b, e\}$, $R = \{A, B, C, D\}$ and the profitable undetected deviation is $c' = f$.

First, the enumeration:

- number 1 - playing $A$ and observing 0
- number 2 - playing $B$ and observing 3
- number 3 - playing $B$ and observing -1
- number 4 - playing $C$ and observing -3
- number 5 - playing $C$ and observing 1
- number 6 - playing $D$ and observing 0

And the matrix:

\[
\begin{array}{ccc}
  & b & e & f \\
1 & 1 & 1 & 1 \\
2 & 1 & 0 & 1 \\
3 & 0 & 1 & 0 \\
4 & 1 & 0 & 1 \\
5 & 0 & 1 & 0 \\
6 & 1 & 1 & 1 \\
\end{array}
\]

The columns of this matrix are not independent, since columns $b$ and $f$ have the same payoffs for player 1.

Note that the number of 1’s in each column always equals the number of rows in $R$. Denote this number by $|R|$. 

3.4.3 Changing $C$ and $R$ when Proper Terminal Payoffs Cannot be Found

In case the columns of the matrix above are linearly dependent, then we can divide the set of the columns into two disjoint sets: $K \cup K' = C \cup c'$, such that there exist $\lambda_j \geq 0$, not all zero, such that for every row $k$ the following holds:

\[ (** ) \sum_{j \in K, k \in S(R)} \lambda_j I_{k,j} = \sum_{k \in S(R), j \in K'} \lambda_j I_{k,j}. \]

We will first show that the weights $\lambda_j$ can be chosen such that both hand-sides of the equality (**) describe convex combinations of columns, i.e., we will show that $\sum_{j \in K} \lambda_j = \sum_{j \in K'} \lambda_j$.

**Lemma 3.2:** $\sum_{j \in K} \lambda_j = \sum_{j \in K'} \lambda_j$.

**proof:** Sum the above (**) equalities over all the rows:

\[ \sum_{k \in S(R)} \sum_{j \in K} \lambda_j I_{k,j} = \sum_{k \in S(R)} \sum_{j \in K'} \lambda_j I_{k,j} \]

Changing the order of the summation we obtain:

\[ \sum_{j \in K} \sum_{k \in S(R)} \lambda_j I_{k,j} = \sum_{j \in K'} \sum_{k \in S(R)} \lambda_j I_{k,j} \]

\[ \sum_{j \in K} \lambda_j |R| = \sum_{j \in K'} \lambda_j |R| \]

\[ \sum_{j \in K} \lambda_j = \sum_{j \in K'} \lambda_j \]

Consider the two convex combinations of actions of player 2:

- the first combination is the set of columns $j$ in $K$ with the weights $\frac{\lambda_j}{\sum_{j \in K} \lambda_j}$;
- the second combination is the set of columns $j$ in $K'$ with the weights $\frac{\lambda_j}{\sum_{j \in K'} \lambda_j}$.

If in the matrix above those combinations are equal then they have the same payoffs (and observations) for player 1, and hence the same terminal payoffs for player 2.

Consider the following change in the probabilities of the actions of player 2: take away an $\varepsilon$ probability from the actions in $K$, with weights according to the convex combination, and
add this $\varepsilon$ to the actions in $K'$ (with weights according to the convex combination). That is, for each column $j$ in $K$, $P(j)$ is changed to $P(j) - \varepsilon \frac{\lambda_j}{\sum_{j \in K'} \lambda_j}$, and for every column $j \in K'$ $P(j)$ is changed to $P(j) + \varepsilon \frac{\lambda_j}{\sum_{j \in K'} \lambda_j}$. Since the two combinations induce the same payoffs for player 1 within each row in $R$, his payoff when playing the rows of $R$ will remain the same.

In addition, for an $\varepsilon$ small enough, the payoffs for player 1, for all actions in $R$ remain zero, while the payoff for all actions in $A_1 \setminus R$ remain strictly less than zero. On top of that, the terminal payoffs for player 2 remain the same.

Note that before adding $c'$, the profitable deviation, there was no dependency between the columns of the matrix (at the end of each iteration there is no such dependency). So the addition of $c'$ created the dependency. Hence, $c'$ has a positive weight either in $K$ or in $K'$. Assume, without loss of generality, that it has a positive weight in the convex combination of $K'$.

Player 2 is currently indifferent between all the actions in $C$, and strictly prefers $c'$ over them. Putting more weight on $K'$ is strictly better for player 2: it increases current payoff, because it increases the probability of $c'$ without changing the terminal payoffs.

We increase $\varepsilon$ and shift the weights of the distribution from the columns in $K$ to the columns in $K'$ until at least one of the following possibilities takes place:

possibility $a$: the probability of at least one column in $K$ went down to zero. Then we will remove this column from $C$. If this change enables finding appropriate terminal payoffs (i.e. if the resulting updating situations matrix has independent columns), then we are done. If not, we have two new equal combinations of columns, and we need to keep changing the weights with the new pairs of combinations (because $c'$ remained a profitable deviation).

possibility $b$: the payoff for player 1 for at least one of the rows in $A_1 \setminus R$ went up to zero. We add this row to $R$ (it is now a best-response) and check whether with this additional row the right terminal payoffs can be found. Again, if those cannot be found, then again there are two equal combinations and we keep changing the weights until either possibility $a$ or possibility $b$ takes place.

Rows, in this algorithm, are only added to $R$ and never removed. The algorithm could
get into infinite loop only if it can check for some given set \( R \) the same distribution over
the same set \( C \) more than once. For a given set \( R \) the current payoff of player 2 is strictly
increasing as the algorithm advances, so this cannot happen. Hence the algorithm cannot
get into an infinite loop.\(^6\)

This sums up the construction of the algorithm and therefore the proof of lemma 3.1. ■

3.5 From Terminal Payoffs to Continuation Payoffs

In this section we shall describe how the terminal payoffs, used after punishments, can be
obtained as continuation payoffs in the infinitely repeated game. We do that in three steps:
in the first step we obtain an initial two-dimensional set of payoffs, using strategies such that
any profitable deviation from those strategies is detected with a positive probability, and its
detection is common knowledge. In addition, this set would be such that any payoff on the
strictly efficient frontier Pareto dominates a non-empty subset of it. Call this set \( W \). The set
of games where such a \( W \) can be obtained via sequential equilibria is the one we defined to
be \( \Gamma \). In the second step we describe how we use the set \( W \) as continuation payoffs. Finally,
in the third step, we demonstrate how the players communicate the continuation payoffs.

3.5.1 First Step: A Two-Dimensional Set of Payoffs with Observ-
able Deviations, \( W \)

An initial two-dimensional set of payoffs with detectable deviations

We will first obtain some two-dimensional set, call it \( W^- \), and then we will enlarge it so that
any payoff on the strictly efficient frontier will Pareto-dominate a non-empty subset of it.

\(^6\)For the algorithm to be even more efficient, we can add to it the condition that if for a given set of
columns, there is more than one way to minimax the opponent using a full-support distribution over that
set, then the algorithm should begin with considering the distribution which induces the highest current
payoff to player 2. With this improved version, the algorithm will consider a given set of rows paired with a
given set of columns at most once, hence making it finite.
As mentioned in the introduction, the set $\Gamma$ includes, among other games, all games where at least one of the following three conditions holds. We will explain in this section how $W^-$ can be obtained using each one of those conditions:

Condition 1: there is at least one one-period Nash equilibrium whose payoff is internal to the set of feasible payoffs.

If condition 1 holds, then we can easily obtain the convex-hull that Pareto-dominates the one-period Nash equilibrium as payoffs of sequential equilibrium. Any point on the strictly efficient frontier can be obtained as a sequential equilibrium payoff because any profitable deviation is immediately detected, its detection is common-knowledge, and the one-period Nash equilibrium can be the punishment for such a deviation.

As in Fudenberg and Maskin (1991), we can find a sequence of the one-period Nash equilibrium payoff and the payoffs on the strictly efficient frontier such that the continuation payoff is always within an $\epsilon$-distance away from the required payoff, an $\epsilon$ that approaches zero as $\delta$ does to one.

Condition 2: the strictly efficient frontier is piecewise linear but not linear.

If condition 2 holds, then there is the two-dimensional convex-hull of the strictly efficient frontier as a set that can be obtained via strategies such that any profitable deviation is detected (and its detection is a common knowledge). As a punishment for such a deviation we will use our construction of minimaxing, and we will prove, in section 3.5.2, that this punishment is indeed effective.

Condition 3: there exists a pair of distributions, one over the actions of player 1, $p = (p_1, \ldots, p_{|A_1|})$, and one over the actions of player 2, $q = (q_1, \ldots, q_{|A_2|})$, that satisfy all the following conditions:

a. $u(p, q)$ is internal to the set of feasible payoffs.

b. all actions that are played with positive probabilities have the same expected payoff when played against the opponent’s distribution, that is,
\[ \forall a_1, a'_1, \quad p(a_1) > 0, p(a'_1) > 0 \Rightarrow u_1(a_1, q) = u_1(a'_1, q) \]
\[ \forall a_2, a'_2, \quad q(a_2) > 0, q(a'_2) > 0 \Rightarrow u_2(p, a_2) = u_2(p, a'_2) \]

c. playing an action that is not in the support of \( p \) or \( q \), and that has a higher payoff than those that are in the support (when played against the distribution of the opponent) induces a positive probability for both players to observe simultaneously a payoff that cannot be observed if only actions in the support was played. Formally:

\[ \forall a'_1, p(a'_1) = 0, u_1(a'_1, q) > u_1(p, q) \Rightarrow \exists a_2, q(a_2) > 0, \forall a_1, p(a_1) > 0, \ u_1(a'_1, a_2) \neq u_1(a_1, a_2), u_2(a'_1, a_2) \neq u_2(a_1, a_2) \]
\[ \forall a'_2, q(a'_2) = 0, u_2(p, a'_2) > u_2(p, q) \Rightarrow \exists a_1, p(a_1) > 0, \forall a_2, q(a_2) > 0, \ u_1(a_1, a'_2) \neq u_1(a_1, a_2), u_2(a_1, a'_2) \neq u_2(a_1, a_2) \]

If condition 3 holds, then again, we have the required set as the convex-hull of \( u(p, q) \) and the strictly efficient frontier.

Note that we keep insisting that a profitable deviation from strategies that induce payoffs in \( W^- \) will not only be detectable, but also that it’s detection will be common-knowledge. The reason we do that is not because a player cannot let it’s opponent know that a deviation was detected (on contrary, we show how players can communicate in section 3.5.3). The problem is rather that we cannot force the players to communicate truthfully. A player who observes a deviation and knows that the deviator does not know whether his deviation was detected or not, can decide not to report this deviation. In order to give incentives to the player to report detections of deviation, they need to (at least weakly) prefer punishing over not punishing.

A player’s continuation payoff, following his own punishment, is low, so it is easy to give incentives to such a player to report detections of deviations of his opponent.

However, a player who was punishing has a range of continuation payoffs, that we implement using the set \( W \). If such a player has the highest continuation payoff, then in order to give him incentives to prefer punishing again, his set of continuation payoffs following the next punishment, should include payoffs that are even higher than that highest continuation
payoff. That is, we will need a larger set of continuation payoffs. But more deviations can take place, so we will need to constantly increase the set of continuation payoffs. This cannot be done with a given set $W$. 

In case the payoff matrix is one that allows simultaneous transmission of messages between the players, then in some games a "confession and pardon" method, as in the first part of this dissertation, can be employed. If this is the case, then the detections of deviations do not have to be common-knowledge.

Games in which none of the three conditions above hold have some unique features. For example, from any pure action whose payoff is not on the efficient frontier (that is a single line) there is a profitable undetectable deviation, or a deviation whose detection is not common-knowledge. Hence, if we start at such a payoff and follow the “path of deviations that cannot be detected in a common-knowledge fashion ”, we will end up at some point on the efficient frontier. In addition, there is no one-period Nash equilibrium whose payoff is internal to the set of payoffs. In cases where both extreme points of the efficient frontier are one-period Nash equilibria, it is trivial to support the entire efficient frontier as a sequential equilibria payoffs, so we are interested only in games where this is not the case. We use those features to show in section 3.6 how one can obtain an internal payoff point as a sequential equilibrium payoff in many of those cases.

**Enlarging the set $W^-$**

We can enlarge $W^-$ by convexifying $W^-$ with the strictly efficient frontier. We can do that, for example, by playing alternatingly one period a strategy which yields a point in $W^-$ and the next period a strategy whose payoff is on the strictly efficient frontier. Of course, we can alternate every three, four, five periods, etc. We may need the players to have a higher discount factor in order to support some of those alternating strategies.
CHAPTER 3. OBSERVABLE DETERMINISTIC PAYOFFS

3.5.2 Second Step: Using the set $W$

As we saw in the previous section, the observations of the player who is being punished during the punishment, induce changes in the continuation payoffs for the punishing player. For each observation the continuation payoff may change. Let the maximal over all $\Delta_k$’s be $\Delta_{\text{high}}$ and the minimal be $\Delta_{\text{low}}$. Let the highest possible continuation payoff in case no deviation of the punishing player was detected be $Y_{\text{high}}$ and the lowest be $Y_{\text{low}}$. In case a deviation of the punishing player during the punishment was detected, the continuation payoff will be $Y_{\text{punish}}$, that will be even lower than $Y_{\text{low}}$. In addition, as will be explained in the next section, there will be at most three periods during which the continuation payoff is communicated. Let $X_{\text{com}}$ be the payoff during the communication period, (its value will be clear in section 3.5.3) and let $G$ be the highest absolute value of all payoffs that appear in the payoff-matrix.

