What You Get is What You See: Cooperation in Repeated Games with Observable Payoffs[*]

Galit Ashkenazi-Golan[†] and Ehud Lehrer[‡]

January 19, 2019

Abstract: We consider two-player repeated games, where the players observe their own payoffs with a positive probability. Typically, a player observes neither the other's actions nor her own payoffs. We show that when costly communication is available to the players and when they are patient enough, being aware of her own payoffs suffices to provide the players with any strictly efficient payoff by sequential equilibrium.

Keywords: Discounted repeated games, observable payoffs, efficient payoffs, information matrix, sequential equilibrium.

[*] This research was supported in part by ISF through Grant #963/15. The authors wish to thank David Lagziel and Eilon Solan.

[†] Tel Aviv University, Tel Aviv 69978, Israel. e-mail: galit.ashkenazi@gmail.com.

[‡] Tel Aviv University, Tel Aviv 69978, Israel and INSEAD, Bd. de Constance, 77305 Fontainebleau Cedex, France. e-mail: lehrer@post.tau.ac.il.

1 Introduction

The main result of the theory of repeated games is the Folk Theorem (Aumann and Shapley (1994), Rubinstein (1994), Fudenberg and Maskin (1986)). It states that in infinitely repeated games, when the players are sufficiently patient every feasible and individually rational payoff can be sustained by a subgame-perfect equilibrium. In particular, all efficient and individually rational payoffs can be obtained as equilibrium payoffs. This classical result relies on the assumption that the players have perfect monitoring: each player's actions are perfectly observable by his opponents. The extent to which the players can cooperate when they do not observe the opponents' actions, but rather receive noisy signals that depend on the actions taken, is still mostly unknown.

In this paper, we analyse two-player repeated games in which the players cannot fully monitor each other's actions. Rather, players receive noisy signals that reveal with a positive probability their own payoffs. This model encompasses in particular the cases of full monitoring and the case where the players always observe their own payoffs, as discussed in Lehrer (1992c). When a player observes only his own payoffs, he cannot fully monitor but he can obtain partial information about the other player's actions. A natural question arises as to whether this information is sufficiently rich to enable the players to sustain efficient payoffs in equilibrium. The main contribution of the current paper is showing that any strictly Pareto efficient payoff can be supported by a sequential equilibrium, if costly communication is available and the players are sufficiently patient. A simple consequence of this result is that any combination of a strictly Pareto efficient payoff and Nash equilibrium payoff of the one-shot game, can be also supported by a sequential equilibrium.

In general, repeated games with imperfect monitoring may be divided into two types. The first type consists of games where players obtain private signals and their strategies may depend on these signals. In a series of papers[1], Lehrer discusses this subject and fully characterizes the set of equilibrium payoffs in certain families of undiscounted games. Our model is closest to the one explored in Lehrer (1992c). In undiscounted repeated games any action that a player takes seldom enough has no effect on his total payoff. Indeed, in Lehrer (1992c) the equilibrium construction relies on statistical tests that become rare as the game unfolds, and therefore have no impact on the payoffs. None of the techniques employed to sustain equilibrium payoffs in undiscounted games could be employed in the current paper.

The second type of repeated games with imperfect monitoring consists of games with public monitoring. In these games, after each period, all players observe the same signal which determines, along with the players' own actions, their payoffs. The solution concept typically

[1] See Lehrer (1989, 1990, 1992a,b).

employed in such games is public equilibrium, where players cannot use their private information; the strategies may depend only on the public signals (see, Abreu et al. (1986) Abreu et al. (1990) and Fudenberg et al. (1994)). This paper belongs to the first type.

Characterizing the set of equilibrium payoffs for infinitely discounted repeated games with private monitoring was described by Kandori (2002) as 'a simple hard open question'. More than ten years later, this question still remains difficult and open both for discounted and for undiscounted games. The difficulties in analysing repeated games with private signals might stem from different sources: (i) identifying the power of the correlation between the players that might be