We will use a subset of $W$ in order to implement the continuation payoffs. Let $(v^*_1, v^*_2)$ be the payoff on the strictly efficient frontier that we wish to support via sequential equilibrium. Let the convex hull of $\{(v_1, v_2), (v'_1, v_2), (v_1, v'_2), (v'_1, v'_2)\}$ be a subset of $W$ such that $v'_i < v_i < v^*_i$ for $i \in \{1, 2\}$.

The different continuation payoffs for player $i$ will be applied by randomizations over $v_i$ and $v'_i$. That is, the expected continuation payoff will be $\alpha v_i + (1 - \alpha)v'_i$ for some $\alpha$ that depends on the observations of the punished player during the punishment. Formally,

\[
Y^i_{\text{high}} = \alpha_h v_i + (1 - \alpha_h)v'_i
\]

\[
Y^i_{\text{low}} = \alpha_l v_i + (1 - \alpha_l)v'_i
\]

\[
Y^i_{\text{punish}} = \alpha_p v_i + (1 - \alpha_p)v'_i
\]

Payoffs between $Y^i_{\text{low}}$ and $Y^i_{\text{high}}$ are implemented using $\alpha$s between $\alpha_l$ and $\alpha_h$.

For reasons detailed in the next sub-section, we wish all $\alpha$s to be bounded away both from 0 and 1. For now, we assume only that $\alpha_p < \alpha_l < \alpha_h$ (which implies $Y_{\text{punish}} < Y_{\text{low}} < Y_{\text{high}}$). We will show that for any such choice of $\alpha$s, if $\delta$ is close enough to 1 then we can find $N$ such that $\delta^N$ satisfies all the conditions that are needed for our description to be an equilibrium.
(conditions 1-4 in the next section).

We will use, for the player being punished, \( v' \) as the continuation payoff, so, for example, in case player 1 was being punished and he detected a deviation of player 2 during that punishment, the continuation payoff will be \( (v'_1, Y_{\text{punish}}^2) \), and if no such deviation was detected then the continuation payoffs will range from \( (v'_1, Y_{\text{low}}^2) \) to \( (v'_1, Y_{\text{high}}^2) \).

(\( v'_1, v'_2 \))
(\( v'_1, Y_{\text{high}}^2 \))
(\( v'_1, Y_{\text{low}}^2 \))
(\( v'_1, Y_{\text{punish}}^2 \))
(\( v'_1, v'_2 \))
(\( Y_{\text{punish}}, v'_2 \))
(\( Y_{\text{low}}, v'_2 \))
(\( Y_{\text{high}}, v'_2 \))

For the description above to be an equilibrium, four conditions should hold:

1. Any profitable deviation from the equilibrium path becomes un-profitable if it is followed by a punishment.

Recall that profitable deviations from the equilibrium path are detected with probability 1. In addition, the equilibrium path is designed so that the continuation payoff is always within an \( \varepsilon \)-distance from the target payoff (an \( \varepsilon \) which decreases as the players are more patient)\(^7\).

For the first condition to hold it is sufficient that:

\[(1 - \delta)2G + (\delta - \delta^{N+1}) 0 + (\delta^{N+4} - \delta^{N+1}) 2G + \delta^{N+4}v'_i < v^e_i - \varepsilon\]

Recall that \( v'_i < v^e_i \), so for \( \delta \) close enough to 1, which, in turn, enables us to choose a small enough \( \varepsilon \) so that

\(^7\)for details see section 3.6
$v_i' < v_i^c - \varepsilon$

the inequality holds.

2. The set of continuation payoffs, $W$, is sufficiently large to allow balancing the payoffs of the punishing player for any possible realization during the minimaxing periods.

For the second condition to hold, it is sufficient that:

$$(ii) \sum_{t=0}^{N-1} \left( \frac{\Delta_{high}}{\delta^t} \right) + \delta^N (1 - \delta^3) G + \delta^{N+3} Y_{low}' \leq \sum_{t=0}^{N-1} \left( \frac{\Delta_{low}}{\delta^t} \right) + \delta^N (1 - \delta^3) (-G) + \delta^{N+3} Y_{high}'$$

put in another way, we need to be able to choose $Y_{high}$ and $Y_{low}$ far enough from each other such that:

$$(ii) \frac{(1 - \delta^N)}{\delta^N} (\Delta_{high} - \Delta_{low}) + \delta^N (1 - \delta^3) 2G \leq \delta^{N+3} (Y_{high}' - Y_{low}')$$

3. Any profitable deviation of the punishing player during the punishment (all of those are detected with some positive probability) becomes un-profitable because of this probability of being detected, a detection that will lead to obtaining the lowest continuation payoff, $Y_{punish}'$

Let $\rho$ be the smallest probability of such a detection, (which is at least the smallest probability in the distribution over the actions of the player being punished).

For the third condition to hold it is sufficient that:

$$(iii) (1 - \delta^N) G + \delta^N (1 - \delta^3) G + \rho \delta^{N+3} Y_{punish}' + (1 - \rho) \delta^{N+3} Y_{low}' < (1 - \delta^N) (-G) + \delta^N (1 - \delta^3) G + \delta^{N+3} Y_{low}'$$

re-arrange:

$$(iii) (1 - \delta^{N+3}) 2G+ < \rho \delta^{N+3} (Y_{low}' - Y_{punish}')$$

4. When the players finish the punishment and play the strategies that induce payoffs in $W$, profitable deviations are, again, detected with positive probabilities (and their detection is common knowledge). Such profitable deviations should become un-profitable thanks to this probability of being detected and punished.

Those probabilities of detecting deviations are bounded from below over all periods and all deviations. Let their infimum be $\nu$

\[\frac{1}{\delta^N} (\delta + \delta^2 + ... + \delta^N) = \frac{1 - \delta^N}{\delta^N(1 - \delta)}\]

\[\text{In fact, in our construction, they are fixed over all periods, so that a minimum exists.}\]
CHAPTER 3. OBSERVABLE DETERMINISTIC PAYOFFS

Let $v_i^c$ be the continuation payoff of player $i$ according to the equilibrium path that has the payoff in $W$ that we wish to support. This continuation payoff will be positive. Denote the expected payoff for player $i$, from the period after the deviation and on, in case the deviation is not detected to be $X$. This $X$ is at least such that:

$$v_i^c = (1 - \delta)(-G) + \delta X$$

it is sufficient for the fourth condition to hold that:

$$(iv) (1 - \delta)G + \nu \left[ (\delta - \delta^{N+1})(1 + (\delta^{N+4} - \delta^{N+3})2G + \delta^{N+4}v_i') + (1 - \nu)\delta \left( \frac{v_i^c - G(1 - \delta)}{\delta} \right) \right] < v_i^c$$

recall that the continuation payoff $v_i^c$ is within an $\epsilon$ distance from the target payoff, which means that $v_i^c > v_i' - \epsilon$. Hence, it is sufficient that:

$$(iv) (1 - \delta) [2G - \nu G] + \nu \delta^N (1 - \delta^3) 2G + \nu \epsilon < \nu (1 - \delta^{N+4}) v_i'$$

which is:

$$(iv) \frac{(1-\delta)[2G - \nu G] + \nu \delta^N (1-\delta^3) 2G + \nu \epsilon}{\nu v_i'} < 1 - \delta^N$$

Conditions 1 and 4 require that the weight of the punishment phase is high enough, so that the low payoff during the punishment phase will cancel any profits from deviations. Condition 2 requires that the weight of the punishment phase is low enough, so that all realizations can be balanced with continuation payoffs. Condition 3 also requires that the punishment phase weight is low enough, because punishments for deviations of the punishing player is by a low continuation payoff inside $W$, and not by yet another punishment phase.

Condition 4 is stronger than condition 1 because $W$ is Pareto-dominated by $v_i^c$, that is, $v_i^c < v_i'$, and because the probability of detection, $\nu$, can be lower than 1.

In order to satisfy conditions 2 and 4 simultaneously, we will need to find $\delta^N$ such that:

$$\frac{(1-\delta)[2G - \nu G] + \nu \delta^N (1-\delta^3) 2G + \nu \epsilon}{\nu v_i'} < 1 - \delta^N \leq \frac{\delta^{N+3} - 2G\delta^N (1 - \delta^3)}{\Delta_{high} - \Delta_{low}} \delta^{N-1}$$

And in order to satisfy conditions 3 and 4 simultaneously, $\delta^N$ will need to satisfy:
\[
\frac{(1-\delta)[2G-\nu G] + \delta^N (1-\delta^3) 2G \nu + \nu \epsilon}{\nu v_i'} < 1 - \delta^N < \frac{\rho \delta^N (Y_{\text{low}} - Y_{\text{punish}}) - \delta^N (1-\delta^3) 2G}{2G}
\]

But for every given \( \rho \) and \( \nu \), the left-hand side of the inequalities goes to zero as \( \delta \) goes to 1, while the right hand sides do not. This means that for \( \delta \) close enough to 1, we can find \( N \) such that \( \delta^N \) satisfies both inequalities simultaneously.

### 3.5.3 Step 3: Communicating the Continuation Payoff

In this section we show how messages regarding the continuation payoffs can be exchanged. Since we assume that no direct communication is allowed, the messages can only be transmitted through the actions of the players. As mentioned above, there are, in fact, only two possible continuation payoffs for player \( i \), \( v_i' \) and \( v_i \). The segment of expected continuation payoffs, between \( Y_{\text{punish}} \) and \( Y_{\text{high}} \) is obtained by randomizations over \( v_i' \) and \( v_i \). The message, then, is a binary one - either the high continuation payoff, \( v_i \), or the low continuation payoff, \( v_i' \).

We will show how player 1 can convey a message to player 2. Player 2 can convey a message to player 1 in a similar manner. We divide the discussion into the following three cases: there is at least one column such that there are two different efficient payoffs on that column; there is at least one row such that there are two different efficient payoffs on that row; and there are two different efficient payoffs in different columns and rows.\(^{10}\)

**case 1: two efficient payoffs in the same column**

In this case, let the column where there are two efficient payoffs be \( c^* \). Let the actions of player 1 which induce, combined with \( c^* \), the first and the second efficient payoff to be \( \hat{a}_1 \) and \( \hat{a}_2 \) respectively. We will instruct player 2 to play column \( c^* \), and player 1 to play either \( \hat{a}_1 \) or \( \hat{a}_2 \). The first action, \( \hat{a}_1 \), will convey the message that the continuation payoff for player

\(^{10}\)obtaining the efficient frontier as sequential equilibrium payoff in the case where there is only one efficient payoff is trivial.
2 is \( v_2 \) and the second action, \( \hat{a}_2 \), that it is \( v'_2 \). Since both actions induce an efficient payoff, none of the players can profit from deviating without being detected.\(^{11}\)

The two messages induce different payoffs also for player 1. We want player 1 to be indifferent between the two messages. Let the efficient payoffs be \((u_1(\hat{a}_1, c^*), u_2(\hat{a}_1, c^*))\) and \((u_1(\hat{a}_2, c^*), u_2(\hat{a}_2, c^*))\). Assume, without loss of generality, that \( u_1(\hat{a}_1, c^*) > u_1(\hat{a}_2, c^*) \). Then the continuation payoff in case player 1 played \( \hat{a}_1 \) will be \( (v'_1 - \delta \left( u_1(\hat{a}_1, c^*) - u_1(\hat{a}_2, c^*) \right), v_2) \) and in case he played \( \hat{a}_2 \) it is \( (v'_1, v'_2) \).

**case 2: two efficient payoffs in the same row**

In this case, transmitting a message will have three stages: first player 1 will communicate a message, then player 2 will “echo” back the message to player 1, and then player 1 will confirm that the echo was accurate.

Let the set of columns where the highest payoff of player 2 appears be \( C_{max} \). We shall look, within the payoffs of the columns \( C_{max} \) for the column with the highest payoff of player 2 that is not that maximal payoff (the “second best” within those columns). There can be more than one column within \( C_{max} \) with that ”second-best” payoff. Pick one and denote it \( c'' \). Note that at least one payoff that is different from the maximal payoff appears at each column in the row that minimaxes player 2, so such a “second best” payoff must exist. In addition, player 1 may have more than one action which gives player 2, against \( c'' \), player 2’s highest payoff, and his ”second-best” payoff. We will allow player 1 to choose the actions which give him (player 1) the highest payoffs among those actions. Let the row that gives player 1 the highest payoff against \( c'' \) while inducing the highest payoff to player 2, and the one that gives player 1 the highest payoff against \( c'' \) while inducing that ”second best” payoff to player 2, be \( r' \) and \( r'' \) respectively.

To understand better, consider example 3.1. The set \( C_{max} \) is \( \{b, c, e, f\} \). Among them, the ”second-best” payoff, payoff 2, appears in columns \( b, c \) and \( f \). Pick \( f \) to be \( c'' \), and then

\(^{11}\)even if the belief of player 2 is that with a probability 1 one of the messages will be transmitted, he still would have no incentive to deviate.
CHAPTER 3. OBSERVABLE DETERMINISTIC PAYOFFS

$r'$ is $A$ and $r''$ is $B$.

We instruct player 1 to play $r'$ when he wants to convey the message that player 2’s continuation payoff is $v_2$, and $r''$ when the message is that the continuation payoff is $v_2'$.

In the following period, player 1 should play the row in which there are two efficient payoffs, let that row be $\hat{r}$, and player 2 should play the column which induces the highest payoff for himself of the two efficient payoffs in case message he received was that the continuation payoff is $v_2$, and the action that induces the lower self-payoff in case it is $v_2'$.

At the first period of the communication, we want player 1 to play either $r'$ or $r''$ and player 2 to play $c''$. Player 2, by this, will obtain either his highest payoff or a payoff that is the second-best in the columns of $C_{max}$, respectively. Note that there could be a payoff higher than $u_2(r'', c'')$ in columns that are not in $C_{max}$, so profitable deviations of player 2 may exist. To avoid them, we need the probability of $r'$ to be high enough. The lowest probability of $r'$ is $\alpha_p$. Thus, We should keep $\alpha_p$ high enough such that the following hold:

$$\forall c \in A_2 \text{ such that } u_2(r'', c) < u_2(r', c),$$

$$\alpha_p u_2(r', c'') + (1 - \alpha_p)u_2(r'', c'') > \alpha_p u_2(r', c) + (1 - \alpha_p)u_2(r'', c)$$

Let that minimal $\alpha_p$ needed be $\alpha_p^*$. In example 3.1, $u_2(r'', a) < u_2(r', a)$ but both payoffs are lower than those player 2 obtains when playing $f$ against $r'$ and $r''$, so the condition holds trivially for and $\alpha_p$, so $\alpha_p^* = 0$.

Player 2 can still profitably deviate, but only to columns that induce the highest payoff both against $r'$ and $r''$. This means that such a deviation will result in loosing information, a loss that could be revealed at the second stage, when player 2 is required to “echo” the message. In order to keep him from deviating and loosing that information, his belief regarding his continuation payoff should not be “too firm”, that is, he shouldn’t believe given any of his histories that the probability to observe one of the messages is “too close” to 1. This is because if he does believe that the message is that the continuation payoff is, for example, $v_2$, then he may deviate during the period when the message is being conveyed to an action that does not allow him to differentiate between the two messages, and then he will “echo” the message ”$v_2$", being almost certain that he is correct. In order to avoid such
a scenario we will keep the probabilities of the two continuation payoffs bounded away from 1.

It is sufficient that \( \alpha_h \) will be low enough such that:

\[
(v1) \quad (1 - \delta^2)2G + (1 - \alpha_h) \left[ (\delta^{N+3} - \delta^N) G + \delta^{N+3}v_2' \right] + \delta^2 \alpha_h v_2 < \delta^2 \left[ (1 - \alpha_h)v_2' + \alpha_h v_2 \right]
\]

that is,

\[
(v1) \quad \frac{(1 - \delta^2)2G + (1 - \delta^3)(1 - \alpha_h)G}{\delta^2(1 - \alpha_h)v_2'} < 1 - \delta^{N+1}
\]

For similar reasons, we need \( \alpha_p \) to be low enough so that:

\[
(v2) \quad (1 - \delta^2)2G < \alpha_p \left[ (\delta^{N+3} - \delta^N) (-G) + \delta^2 (1 - \delta^{N+1}) v_2' \right]
\]

which is:

\[
(v2) \quad \frac{(1 - \delta^2)2G + \alpha_p (-G)(1 - \delta^3)}{\alpha_p \delta^2 v_2'} < 1 - \delta^{N+1}
\]

For any \( \alpha_h \) and \( \alpha_p \), as \( \delta \) approaches 1, the left hand side of both inequalities approaches zero, so it will be possible to find \( N \) that satisfies \((v1)\) and \((v2)\) together with inequalities \((i)-(iv)\), for \( \delta \) close enough to 1.

If player 2 deviated and chose during the first period of the communication a strategy which did not allow him to differentiate between the two actions of player 1 (he “did not listen”), then only player 1 knows whether the message that is echoed back was the correct one. If it is the wrong message, then player 1 should punish player 2. Player 2 does not know whether his deviation was detected, and hence he does not know whether he should be punished. We cannot instruct player 1 to simply begin the punishment, because if player 2 does not know that he is being punished, then he will not necessarily play the set of best-responses, and he will not be able to design the continuation payoffs for player 1. Therefore, player 1 does not have the proper incentives to randomize during the punishment.

Hence, our problem is that player 1 knows that player 2 ”did not listen”, and he prefer to punish player 2 (because current future payoff of player 1 is \( v_1' \), and punishing and then obtaining \( v_1 \) is better), but player 1 is not sure that player 2 knows that he echoed the wrong message and therefore was ”caught not listening”.
In the following period, we will instruct the players to use strategies such that player 2 obtains his highest payoff on the strictly efficient frontier if the communication went well. Player 2 will conform to that strategy both because it gives him his highest payoff and because he can hope that that he got the message right, and he can continue playing as if he did not deviate. Player 1 will conform to this strategy if indeed the message was correct, but if the echo was mistaken, then he is instructed to play some action that gives him (player 1) his highest payoff while inducing a payoff for player 2 that is different from the one he is expecting (player 2’s highest). Player 1 will do it because he prefers to punish\textsuperscript{12}.

**case 3: there are two efficient payoffs in different rows and columns**

In this case, player 1 can convey the message to player 2 in a similar way to the one described in the above subsection. The problem is to find a way for player 2 to echo the message back, without any profitable deviations for any of the players. The “echoing” will consist of an order over two action profiles which induce the two efficient outcomes, that is, if the message is that the continuation payoff is $v_2$, then there is one order over the two actions with efficient payoffs, and if the message is that the continuation payoff is $v'_2$ then the order is reversed.

Again, we will keep the probabilities of the two continuation payoffs bounded away from 1. Consider, within the payoff-matrix, the $2 \times 2$ sub-matrix which includes both efficient payoffs. Player 1, during the echoing periods, can detect a deviation of player 2 with a positive probability in all payoff matrices, except the one where the efficient payoffs are organized as in Example 3.2. The notation $(x, e(x))$ represents a pair of payoffs that is efficient, and $(x, e(x)^-) \) represents a pair of payoffs that is not efficient:

---

\textsuperscript{12}Player 1 will not try and claim that the echo was mistaken when in fact it was correct. This is because if he obtained the right echo, then he believes with probability 1 that player 2 indeed listened. So if player 1 tries to wrongfully report a mistaken echo, then both players know that player 1 is the real deviator.
We will show how player 1 can convey a message to player 2 when the sub-matrix in example 3.2 appears. In this specific case, the communication will consist of four periods, but it is easy to see that all inequalities can still be satisfied with four periods of communication.

Assume, without loss of generality, that \( \alpha > \beta \), hence \( e(\alpha) < e(\beta) \) and \( e(\alpha)^- < e(\beta) \). If player 2 will play \( R \), then player 1 will be able to convey a message by playing either \( T \) or \( B \). If the probability with which player 1 will play \( B \) is kept high enough (regardless of history) then profitable undetectable deviations of player 2 can be only to an action that gives him a payoff of \( e(\beta) \) against \( B \), and a payoff higher than \( e(\alpha)^- \), say \( e(\alpha)^- + \), paired with a payoff of \( \alpha \) against action \( T \).\(^{13}\)

We can allow player 2 to play that deviation instead of \( R \).

We can instruct player 2 to play that deviation, and instruct player 1 to play the action \( B \), or the action that gives player player 1 the highest payoff among all actions which give player 2 the payoff \( e(\alpha)^- + \). We can iterate such a sequence of changes in the instructions until there are no profitable undetectable deviations. We need to verify two things: first, that these iterations will eventually end, and second, that when it does end, player 2 will be able to differentiate between the actions of player 1.

Note that when player 2 deviates during the iterations above, he increases his own payoffs while keeping the payoffs of player 1 the same as before. The same goes for when we allow player 1 to choose the actions that give himself the highest payoffs while keeping the payoffs of player 2 the same. Hence, each iteration leads to a payoff that is the same as before for player 2 cannot improve his payoff against action \( B \) without being detected. An improvement that is not detectable will have to be such that it gives a higher payoff against \( T \). If the probability of \( B \) is high enough, then for a deviation to be profitable it has to first not decrease the payoff against \( B \), and only second to improve against \( T \).
one player and strictly higher for the opponent. The payoff matrix is finite so the iterations have to end.

In addition, since the payoff of player 1 cannot decrease, when the iterations stop, it has to be at least $\alpha$ for the actions that do not induce the payoffs $(\beta, e(\beta))$. And since the payoff of player 2 is paired with a payoff that is at least $\alpha$, it cannot exceed $e(\alpha)$ which is smaller than $e(\beta)$. Therefore, player 2 can differentiate between the actions.

Again, if player 2 deviates during the first period of the communication (he “does not listen”) and then he tries to echo back one of the messages, hoping that he will be correct, then it is only player 1 who knows that player 2 was wrong. As in case 2, the following period will consist of player 2 obtaining his highest payoff, unless he is to be punished.

3.5.4 Off Equilibrium Path Beliefs

Any profitable deviation from the equilibrium path is immediately detected with probability 1. So as long as such a deviation does not take place, the players believe that the opponent is conforming to the equilibrium path.

Once a deviation does take place, it is a common-knowledge that a punishment phase should begin.

During the punishment phase, the player being punished is playing only best responses, and the punishing player is indifferent between all the actions he is instructed to play. Any profitable deviation from those actions is detectable with a positive probability. We don’t give any specific instructions as to how should the players play from the detection of a deviation (from the punishment) until the communication phase. We are not interested in the payoffs that the players obtain during those periods, since the reduced continuation payoffs that results from the first detected deviation during the punishment phase makes such a deviation un-profitable regardless of the payoffs obtained until the end of the punishment phase. It is sufficient to know that for this limited amount of periods, there exists a sequential equilibrium, (Kreps and Wilson (1982)), and that it is a best response to all players, regarding
of what they believe that happened during those periods, to act according to the instructions of the communication phase once the punishment phase is over.

3.6 Obtaining an Internal Point in Other Matrices

3.6.1 A Simple Randomization with Continuation Payoffs

Consider the following matrix

\[
\begin{array}{c|ccc}
  & L & C & R \\
  \hline
  T & -1, -1 & -1, 0 & -1, 1 \\
  M & 0, -1 & -2, 0 & 0, 6 \\
  B & 1, -1 & 0, 6 & 6, 1 \\
\end{array}
\]

Example 3.3 has only one one-period Nash-equilibrium, \((B, C)\), with the payoff \((0, 6)\), it’s strictly efficient frontier is the line connecting \((0, 6)\) and \((6, 1)\), and from each common action which induces a payoff that is not on the efficient frontier there is an un-detectable profitable deviation (for example, from the common action \((T, L)\) there is a profitable un-detectable deviation for both players - for player 1 to either \(M\) or \(B\), and for player 2 from \(L\) to either \(C\) or \(R\)). As for other sub-matrices, row \(B\) strictly dominates \(T\) and \(M\), so there is no distribution over the columns that can make both players indifferent between their actions.

However, there is a way to obtain a point that is not on the strictly efficient frontier as a sequential equilibrium payoff. We will instruct player 2 to play \(C\), and player 1 to randomize between \(T\) and \(B\), at the probability at least \(\frac{1}{6}\) to \(B\), all this for \(K\) periods. The payoffs for player 1 when he plays actions \(M\) and \(B\) are different. In order to make player 1 indifferent between playing those actions, they need to have different continuation payoffs.

The different continuation payoffs will take place after those \(K\) periods, and will be on the strictly efficient frontier. Since player 2, when playing \(C\), can tell whether player 1 is
playing $T$ or $B$, both players know what should the continuation payoffs that will make player 1 indifferent.

In order to obtain continuation payoffs that are within an $\epsilon$ distance from the intended payoff, we can iterate a similar construction of $K$ periods of randomizations followed by payoffs in the efficient frontier. Small adjustment in the payoffs on the efficient frontier might be needed in order to be accurate with the continuation payoff of each iteration.

### 3.6.2 A Complicated Randomization with Continuation Payoffs

Consider the following matrix

**Example 3.4:**

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>C</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>-1,-1</td>
<td>-1,0</td>
<td>-1,0</td>
</tr>
<tr>
<td>$M$</td>
<td>0,-1</td>
<td>6,0</td>
<td>0,6</td>
</tr>
<tr>
<td>$B$</td>
<td>0,-1</td>
<td>0,6</td>
<td>6,0</td>
</tr>
</tbody>
</table>

In this example as well, the only one-period Nash-equilibrium is $(3, 3)$ on the efficient frontier. In addition, there is a profitable un-detectable deviation from every sub-matrix that contains an internal point (from a distribution such that the players are indifferent between the actions in that sub-matrix). However, in this matrix we cannot employ the same method as in example 3.2. This is because from every observable randomization over two actions for one player and one action of the opponent that induces a positive probability for a payoff not on the efficient frontier, there is a profitable undetectable deviation. For example, if we instruct player 2 to play $C$, and player 1 to randomize between $T$ and $B$, then player 1 can deviate from $T$ to $B$. 
There is still a way to obtain a payoff not on the efficient frontier in this matrix. We will use randomization and continuation payoffs on the efficient frontier, as in example 3.2, only the randomization and the construction of the continuation payoffs will be slightly more complicated. We will ask player 1 to randomize between all his actions \((T, M \text{ and } B)\) and player 2 to randomize between \(C\) and \(R\). The players should have common knowledge regarding the continuation payoffs, so the different continuation payoffs are designed by the following common-knowledge equivalence-classes division:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>(x)</td>
<td>(x)</td>
</tr>
<tr>
<td>M</td>
<td>(x)</td>
<td>(y)</td>
</tr>
<tr>
<td>B</td>
<td>(z)</td>
<td>(x)</td>
</tr>
</tbody>
</table>

We will show that proper continuation payoffs exist.

Let the probability with which player 1 plays \(T, M\) and \(B\) to be \(p_1, p_2\) and \(1 - p_1 - p_2\) respectively, and the probability with which player 2 plays \(C\) and \(R\) to be \(q\) and \(1 - q\) respectively.

Let the reference continuation payoff, after \(K\) periods be \((\beta, 6 - \beta)\). This payoff in on the efficient frontier. The players will adjust this payoff after each period, according to the equivalence class of the cell that was played. Let the change in player 1’s payoff given that a cell marked by \(x\) is played to be \(x_1\). The corresponding change in player 2’s payoff is \(x_2 = -x_1\). In the same fashion, let the changes in the players’ payoffs given that the cells marked \(y\) and \(z\) are played to be \((y_1, y_2 = 6 - y_1)\) and \((z_1, z_2 = 6 - z_1)\) respectively.

In order to make player 1 indifferent between playing his three actions the following two inequalities have to hold:

\[
(i) \quad (1 - \delta)(-1) + \delta^M x_1 = (1 - \delta)q6 + \delta^M [qx_1 + (1 - q)y_1]
\]

or:

\[
(1 - \delta)(-1) + \delta^M x_1 = (1 - \delta)q6 + \delta^M [qx_1 + (1 - q)y_1]
\]
(i) \((1 - q)\delta^M (x_1 - y_1) = (1 - \delta)(1 + q6)\)

and:

(ii) \((1 - \delta)(-1) + \delta^M x_1 = (1 - \delta)(1 - q)6 + \delta^M [qz_1 + (1 - q)x_1]\)

or:

(ii) \(q\delta^M (x_1 - z_1) = (1 - \delta) [1 + (1 - q)6]\)

In order to make player 2 indifferent between his two actions the following equality has to hold:

(iii) \((1 - \delta)(1 - p_1 - p_2)6 + \delta^M [(p_1 + p_2)x_2 + (1 - p_1 - p_2)z_2] =
(1 - \delta)p_26 + \delta^M [p_2y_2 + (1 - p_2)x_2]\)

or:

(iii) \((1 - \delta)(1 - p_1 - 2p_2)6 =
\delta^M [p_2y_2 + (1 - p_1 - 2p_2)x_2 - (1 - p_1 - p_2)z_2]\)

Let \(\Delta^1_{xy}\) be \(x_1 - y_1\), \(\Delta^1_{xz}\) be \(x_1 - z_1\) and \(\Delta^1_{yx}\) be \(y_1 - x_1\) and we use similar notations for player 2.

put differently:

(i) \(\Delta^1_{xy} = \frac{(1 - \delta) [1 + q6]}{\delta^M} \frac{1}{1 - q}\)

(ii) \(\Delta^1_{xz} = \frac{(1 - \delta) [1 + (1 - q)6]}{\delta^M} \frac{q}{q}\)

(iii) \(\frac{(1 - \delta)}{\delta^M} (1 - p_1 - 2p_2)6 = p_2\Delta^2_{yx} + (1 - p_1 - p_2)\Delta^2_{xz}\)

Now, given that \(\Delta^2_{yx} = -\Delta^1_{yx}\) and \(\Delta^2_{xz} = -\Delta^1_{xz}\), we obtain:

(iii) \((1 - p_1 - 2p_2)6 = p_2 \frac{[1 + q6]}{1 - q} + (1 - p_1 - p_2) \left[-\frac{1 + (1 - q)6}{q}\right]\)

For any given \(p_1\) and \(p_2\), when we take \(q\) to 1, the right hand side approaches infinity, while when we take it to zero, it approaches minus infinity, so we can always find a value of \(q\) that will make equality (iii) hold. Equalities (i) and (ii) can be solved afterwards for value of \(q\).
Chapter 4

Repeated Two-Bidders Auction

4.1 Introduction

Collusive behavior among bidders is a well recognized problem in auctions. One branch of the theoretical literature tries to analyze the extent to which a group of bidders can successfully collude. Most of the early theoretical literature on collusion focuses on single period auctions (Graham and Marshall (1987), McAfee and McMillan (1992)). In their important paper McAfee and McMillan show that extent to which a group of bidders can successfully collude is crucially related to the possibility of ex-post transfers. A collusion is successful if it can maximize the surplus. Surplus is maximized if the bidder with the highest valuation gets the object and pays the reserve price. In a model where a bidder’s valuation is private information, the identification of the bidder with the highest valuation is an issue. McAfee and McMillan show that if the cartel members can engage in side payments, a mechanism such as running an internal auction and then bidding the reserve price can effectively sustain perfect collusion i.e., maximize the surplus (in the literature this is called efficient collusion). However in the absence of side transfers the efficiency of a cartel is severely limited.

One limitation of the analysis discussed above is that the model is one of static collusion (that is, a one-shot auction framework). However, if collusion is a result of repeated inter-
actions, a more appropriate framework is provided by the theory of repeated games under imperfect monitoring. The analysis of collusion in auctions along these lines is relatively recent. Aoyagi (2002) considers a model of repeated auction where the agents communicate their bids at the end of each round. Aoyagi develops a dynamic bid-rotation scheme (as against the static bid-rotation scheme of M and M) and shows that even without side payment of money, it is possible to improve cartel efficiency over the static bid-rotation scheme, through intertemporal payoff transfers. Aoyagi considers explicit communication among bidders. In many real life auctions however, cartel formation among bidders is illegal and explicit communication is not allowed. In an important paper Skrzypacz and Hopenhayn (2004) prove the existence of a collusion scheme without communication that performs strictly better than the static scheme of McAfee and McMillan. They construct collusion schemes that have asymmetric continuation equilibria, and through these asymmetric equilibria they are able to generate implicit transfer schemes. Skrzypacz and Hopenhayn consider strategies of the players that depend on the history of publicly observable signals (in their case the history of wins). The paper constructs a Public Perfect Equilibrium (PPE) of the repeated game in which the cartel extracts a surplus that is strictly greater than the one obtainable under bid-rotation scheme. However, the equilibrium surplus is bounded away from the fully efficient surplus.

A recent paper Horner and Jamison (2007) also considers the possibility of collusion when there is almost no information transmission.

The studies so far analyzes collusion by focusing on the PPE of the resulting repeated game. In a PPE, the players are restricted to use public strategies i.e., strategies that depend only on the history of the publicly observable signals. In this chapter, we take the analysis one step further by considering private equilibria of the repeated game. In a private equilibrium, the (private) strategies of a player may depend on the history of his past actions which is private information, in addition to the history of the publicly observable signals. The interest in private equilibria in repeated games whose monitoring is public is fairly recent. Mailath et. al. (2002) give three examples in the case of finitely repeated games where
the equilibria in private strategies Pareto dominates every PPE. In the context of infinitely repeated games Kandori and Obara (2003) extends this analysis and illustrates cases where players make better use of information by using private strategies. Both the papers consider games with finite action spaces. In this paper we study collusion in auction games where the action spaces for the agents are intervals in the real line. Most of the literature on auctions deals with models where action spaces are intervals on the real line. In such a case, even with communication, under imperfect public monitoring, the best achievable pay-off configuration is bounded away from the efficient pay-off (Aoyagi, 2003). Moreover, most of the literature on repeated games deals with sequential equilibrium. However, sequential equilibrium is only defined for games with finite action spaces.

In this chapter we consider a model of first price auctions where there are two bidders with Independent Private Valuations (IPV). The two bidders are engaged in repeated auction. Their private valuations are randomly and independently drawn before each round according to a common distribution over \([0, 1] \times [0, 1]\). At the end of each round the auctioneer announces the identity of the winner in that round. Although we are unable to construct an exact equilibrium that induces efficient outcome, we show that there exists an \(\epsilon\)-equilibrium, where the outcome is close in an appropriate sense to the efficient outcome.

Specifically we show that, if players are sufficiently patient, then using private strategies it is possible to construct an \(\epsilon\)-equilibrium in which the payoff to the bidders is arbitrarily close to what they would get if the object is allocated to the highest valuation bidder at the reserve price in every period. In their celebrated paper Fudenberg, Levine and Maskin (1994) show that it is possible to obtain full efficiency under imperfect public monitoring as long as the space of public signals is rich enough relative to the action spaces of the agents. Aoyagi (2003) obtains a similar result for auctions when the agents have finite action spaces. Note that, in our model the action spaces of the bidders are intervals on the real line. Moreover, the only public signals available at the end of each period are the identities of the winner. As in Skrzypacz and Hopenhayn we show that with such an uninformative public signal the cartel’s surplus is strictly bounded away from the efficiency surplus.
The specific cartel mechanism we consider consists of three phases: the cooperation phase, the communication phase and the punishment phase. The game starts in the cooperation phase where both the bidders bid an amount that is proportional to their valuation for that period. Choosing a proportional factor close to zero makes the bids close to zero (the normalized reserve price in our model). The cooperation phase lasts for \(N\) periods (\(N\) predetermined) at the end of which the bidders conduct a statistical test. The statistical test is to check whether the other bidder has played according to the prescribed strategy. After the statistical test the game moves to the communication phase where the bidders communicate the results of the test. In manner that will be made clear in the body of the chapter, the communication takes place through bids, and hence makes use of the available public signals only. No other external means of communication is available. If both players communicate that the opponent has passed the statistical test the game returns to the cooperation phase. Otherwise the game moves to the punishment phase. The punishment phase has two stages — the punishment for player 1 (if he has failed the test) and the punishment of player 2 (if he has failed the test). Observe that the transition from cooperation phase to the punishment phase depends on statistical test, the result of which is private information. Each player knows that by mis-reporting the result of the statistical test he can induce or prevent a punishment for the other player and can favorably alter his overall payoff. The challenge is to ensure that each bidder has incentive to report the result of his statistical test truthfully. The trick is to choose a mixed action for the punished player so as to make the punishing player indifferent between punishing and not punishing. Similar ideas appear in Piccione (2002), Ely and Valimaki (2002) in the context of repeated games with private monitoring and in Kandori and Obara (2003) in the context of repeated games with imperfect public monitoring and private strategies. The details of the construction follow in the body of the chapter.

The repeated game that we consider is one with discounting. Every specific discount factor \(\delta\) entails a repeated game. By changing the discount factor we get a sequence of such repeated games. For each such game we construct the epsilon equilibrium and show that as
CHAPTER 4. REPEATED TWO-BIDDERS AUCTION

the discount factor tends to 1 (i.e., players become more patient), the $\epsilon$-equilibrium outcomes tend to the efficient outcome.

As mentioned before we are unable to give a precise equilibrium argument. The reason is that even though we consider the augmented set of private histories, we still cannot account for all possible deviations. If one has to take account of all possible deviations, it comes at the cost of efficiency. What we show is that such deviations are profitable only up to $\epsilon$. In a sense made precise later, our solution is more than an $\epsilon$-equilibrium. It is an $\epsilon$-consistent equilibrium (Lehrer and Sorin (1998), which is equivalent to contemporaneous perfect $\epsilon$-equilibrium of Mailath et al. (2003)).

The chapter, for its major part, focuses on Independent Private Valuations model with uniform priors. The results hold for independent priors that are not uniform as long as the collusive payoff strictly Pareto dominates the one period Nash equilibrium payoff. Models of correlated priors are discussed in section 4.6.

4.2 Preliminaries

The set of bidders is $N = \{1, 2\}$. The bidders are symmetric and risk-neutral. A single indivisible object is sold in every period through a fixed auction format. In our model, it is a first-price auction. At the beginning of each period nature draws a valuation for each player that is independent of the past draws and each component is independently distributed of the other. The vector of private valuations is denoted by $v = (v_1, v_2)$. We assume that each player’s valuation $v_i$ is independently and uniformly distributed over the interval $[0, 1]$.

The participation in the auction is voluntary, so in any period the set of each bidder’s generalized bids is given by the set $B = \{\emptyset\} \cup \mathbb{R}_+$, where $\emptyset$ represents no-participation.

4.2.1 The Stage Game

In every period the object is sold through a first-price sealed bid auction (FPA). We normalize the reserve price to zero. The auction mechanism is described by the measurable mappings
\( p_i \) and \( t_i (i = 1, 2) \) on the set \( B^2 \) of bid profiles \( b = (b_1, b_2) \) : \( p_i(b) \) is the probability that bidder \( i \) is awarded the good and \( t_i \) is his expected payment. Each agent has a strategy \( \phi : [0, 1] \rightarrow B \). The stage-game expected payoff of agent \( i \) is,

\[
    r_i(\phi_1, \phi_2) = E(p_i(\phi_i, \phi_j)v_i - t_i(\phi_i, \phi_j))
\]

The functions \( p_i \) and \( t_i \) are symmetric and satisfy the following conditions:

**Assumption 4.1:**

i. A bidder makes no payment when he chooses not to participate i.e., \( t_i(b) = 0 \) if \( b_i = 0 \).

ii. If only one bidder participates and bids zero, then he wins the object at price zero.

A second assumption we make is that after each period only the identity of the winner is publicly announced (or, of course, that this information can costlessly be discovered).

Let \( r^0 \) be the (ex-ante) symmetric Bayesian Nash Equilibrium payoff to each bidder in the stage auction. Also let \( r^* \) be the expected payoff to each bidder under truthful information sharing and efficient allocation with bidder \( i \) winning the object and paying price 0 if and only if \( v_i > v_j \). In other words,

\[
    r^* = E[1_{\{v_i > v_j\}}(v_i)]
\]

### 4.2.2 The Repeated Game

The repeated game \( G^\delta \) is the repetition of the stage game \( G \). Given a stage \( t \in \{1, 2, \ldots\} \) let \( h_t \) denote the history of the game up to stage \( t \). We denote by \( \mathcal{H}_t \) the set of all such \( t \)-stage histories. We would like to distinguish between two kinds of histories. The *public* history of a game consists of the data up to stage \( t \) that is publicly observable. Formally let \( h^p_t \) denote the public history of a game up to stage \( t \). In our game, the public history consists of the sequence of winners up to the current period. Let \( \mathcal{H}^p_t \) be the set of all such histories. For any player \( i \), a *private* history up to stage \( t \), on the other hand, includes the private information that player \( i \) may have up to stage \( t \), in addition to the publicly observable data. For example, in the game we consider here, a \( t \)-stage private history for
player $i$ records for each of the first $t - 1$ periods, bidder $i$’s valuation $v_{i,\tau}$, his bids $b_{i,\tau}$ as well as the data on the publicly observable signals in period $\tau$. Let $h_{i,t}$ denote the private history of player $i$ up to stage $t$ and $\mathcal{H}_{i,t}$ be the set of all such private histories. A *private* behavioral strategy for player $i$ in period $t$ is a function $\sigma_{i,t} : \mathcal{H}_{i,t} \rightarrow \Delta(B)$ that maps $t$-stage private histories into probability distributions over set $B$. A *private* strategy for a player is a collection $\sigma_i = (\sigma_{i,t})_{t=1}^{\infty}$. Given that $\sigma = (\sigma_i, \sigma_j)$ is the strategy profile, the payoff to player $i$ in the repeated game $G^S$ is,

$$\pi_i(\sigma_i, \sigma_j) = (1 - \delta) \sum_{t=1}^{\infty} \delta^t r_i(\sigma_{i,t}, \sigma_{j,t})$$

### 4.2.3 The Equilibrium Scheme

In this section we set out the collusive scheme. We are considering a scenario where the bidders’ valuations are uniformly and independently distributed. In this case, the expected payoff of each bidder in the one period symmetric Nash equilibrium (NE) is $r^0 = \frac{1}{6}$. The most efficient collusive payoff is $r^* = \frac{1}{3}$. The collusive mechanism that we consider has four phases: the cooperation phase $C$, the statistical test phase $S$, the communication phase $\bar{C}$ and lastly the punishment phases $P_1$ and $P_2$.

**Cooperation Phase:** Play begins in the cooperation phase which lasts for $N$ periods ($N$ pre-specified). In each period $\tau \in \{1, \ldots, N\}$, the players are advised to bid $\alpha v_{i,\tau}$, $\alpha > 0$. In other words, the advised strategy $\sigma^*$ has the following feature: for all $\tau \in \{1, \ldots, N\}$ and for all $h_{i,\tau} \in \mathcal{H}_{i,\tau}$ it is the case that,

$$\sigma^*_i(h_{i,\tau}) = \alpha v_{i,\tau}$$

Observe that if the bidders stick to the advised strategy, the highest valuation bidder will be the winner in each of the $N$ periods. Moreover, by fixing $\alpha$ close to zero, the payment to the auctioneer will be close to zero — the reserve price in our model. At the end of phase $C$ the game moves to phase $S$ where the players conduct a statistical test.
Statistical Test Phase $S$: In this phase each player conducts a statistical test to verify whether the other player has adhered to the advised strategy in phase $C$. The critical part here is the construction of the test function. Suppose that player $i$ is conducting the test on $j$. Fix a $K \in \mathbb{N}$, $K < \infty$ and divide the interval $[0, 1]$ into $K$ sub-intervals $\{[x_k, x_{k+1}]\}_{k=0}^{K-1}$ where $x_k = \frac{k}{K}$. Define $p_i(k, k+1, \sigma_i, \sigma_j)$ to be the number of times player $i$ won when he bid in the interval $[\frac{ak}{K}, \frac{a(k+1)}{K}]$ given that the players $i$ and $j$ are using strategies $\sigma_i$ and $\sigma_j$ respectively. Similarly let, $n_i(k, k+1, \sigma_i)$ to be the number of times player $i$ bid in the interval $[\frac{ak}{K}, \frac{a(k+1)}{K}]$ while using the strategy $\sigma_i$. Then the test function that player $i$ employs to test whether player $j$ has conformed to the advised strategy is defined as,

$$t_{ij}(\sigma_i, \sigma_j) = \max_{k \in K} \left[ \left( \frac{k+1/2}{K} - \frac{p_i(k, k+1, \sigma_i, \sigma_j)}{n_i(k, k+1, \sigma_i)} \right) \frac{n_i(k, k+1, \sigma_i)}{N} \right]$$

The statistical test measures a (weighted) distance between the proportion of times a player won when he bid in a certain segment and the theoretical proportion of times he should win, assuming that the opponent bids according to $\sigma_j$. A player will fail the test if this distance is too large, i.e., if $t_{ij}$ is above some thresh-hold. We will set this thresh-hold be $T = \frac{1}{K^3}$.

The Communication Phase: After the statistical test the game moves to the communication phase where the players communicate the results of the statistical test to their opponents. The communication phase is for two periods $C_{ij}^1$ and $C_{ij}^2$. In period $C_{ij}^i$ player $j$ communicates to player $i$ whether the latter has passed the statistical test or not. Player $i$’s advised action in the period $C_{ij}^i$ is,

$$\sigma_i^*() = \emptyset$$

For player $j$, the advised strategy in period $C_{ij}^i$ is,

$$\sigma_j^*(.) = \begin{cases} \emptyset, & \text{if agent } i \text{ has failed } j'\text{'s test;} \\ 0, & \text{otherwise.} \end{cases}$$
However, player $i$ can deviate and bid any $b \in [0, 1]$ and win the round. Thus there are three possible publicly observable signals:

- $s_i = \text{player } i \text{ won the round}$
- $s_j = \text{player } j \text{ won the round}$
- $s_0 = \text{no body won the round}$.

If the signal is $s_0$ player $i$’s punishment phase follows after the communication phase for player $j$. If the signal is $s_j$, player $i$ has passed player $j$’s test. The signal $s_i$ denotes a deviation by player $i$ from the advised strategy and leads to the one-period NE forever.

In period $C^j_i$ the roles of the two players are reversed.

**Punishment Phases:** If both agents have passed the statistical test, the play returns to the cooperation phase. If only one player, say player 1, has failed the statistical test, then punishment phase $P_1$ ensues that last for $M$ periods. If both players have failed their statistical tests, then two punishment phases, $P_1$ and $P_2$, take place. The strategy of the punished player is a mixed strategy. During, for example, the punishment of player 1, in the first period of the punishment phase, player 1 (the punished player) conveys a message to the punishing player, player 2, through the sum he bids. If player 1 lets player 2 win the first bid of the punishment phase, then player 2 will be allowed to win all bids in the remaining periods of the punishment phase (and player 1 will not participate). If player 1 wins the bid, then both players will not participate in all those remaining periods. Which message to convey is decided by a randomization conducted by the player 1, with probability $p$ to allow the opponent to win the remaining periods.

The strategies of the two players are defined below. The strategy of player 1 is,

$$
\sigma^1(.) = \begin{cases} 
0, & \text{with probability } p \text{ in the first period of punishment;} \\
1, & \text{with probability } (1 - p) \text{ in the first period of punishment;} \\
\emptyset, & \text{for the remaining } M - 1 \text{ periods.}
\end{cases}
$$

The strategy of the punishing player (in this case player 2) is,
\[ \sigma_2^*(.) = \begin{cases} 
\alpha, & \text{in the first period of player 1's punishment phase;} \\
0, & \text{in the remaining } M - 1 \text{ periods if player 2 won in the first period;} \\
0, & \text{otherwise.}
\end{cases} \]

If both players fail the statistical test, the punishment phase \( P_1 \) is followed by a second punishment phase \( P_2 \) for the second player. Suppose that player 2 has also failed the statistical test. Then after player 1 is punished in phase \( P_1 \) it is the turn of player 2. The punishment phase \( P_2 \) also lasts for \( M \) periods. The strategies of the players are symmetric to those above, only the probability with which player 2 randomizes between the two options, (either that player will win the remaining bids of the punishment phase, or that none of the players will participate in those bids) is probability \( q \), which is not equal to \( p \). The values of \( p \) and \( q \) will be defined later.

To conclude, there are six parameters defining the scheme:
- \( \alpha \) - the (small) constant by which the players multiply their values when they bid.
- \( N \) - the length of the cooperation phase
- \( K \) - the number of sub-intervals to which we divide the interval \([0, 1]\)
- \( M \) - the length of the punishment phase
- \( p \) and \( q \) - probabilities for the punished player to randomize with, at the beginning of the punishment phase.

### 4.3 \( \epsilon \)-Consistent Equilibrium

When we consider games with private monitoring and a continuum of actions at each period, the "perfectness" concept which applies is that of Perfect Bayesian.\(^1\)

Denote the game by

\[ G = \langle N = \{1, 2\}, \{\mathcal{H}_i\}, \{\Theta_i\}, \{\Sigma_i\}, \{\pi_i\} \rangle \]

---

\(^1\)Sequential Equilibrium requires convergence of beliefs over histories, thus one would need to go into details of convergence over histories with continuum of actions.
Here $H_i$ is the set of history for player $i$, $\Theta_i$ is the set of types, $\Sigma_i$ the set of strategies and $\pi$ the pay off function. The histories are private. So player $i$’s conditional beliefs at stage $t$ is denoted by $\mu_i(\cdot|h_t^i)$. In our case $h^i_t$ contains all the relevant information up to stage $t$ including realization of player $i$’s type in stage $t$. Given that in our model actions are not observable, but consequences are, the conditional probabilities are on the consequences and on the actions taken by the player whose beliefs we discuss i.e., on the private history of the game. Now given an action pair $a = (a_1, a_2)$, a stage $t$ and a public history $h_{t-1}^t$, $c(h_{t-1}^t, a)$ is the unique concatenation. We denote the belief profile by $\mu = (\mu_1, \mu_2)$

**Definition 1:** An assessment $(\mu, \sigma)$ is reasonable if for all history profiles $h^t = (h_1^t, h_2^t)$:

1) Bayes Rule is used to update beliefs whenever possible: i.e. Given any stage $t$ and any history $h^t = (h_1^t, h_2^t)$, and any action pair $a$, and public history $h_{t-1}^t$ that is compatible with $h^t$,

$$\sigma(a|h^t) = \mu(c(h_{t-1}^t, a)|h^t)$$

**Definition 2:** A PBE of a game defined above is an assessment satisfying

1) $(\mu, \sigma)$ is reasonable

2) for each period $t$ and history profile $h^t$, the continuation strategies $\sigma(\cdot|h^t)$ are a Bayes Nash Equilibrium given the beliefs, $\mu(\cdot|h^t)$.

For reasons that we will detail later, we use epsilon-equilibrium, i.e., the players’ strategies are epsilon-best responses, given their beliefs. We can use, however, the more restrictive form of epsilon-equilibrium - $\epsilon$-consistent equilibrium (Lehrer and Sorin, also known as contemporaneous epsilon equilibrium (Mailath et. al.)). In such an equilibrium the ”allowed loss” of epsilon is weighted always from the period being played now and forward. So, for example, there is no period ”far enough in the future” such that from that period on the players can be instructed to play anything since the weight (of this ”future”) is insignificant.

**Definition 3:** An $\epsilon$-Consistent Bayes Equilibrium of a game defined above is an assessment satisfying

1) $(\mu, \sigma)$ is reasonable
2) for each period $t$ and history profile $h^t$, the continuation strategies $\sigma(\cdot|h^t)$ are $\epsilon$-Consistent Bayes Equilibrium given the beliefs, $\mu(\cdot|h^t)$.

### 4.4 Main Result

In this section we derive the main result of the chapter. First we show that by playing according to the advised strategy any player can pass the statistical test with a very high probability. As noted before, there are six parameters defining the scheme, $D = \langle \alpha, N, K, M, p, q \rangle$. In the following lemma we show that if player $j$ is following the advised strategy, the probability that he will fail player $i$’s test goes to zero as the length of the cooperation phase increases.

**Lemma 4.1:** For all $\epsilon > 0$ and $K'$ there exists $N'$ such that for all $N > N'$, if the players are patient enough and both players are using the **advised strategies** $(\sigma_1^*, \sigma_2^*)$, we have

$$\text{Prob}(t_{ij}(\sigma_1^*, \sigma_2^*) > T) < \epsilon$$

### 4.4.1 Proof of Lemma 4.1

Let $n$ be the number of periods out of the $N$ periods of the cooperation phase, where player $j$ bid within segment $[\frac{k}{K}, \frac{k+1}{K}]$. Let $X$ be the number of times out of those $n$ periods when player $j$ won. If both players act according to the advised strategy then $n \sim \text{Bin}(N, \frac{1}{K})$, and $X|n \sim \text{Bin}(n, \frac{k+0.5}{K})$.

In order for player $i$ to pass the test, it has to be that for each one of the segments $k = 0, 1, ..., K - 1$

$$\left( \frac{k + 0.5}{K} - \frac{X}{n} \right) \frac{n}{N} < \frac{1}{K^3}$$

We will show that if both players are following the advised, then for any player $i$, the probability of failing the test approaches 0 for large enough $N$. 
Failing the test is a union of $K$ events (one event for each segment). Each of those events depends on the two random variables $n$ and $X$.

For passing the test in segment $k$, it is sufficient that both of the following events take place:

1. $\frac{n}{N} < \frac{1}{K} + \frac{1}{K}$
2. $\frac{k+0.5}{K} - \frac{X}{n} < \frac{1}{2K^2}$

In order to prove that the probability of event 1 and event 2 approaches 1, we will use Chernoff’s bound. From Chernoff’s bound, if $X \sim Bin(n, p)$ then for any $\delta > 0$,

$$\text{Prob} \left[ X > (1 + \delta)\mu_X \right] < \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^{\mu_X}$$

where $\mu_X$ is the expected value of the random variable $X$. Thus for event 1:

$$\text{Prob} \left[ \frac{n}{N} < \frac{1}{K} + \frac{1}{K} \right] = \text{Prob} \left[ n < \frac{N}{K} + \frac{N}{K} \right] = 1 - \text{Prob} \left[ X > 2\mu_X \right]$$

Using Chernoff’s bound:

$$> 1 - \left( \frac{e^\delta}{2\delta^2} \right)^{\frac{N}{K}}$$

which approaches 1 as $\frac{N}{K} \to \infty$

Now we wish to show that the probability that $\frac{k+0.5}{K} - \frac{X}{n} < \frac{1}{2K^2}$ happens approaches 1.

Event 1 puts an upper bound on $n$. However for the probability of event 2 to approach 1, $n$ cannot be “too” small. Specifically we want $\frac{N}{K}$ to be large. However, we have already established that with a probability that approaches 1 (see step 2)

$$n \geq \frac{N}{2K}$$

Assuming $n \geq \frac{N}{2K}$, we wish to show that the probability that $\frac{k+0.5}{K} - \frac{X}{n} < \frac{1}{2K^2}$ approaches 1. Now,

$$\text{Prob} \left( \frac{k + 0.5}{K} - \frac{X}{n} < \frac{1}{2K^2} \right) = \text{Prob} \left( X > n \left( \frac{k + 0.5}{K} - \frac{n}{2K^2} \right) \right)$$
Let $Y$ be the number of periods in which player $j$ bid within segment $k$ and lost, that is, $Y = n - X$. If both players are following the advised strategy, then, $Y \sim Bin(n, 1 - \frac{k+0.5}{K})$. Therefore we have,

$$\Pr(X > n \left(\frac{k+0.5}{K} - \frac{n}{2K^2}\right)) = \Pr(Y < n - n \frac{k+0.5}{K} + \frac{n}{2K^2})$$

$$= 1 - \Pr(Y \geq n \left(1 - \frac{k+0.5}{K} + \frac{n}{2K^2}\right))$$

$$= 1 - \Pr(Y \geq \mu_Y + \frac{n(K-k-0.5)}{K} \frac{n}{2K^2} n(K-k-0.5))$$

$$= 1 - \Pr\left(\mu_Y \left(1 + \frac{1}{2K(K-k-0.5)}\right) \geq \mu_Y \left(1 + \frac{1}{2K(K-k-0.5)}\right)\right)$$

$$= 1 - e^{\frac{n}{K}3(4(K-k-0.5)^2)}$$

When $n \geq \frac{N}{2K}$, the expression in the above equation tends to 1 as $\frac{N}{K^4}$ tends to infinity.

For the player to pass the test we need those events to take place in any segment, but as long as $\frac{N}{K^4}$ faster than $ln(K)$ does, then multiplying this expression by $K$ (which is the bound for the union of the $K$ events), will still give us a probability of passing the test which approaches 1.

### 4.4.2 The Punishment Phase

We first consider the case where only one player (say player 1) has failed the test. During the punishment phase player 1 will receive nothing and in the first of his $M(\delta)$ periods of punishment he will randomize over whether player 2 will win the entire set of bids during player 1’s punishment phase or not. Any profitable deviation during the punishment phase is publicly detected with probability 1. Following such a deviation the players will move to playing the one period NE for ever.

Both players announce the result of their statistical tests, and then the punishments take
place. In case no body failed the test, the players restart on the equilibrium path. Assume that player 1 failed the test. We need to show that there exists a $p^*$ such that when player 1 (the punished player) is randomizing over whether player 2 will win the bids during the punishment or not with probability $p^*$, player 2 is indifferent between punishing and not punishing, when we disregard the period which consists the statistical test.

**Lemma 4.2:** There exists a $p^* \in (0, 1)$ such that player 2 is indifferent between punishing and not punishing (disregarding the statistical test phase).

### 4.4.3 Proof of Lemma 4.2

Player 2 is the punishing player. From the strategy of player 1 it follows that for $p = 1$, given that player 2 is following the advised strategy, player 2’s payoff for the entire $M$ periods of punishment is 0. For $p = 0$ player 2’s payoff is $\frac{1}{2}$. Since the expected payoff of player 2 in the continuation in case he did not punish player 1 is between 0 and $\frac{1}{2}$, it follows that, there exists a $p^*$ such that player 2 is indifferent between punishing and not punishing.

Now we consider the case where both players fail the test. For the player who is punished first (in this case player 1) should have the same continuation payoff in case he is punishing player 2 and in case he is not. This is the same as before, (Lemma 4.2). The tricky part is when it is turn for the player who is punishing first and being punished later (player 2 in this case), to convey his message.

A player who is punishing first and being punished later can use punishing in order to delay his own punishment. Keeping the same $p^*$ as before will make the first player favor punishing in any case when he believes that there is a positive probability that he failed his statistical test. We will choose a different probability $q^*$, that player 1 will randomize with to decide whether player 2 will win the bids during player 2’s punishment, in case both players fail the statistical test. Note that such randomization takes place during the punishment phase when the results of the statistical tests are already known. Using the different probabilities in the different situations will make player 2 indifferent to punishing
whether he failed or passed the test, and hence indifferent for every belief he has regarding failing the statistical test.

**Lemma 4.3:** Suppose that both players have failed the statistical tests. Suppose that player 2 is to be punished second. Also suppose that both players are following the advised strategies. Then there exists a $q^*$ such that player 2 is indifferent between punishing and not punishing.

### 4.4.4 Proof of Lemma 4.3

By punishing first, player 2 postpones his punishment by $M$ periods. This results in a gain of $\frac{1}{3}(1 - \delta^M)^2$. We now prove that a probability $q^*$ (with which player 1 randomizes over whether player 2 wins the bids or not during player 1’s punishment phase in case both players fail the test) which will make player 2 indifferent to punishing when failing his own test exists. Observe that using $q = p^*$ as before player 2 gets $\frac{1}{3}(1 - \delta^M)$ during player 2’s punishment. With $q = 0$ his payoff during player 2’s punishment is 0. We wish his payoff to be $\frac{1}{3}(1 - \delta^M) - \frac{1}{3}(1 - \delta^M)^2$, and since $0 < \frac{1}{3}(1 - \delta^M) - \frac{1}{3}(1 - \delta^M)^2 < \frac{1}{3}(1 - \delta^M)$ we have shown that there exists a $q^*$ that does the job.

**Proposition 4.1:** At the punishment phase, the advised strategy pair $(\sigma^*_1, \sigma^*_2)$ are mutually $\varepsilon$-best responses.

**Proof:** First we show that the punishing player will not gain by deviating from the advised strategy. From lemma 4.2 and 4.3 it follows that the punishing player is indifferent between punishing and not punishing. A punishing player can benefit by deviating from the advised strategy at the punishment phase, only by bidding when he is not supposed to bid. Except for the first period of punishment, the same is true also of the punished player. Such an action is immediately detected, and the game moves to the one period NE for ever. We need to show that such a deviation is unprofitable. The lowest average continuation payoff for a player approaches $\frac{1}{3}\delta^N$ as $\delta$ approaches 1. The highest payoff from one period deviation
is \((1 - \delta)\). The one period NE is \(\frac{1}{6}\). For \(\delta\) close enough to 1 and \(\delta^N\) strictly larger than \(\frac{1}{2}\) the following inequality holds:
\[
(1 - \delta) + \delta \frac{1}{6} < \frac{1}{3} \delta^N
\]
(4.1)

We now come to the first period of punishment. In this period, the punished player is advised to bid 1 with positive probability. Given his valuation the player has an incentive to deviate. However, this is only one period, and as the discount factor grows, the weight of this period decreases, and for large enough \(\delta\), the gain from deviation will be less than \(\epsilon\).

We now move on to the Statistical Test phase. There are two sub-phases: the communication of the statistical test result and the test phase.

4.4.5 The Communication Phase

Lemma 4.4: At the communication phase, the advised strategy pair \((\sigma^*_1, \sigma^*_2)\) are mutually \(\epsilon\)-best responses.

**proof:** At the communication phase there is one player, say player 1, who should communicate the result of the statistical test by bidding zero in case the opponent failed the test, and bidding 1 otherwise. From lemma 4.2 (and lemma 4.3) player 1 (and, in turn, player 2) is indifferent between the two possible continuations. Of course, given his valuation during the communication player 1 is no longer indifferent. However, this is only one period, and as the discount factor grows, the weight of this period decreases, and for a large enough \(\delta\) will be below \(\epsilon\).

As for player 2 (the player whose statistical-test result is communicated), he is not supposed to bid i.e. bid 0. As mentioned before there are three possible public signals \(s_1, s_2\) and \(s_0\). Player 2 by deviating can only change the signal from either \(s_1\) or \(s_0\) to \(s_2\). Such a deviation is publicly detected and hence, it will lead to switching to the one-period Nash equilibrium.

**Note:** At the test phase, the advised strategy pair \((\sigma^*_1, \sigma^*_2)\) are mutually best responses.
This is because that during the test phase the players simply calculates whether the opponent passed or failed the test. There are no profitable deviations here.

4.4.6 The Cooperation Phase

We now move to the crucial cooperation phase. We have shown so far that the advised strategy combination \((\sigma^*_1, \sigma^*_2)\) are mutually \(\epsilon\)-best responses in the statistical test, communication and punishment phases. We want to show now that in the cooperative phase they are \(\epsilon\)-best responses.

Fix a \(K\). This implies that the interval \([0, 1]\) has \(K\) segments. Consider now the segment \([\frac{k}{K}, \frac{k+1}{K}]\), \(k \in \{1, ..., K\}\). Suppose that player \(j\) is using the advised strategy \(\sigma^*_j\). This implies in each period \(t\), \(t = 1, ..., N\) of the cooperation phase, player \(j\) is using the strategy,

\[\sigma_{j,t} = \alpha v_{j,t}\]

If player \(i\) was using the advised strategy as well, then if we consider the interval \([\frac{k}{K}, \frac{k+1}{K}]\), player \(i\) will be bidding on average about \(N \left(1 - \frac{k+1}{K}\right)\) times above the interval, about \(N \left(\frac{k}{K}\right)\), below the interval and \(N(\frac{1}{K})\) times within the interval.

There are two ways in which a player may deviate from the advised strategy - either bidding in the same segment as in the advised strategy (but maybe not bidding the exact sum within that segment), or bidding in a different segment. We will first show that by bidding in the same segment (perturbing the bids within the intervals) a player cannot gain much, and then we will show that a player cannot gain much by bidding in different segments altogether without being punished.

The first step is to show that in the cooperation phase if agent \(i\) is perturbing his bids only within the interval, then he cannot gain much. For that we consider the following kind of strategies for agent \(i\): the strategy \(\sigma_i\) agrees with \(\sigma^*_i\) in the statistical test, communication and punishment phases and in the cooperation phase is defined as follows: for all \(t \in N\),

\[\sigma_{i,t} = \alpha z_{i,t}\]
where $z_{i,t} \in \left[ \frac{k}{K}, \frac{k+1}{K} \right]$ whenever, $v_{i,t} \in \left[ \frac{k}{K}, \frac{k+1}{K} \right]$. Let $\Sigma_i^1$ be the set of all such strategies, that is the set of strategies where the player is allowed to perturb his bids within the advised segment.

**Lemma 4.5:** For any strategy $\sigma_i \in \Sigma_i^1$, it is the case that

$$\pi_i(\sigma_i, \sigma_j^*) - \pi_i(\sigma_i^*, \sigma_j^*) \leq \frac{2}{K}$$

**Proof:** Suppose that player $i$ is using a strategy $\sigma_i \in \Sigma_i^1$. For any $t \in \{1, ..., N\}$, the stage game payoff for player $i$, given that player $j$ is using the advised strategy $\sigma_j^*$ is

$$r_i(\sigma_{i,t}, \sigma_{i,t}^*) \leq \sum_{k=0}^{K-1} \left[ \frac{1}{K} \frac{k+1}{K} - \alpha \frac{k+1}{K} \right]$$

Therefore,

$$r_i(\sigma_{i,t}, \sigma_{j,t}^*) - r_i(\sigma_{i,t}^*, \sigma_{j,t}^*) \leq \sum_{k=0}^{K-1} \left[ \frac{1}{K} \frac{k+1}{K} - \frac{1}{K} \frac{k+\frac{1}{2}}{K} \right] \leq \frac{1}{2K}$$

The $r_i(\cdot)$'s are the stage game payoffs. The total payoff over the cooperation phase is the discounted sum of the stage game payoffs i.e., the total payoff is the weighted average of the stage game payoffs. Hence, the result follows.

We now show that if a player moves “too many” of his bids above a certain segment (meaning, he will bid in one of the segments above the advised one), then he will fail the test with high probability.

We have already observed that when a player, say 1 is using the advised strategy, for any segment $\left[ \frac{k}{K}, \frac{k+1}{K} \right]$, the player will end up bidding on average $N \left( 1 - \frac{k+1}{K} \right)$ times above the “upper” bound of the interval i.e., above $\frac{k+1}{K}$.

We need to be precise about what we mean by “too many”. The following proposition shows that if agent $i$ moves more than a proportion of $\frac{2}{K}$ of the bids above the border of some segment, then he risks a high probability of failing the test. In fact, the probability of failing the test approaches 1 as $N$ grows.
Let \( \Sigma'_i \) be the set of all strategies for player \( i \) such that there exists a segment \( k, \left[ \frac{k}{K}, \frac{k+1}{K} \right] \) such that player \( i \) deviates at least \( N \frac{\epsilon}{2K} \) of the periods and moves his bids above that segment (when his valuation is on or below that segment).

**Proposition 4.1:** For all \( \epsilon > 0 \), for any strategy \( \sigma_i \in \Sigma'_i \) and for any \( K \), there exists \( N_\epsilon \), such that for all \( N > N_\epsilon \),

\[
\text{Prob} \left( \left[ \left( \frac{k+1/2}{K} - \frac{p_i(k,k+1,\sigma_i,\sigma_j)}{n_i(k,k+1,\sigma_i)} \right) n_i(k,k+1,\sigma_i) \right] < \frac{1}{K^3} \right) < \epsilon.
\]

**proof:**

We will prove it in three steps.

The first step will show that if player \( i \) uses a strategy \( \sigma'_i \in \Sigma'_i \) then with a probability that approaches 1 as \( N \) grows, player \( i \) will bid at least \( \frac{K-k+0.8}{K} \) of the periods above this segment.

The second step is to show that if player \( j \) is using the advised strategy, then given any interval \( k \), with a high probability player \( j \) will bid at least \( \frac{N}{2K} \) times in that interval.

The third step is to show that indeed player \( i \) will fail the test with a probability that approaches 1 (as \( N \) grows).

**Step 1:** if player \( i \) uses a strategy \( \sigma'_i \in \Sigma'_i \) then with a probability that approaches 1 as \( N \) grows, player \( i \) will bid at least \( \frac{K-k+0.8}{K} \) of the periods above this segment.

**Proof of Step 1:**

Player \( i \) deviates during at least \( N \frac{\epsilon}{2K} \) of the periods and bid above segment \( k \), so he has at most \( N \left( 1 - \frac{\epsilon}{2K} \right) \) periods when he correlated his bids and his valuations, relative to that segment, i.e. he bids above the segment when his valuation is above the segment, and below the segment when it is below it\(^2\).

Let \( X \) be the number of periods out of those remaining \( N \left( 1 - \frac{\epsilon}{2K} \right) \) periods, when he bids above the segment \( k \). Then \( X \sim Bin(N \left( 1 - \frac{\epsilon}{2K} \right), \frac{K-k-1}{K}) \). If the total number of periods

---

\(^2\)The most profitable way to deviate by moving a fraction of the bids above a segment is to move the bids of values that are below the values of that segment but close to it, and bid above the segment, and in the rest of the periods to bid above the periods only if the valuation is above the segment and below when below. Otherwise there is a negative correlation between the bids and the values, which cannot be profitable.
when player \( i \) bids above that segment should be at most \( N \left( \frac{K-k+0.8}{K} \right) \), then \( X \) has to be at most \( N \left( \frac{K-k-1.2}{K} \right) \). Using Chebychev’s inequality we obtain:

\[
\text{Prob} \left( X < N \left( \frac{K-k-1.2}{K} \right) \right) = \text{Prob} \left( X - \mu_X < \frac{N}{K} \left( -2.2 + \frac{2k}{K} - \frac{2}{K} \right) \right) \\
\leq \frac{N \left( 1 - 2 \right)^{-1} \left( -2.2 + \frac{2k}{K} - \frac{2}{K} \right)^2}{\left( \frac{N}{K} \right)^2}.
\]

and since for \( K \) large enough we can bound \( (2.2 - \frac{2k}{K} - \frac{2}{K}) \) from above by 0.1, and we can see that this probability approaches zero if \( \frac{N}{K^2} \) approaches infinity.

**Step 2:** Suppose that \( j \) is bidding according to the advised strategy. Given any interval \( \left[ \frac{k}{K}, \frac{k+1}{K} \right] \), \( k \in \{0, ..., K-1 \} \), the probability that \( j \) bids less than \( \frac{N}{2K} \) times in that interval approaches 0 as \( \frac{N}{K^2} \to \infty \).

**Proof of Step 2:**

Let \( Y = \) the number of times agent \( j \) bids within the segment \( \left[ \frac{ak}{K}, \frac{a(k+1)}{K} \right] \). Now, \( Y \sim \text{Bin}(N, \frac{1}{K}) \). Using Chebyshev’s inequality, we get, \( P \left( |Y - \frac{N}{K}| \geq \frac{N}{2K} \right) \leq \frac{N \left( 1 - \frac{1}{K} \right)}{\left( \frac{N}{2K} \right)^2} \). As \( \frac{N}{K^2} \) gets large, the last expression goes to zero. □

**Step 3:** If player \( i \) uses a strategy \( \sigma_i' \in \Sigma_i' \) and player \( j \) is using the advised strategy, then player \( i \) will fail the statistical test with a probability that approaches 1 as \( \frac{N}{K^2} \) grows.

**Proof of Step 3:**

We will show that the probability of the intersection of the following three events approaches one:

Event 1: player \( i \) bids at least \( \frac{K-k+0.8}{K} \) of the periods above the segment \( \left[ \frac{k}{K}, \frac{k+1}{K} \right] \).

Event 2: player \( j \) bids at least \( \frac{N}{2K} \) times in segment \( \left[ \frac{k}{K}, \frac{k+1}{K} \right] \).

Event 3: player \( i \) fails the test.

Step 1 and Step 2 show that event 1 and event 2 take place with a probability that approaches 1, so their intersection also takes place with a probability that approaches 1. What is left to show is that given that events 1 and 2 take place, event 3 takes place as well with a probability that approaches 1.
Let \( Z \) be the number of periods when player \( j \) bids within segment \( \left[ \frac{k}{K}, \frac{k+1}{K} \right] \) and won. Given that event 1 and event 2 take place, we know that \(^3 Z|n \sim Bin\left(n, \frac{k-0.8}{K}\right)\). In order for player \( i \) to pass the test, it has to be that:

\[
\left[ \left( \frac{k+1/2}{K} - \frac{p_j(k,k+1,\sigma,\sigma_j^*)}{n_j(k,k+1,\sigma_j^*)} \right) \frac{n_j(k,k+1,\sigma_j^*)}{N} \right] \leq \frac{1}{K^3}.
\]

or:

\[
Z > n\frac{k+0.5}{K} - \frac{N}{K^3}
\]

Now,

\[
(*) \quad \text{Prob} \left[ Z > n \frac{k+0.5}{K} - \frac{N}{K^3} \right] = \text{Prob} \left[ Z > n \frac{k-0.8}{K} + n \frac{1.3}{K} - \frac{N}{K^3} \right]
\]

\[
\leq \text{Prob} \left[ \left| Z - n \frac{k-0.8}{K} \right| > n \frac{1.3}{K} - \frac{N}{K^3} \right]
\]

For \( n \geq \frac{N}{2K} \) we have:

\[
n \frac{1.3}{K} - \frac{N}{K^3} > \frac{1.3n}{K} - 2 \left( \frac{N}{2K} \right) = \frac{1.3n}{K} - \frac{2n}{K^2} = n \left( \frac{1.3 - 2}{K} \right)
\]

For \( K \) large enough we can bound \( \frac{2}{K} \) so that \( \frac{2}{K} < 0.3 \). Hence we have,

\[
n \frac{1.3}{K} - \frac{N}{K^3} > n \frac{1.3}{K} - \frac{2}{K} = n \frac{1.3 - 2}{K} > n \frac{1}{K}
\]

So combining \((*)\) with the last inequality we get that,

\[
\text{Prob} \left[ Z > n \frac{k+0.5}{K} - \frac{N}{K^3} \right] \leq \text{Prob} \left[ \left| Z - n \frac{k-0.8}{K} \right| > \frac{n}{K} \right]
\]

Using Chebyshev’s inequality we have,

\[
\text{Prob} \left[ Z > n \frac{k+0.5}{K} - \frac{N}{K^3} \right] \leq \frac{n^{k-0.8} K^{k+0.8}}{K^3}
\]

\(^3\)Of course, \( \frac{k-0.8}{K} \) is a bound due to event 1, and the probability of failing the test will be higher if consider other probabilities that agree with event 1.
Given that, \( n \geq \frac{N}{2K} \), the last expression goes to zero as, for any fixed level of \( K \), \( N \) increases relative that \( K \), or in other words, \( \frac{N}{K} \) goes to infinity.

This completes the argument. ■

We are now ready to show the following theorem:

**Theorem 4.1:** For \( \epsilon > 0, \eta > 0 \), there exists \( \delta' \in (0,1) \) and integers \( N' \) and \( K' \) such that for all \( \delta \in (\delta',1) \), \( N \geq N' K \geq K' \), and for all \( i \in \{1,2\} \):

a. for all \( \sigma_i \in \Sigma_i \) \( \Pi_i(\sigma_i, \sigma_j^*) - \Pi_i(\sigma_i^*, \sigma_j^*) < \epsilon \),

and

b. \( \Pi_i(\sigma_i^*, \sigma_j^*) > \frac{1}{3} - \eta \).

We first prove the following intermediate step:

**Step 1:** Consider a strategy \( \sigma_i \in \Sigma_i' \). Then,

\[
\Pi_i(\sigma_i, \sigma_j^*) < \Pi_i(\sigma_i^*, \sigma_j^*)
\]

**proof:** Suppose, player \( i \) is using strategy \( \sigma_i \in \Sigma_i' \). As shown in proposition 4.1, if player \( j \) is using the advised strategy, player \( i \)'s probability of punishment is close to 1. Let \( p_i(N, \sigma_i, \sigma_j^*) \) be the probability that player \( i \)'s passes the statistical test using strategy \( \sigma_i \). Note that \( p_i(N, \sigma_i, \sigma_j^*) \to 0 \) as \( N \to \infty \). With a slight abuse of notation, we drop the arguments in \( p_i(.,.,.) \) and simply denote by \( p_i \). Let \( M \) denote the length of the punishment period. Now,

\[
\sup_{\sigma_i \in \Sigma_i'} \Pi_i(\sigma_i, \sigma_j^*) = (1 - \delta^N) + (1 - p_i)\delta^{N+M}X + p_i\delta^N X
\]

In the last equation, \( X \) is the pay-off to each player over the cooperation phase, if they are both following the advised strategy. Note that \( X \sim \frac{1}{3} \). If both players are following the advised strategy, pay-off to player \( i \) is,

\[
\Pi(\sigma_i^*, \sigma_j^*) = (1 - \delta^N)X + (1 - \epsilon)\delta^{N+M}X + \epsilon\delta^N X
\]
The difference between the two payoffs is,

\[ Z = (1 - \delta^N)(1 - X) + (1 - p_i - \epsilon)[\delta^N M X - \delta^N X] \]

To show \( Z < 0 \), as \( \delta \rightarrow 1 \) is equivalent to showing,

\[ \frac{M}{N} > \frac{1 - X}{X(1 - p_i - \epsilon)} \]

Taking \( M > \frac{N}{2} \) does the job. ■

Proof Of Theorem 4.1:

We need to show that for any given pair of \( \epsilon \) and \( \eta \), we can find \( N, M, T \) and \( K \) such that the advised strategy are \( \epsilon \)-best responses and the payoffs for the players when employing the advised strategy is within an \( \eta \)-distance from the efficient payoffs.

From Proposition 4.1 it follows that for the advised strategy pair \( (\sigma_1^*, \sigma_2^*) \) to be \( \epsilon \) best responses during the punishment phase, we need the discount factor \( \delta \) to be close enough to 1, and \( \delta^N \) to be strictly larger than \( \frac{1}{2} \).

From step 1 above, it follows that, for \( \delta \) close to 1, no player \( i \) is going to use any strategy \( \sigma_i \in \Sigma_i' \) against the advised strategy \( \sigma_j^* \) if the probability \( p_i(N, \sigma_i, \sigma_j^*) \) is close to zero. From the proof of proposition 4.1 we know that \( p_i(N, \sigma_i, \sigma_j^*) \rightarrow 0 \) as \( \frac{N}{K} \) goes to infinity. Also, it must be the case that any player \( i \) using strategy \( \sigma_i^* \) against \( \sigma_j^* \) must be able to pass the statistical test with a high probability. From the proof of lemma 4.1 it follows that such an event occurs with a probability that tends to 1 as \( \frac{N}{K} \) tends to infinity. From step 1 above, as long as \( \delta \) close to 1, for any fixed \( K \), \( \frac{N}{K} \) is large enough and \( M > \frac{N}{2} \), any agent \( i \) is not going to use a strategy \( \sigma_i \in \Sigma_i' \). From lemma 4.5 it follows, that, for any strategy \( \sigma_i \in \Sigma_i \setminus \Sigma_i' \),

\[ \Pi_i(\sigma_i, \sigma_j^*) - \Pi(\sigma_i^*, \sigma_j^*) \leq \frac{1}{2K} \]

Now consider any \( \epsilon > 0 \). Fix \( K \) such that \( \epsilon > \frac{1}{2K} \). Then we get that, for all \( \sigma_i \in \Sigma_i \),

\[ \Pi_i(\sigma_i, \sigma_j^*) - \Pi(\sigma_i^*, \sigma_j^*) < \epsilon \]
We now want to find the conditions under which the payoffs of the players who conform to the advised are within an \( \eta \)-distance from the efficient payoff.

From Lemma 4.1, for any \( K \), there exists an \( N' \) such that for \( N > N' \) the probability that a conforming player will pass the test is small enough. Let that probability be \( \xi \).

Suppose that both players are following the advised strategies. Let the total payoff of a conforming player be \( X \), then \( X \) satisfies

\[
\left(1 - \delta^N\right) \frac{1}{3} + \xi \delta^N (1 - \delta^M) 0 + \xi \delta^{N+M} X + (1 - \xi)\delta^N X = X
\]

Then,

\[
X = \frac{1}{3} \frac{(1 - \delta^N)}{1 - \delta^N + \xi (\delta^N (1 - \delta^M))} = \frac{1}{3} \frac{1}{1 + \xi \frac{\delta^N (1 - \delta^M)}{(1 - \delta^N)}}
\]

For \( \xi \) small enough, and if \( \frac{1 - \delta^N}{1 - \delta^M} \) is bounded, then the payoff is close enough to \( \frac{1}{3} \), the efficient one.

To sum up, for any given pair of \( \epsilon \) and \( \eta \), we need to pick up \( K \) to be large enough, and then \( N \) large enough (so that \( \frac{N}{K^2} \) will be large enough, and so that \( \xi \) will be small enough) and then we need to pick \( M \), so that \( M > \frac{N}{2} \), but also such that \( \frac{1 - \delta^N}{1 - \delta^M} \) is bounded. Using L’hopital rule, we can see that \( M \cong \frac{2}{3} N \) will do the job. Finally, we will need to pick up \( \delta \) large enough so that all conditions are satisfied.

4.5 Public Strategies

In this section we will show that relying only on public signals will lead to a situation where the cartel’s payoff is bounded away from the efficient payoff.

As mentioned in section 4.2, at any stage \( t \), the public history \( h^p(t) \) consists of the sequence of identities of the winners.

Skrzypacz and Hopenhayen (2004) show that, when the public history of the game consists of sequence of identities of winners in the previous rounds, the set of equilibrium payoffs
for the game is bounded away from the efficiency frontier. The following result is similar in spirit to proposition 1 in Skrzypacz and Hopenhayen.

For any given \( \delta > 0 \), let \( \Pi_\delta^i(\sigma^*_1, \sigma^*_2) \) be the payoff to player \( i \) from the strategy profile \( (\sigma^*_1, \sigma^*_2) \). Let \( \Pi_\delta^i(\sigma^*_1, \sigma^*_2) = \Pi_\delta^1(\sigma^*_1, \sigma^*_2) + \Pi_\delta^2(\sigma^*_1, \sigma^*_2) \). For any \( \epsilon > 0 \) there exists \( \delta' > 0 \) such that for \( \delta \in (\delta', 1) \), \( (\sigma^*_1, \sigma^*_2) \) is a \( \epsilon \)-equilibrium and \( \Pi_\delta^i(\sigma^*_1, \sigma^*_2) \in (\frac{2}{3} - \epsilon, \frac{2}{3}) \).

With a slight abuse of notation, let us denote by \( \mathcal{H}^p \) to be the set of public histories. Let \((\beta_1, \beta_2) : \mathcal{H}^p \times \mathcal{H}^p \to B \times B \) be a candidate public strategy profile such that for the given \( \epsilon > 0 \), and for a fixed \( \delta \), \((\beta_1, \beta_2)\) is an \( \epsilon \)-equilibrium. Given \( \epsilon \) and \( \delta \), let \( B(\epsilon, \delta) = \{ \text{the set of } (\beta_1, \beta_2) \text{ that are epsilon equilibria given the } \delta \} \).

\[
\hat{V} = \sup_{\beta_1, \beta_2 \in B(\epsilon, \delta)} \Pi(\beta_1, \beta_2)
\]

In the following we show that \( \hat{V} < \frac{2}{3} - \epsilon \). In order for \( \hat{V} \) to be close to \( \frac{2}{3} \) the following has to be true: over a weight \( \hat{q} \) of the periods, \( \hat{q} \) approaches 1, with a probability \( \hat{p} \) that is close to 1, at each stage the player with the higher valuation must win. We will show that with the limited information, provided by public history alone, that is not possible.

Any public history is a sequence data on the identity of the winners. Given that, at any stage a player has 2 possible actions – \( \{\emptyset, b \geq 0\} \) where as before \( \emptyset \) means non-participation. If at any stage \( t \) a player decides to give up his chance to win the object he will either participate and bid 0 or he will not participate. The next lemma shows that for the scheme to be efficient, the probability of non-participation needs to go down to zero.

**Lemma 4.6:** In an efficient framework, over a weight which approaches 1 of the periods the probability of Non-participation by any player approaches 0.

**proof:** Assume that the probability of Non-participation is \( p > 0 \), over a significant weight of the periods (a weight that does not approach zero) This implies that there exists a range of valuation, the lowest of which can be \([0, p]\) in which the player will not participate. With probability \( p \) the opponent will also have his valuation in the range \([0, p]\). So with probability \( p^2/2 \) the scheme is not efficient, because the bidder with the highest valuation will not win the bid. ■
CHAPTER 4. REPEATED TWO-BIDDERS AUCTION

In a scheme with public strategy when both players are participating, there are only two publicly observable signals. Either player 1 won or player 2 won. Hence, given any history, with a probability that approaches 1, there are only two possible continuation payoffs for each player: the continuation payoff in case he wins the next round and the continuation payoff in case the opponent won.

Observe that, for any given public strategy pair \((\beta_1, \beta_2)\), for \(\Pi(\beta_1, \beta_2)\) to be close to \(\frac{2}{3}\), it must be the case that, with a probability \(q \to 1\), each player’s bid \(b\) is close to zero. Otherwise the scheme is not efficient. In other words, given lemma 4.6, in order for a scheme to obtain efficient outcome, with probability that approaches one both players participate and bid at most \(\omega\), and \(\omega\) should approach zero.

Let us define \(W\) to be the continuation payoff of a player who wins the next round and \(P\) to be the continuation payoff in case the opponent won. In the next lemma we will show that the probability that the difference between winning now and getting \(W\) in the future, and losing now and getting \(P\) in the future is at most \(\epsilon\) tends to zero.

**Lemma 4.7:** The probability \(p\) that for any player, the absolute difference between winning now and getting \(W\) in the future, and losing now and getting \(P\) in the future, is less than \(\epsilon\) tends to zero i.e., for all \(\omega > 0\) there exists \(\eta > 0\) such that

\[
\Pr(|(1 - \delta)(v - b) + \delta W - \delta P| < \omega) < \eta
\]

and both \(\omega\) and \(\eta\).

**proof:** Roughly, if the difference is more than omega, then deviating by bidding more than omega would be profitable for large enough valuations.

Observe that there may be one remaining case where one player does not participate. Since the probability of such an event is going to zero and since the continuation payoff is bounded by 1, the range of valuations for which a player is indifferent between winning and participating still goes to zero.

**Theorem 4.2:** For any discount factor \(\delta < 1\), in any \(\epsilon\)-PPE, the average payoffs of the players are bounded away from the efficiency frontier.
**proof:** From lemma 4.7 it follows that with a probability $\mu \to 1$, either (i) a player strictly prefers winning and paying $\omega$ over participating, or (ii) a player strictly prefers participating and paying 0.

Observe that if a player prefers to win for a certain valuation then he also prefers to win for all valuations above it. Let $R_1$ be such that player 1 prefers to win when his valuations are in $[R_1, 1]$ and define $R_2$ analogically. Observe that the range of valuations where player $i$ strictly prefers to lose approaches $[0, R_i)$. There are four cases which will be considered separately.

**Case i:** $R_1 = R_2 = R < 1$. In this case, when $v_1, v_2 < R$, both players participate and bid 0. So with a probability $R^2/2$, it will not be the player with the highest valuation who will win the object.

**Case ii:** $R_1 < R_2 < 1$. In this case, with a probability $\frac{1}{2}(R_2 - R_1)$ player 2 will have the higher valuation, but he strictly prefers to participate, while player 1 strictly prefers to win. So the object will not be awarded to the highest valuation player.

**Case iii:** In case at least one of the $R_i$s equals 1 and the other is less than 1, say $R_1 = 1, R_2 < 1$ then with probability $R_2(1 - R_2)$ it is the case that player 1 has the higher valuation, but he prefers to only participate.

**Case iv:** The last remaining case to consider is where both $R_1 = R_2 = 1$. In this case again, both players prefer to participate. Hence the good is given “randomly” and not, with probability 1, to the player with the highest valuation.

The four cases together completes the proof of the theorem. ■
4.6 Discussion: Correlated Valuations

In this section we provide a short discussion of the problem with correlated valuations.

In case of correlated types, we will go over the various proofs, and discuss the conditions under which they still apply.

The proof of Lemma 4.1 does not assume independent types, and neither does the proof of Lemma 4.2 and 4.3. The proof of Proposition 4.1 assumes that the one period Nash equilibrium is strictly Pareto dominated by the collusive payoff, and so does the proof of Lemma 4.4.

The proof of Lemma 4.5 is the challenging one. For this proof to read smoothly, we will need to be able to divide the segment \([0,1]\) into \(L(K)\) subsegments that are smaller and smaller such that the two following conditions hold:

a. For any valuation of one player, the probability of the opponent to bid within a segment can be bounded from above by an order of \(\frac{1}{K}\). Otherwise, a player would be able to profit from changing his bids within a distance of one segment.

b. For any valuation of one player the probability of the valuations of the opponent to be within each segment is high enough so that the opponent will bid a significant amount of periods within each segments. Otherwise, a player might gain by shifting his bids above a segment in which he believes the opponent will not bid much, hence the statistical test for this segment will not be accurate enough to detect such a deviation.

It is not clear to us, whether an example where the types are correlated and collusion is not possible exists.

We do not know whether there exists a distribution of correlated values for which the collusion is not possible.

This discussion only gives a rough idea about the issues to consider when facing a correlated priors case, but a more detailed discussion is beyond the scope of this work.
Bibliography


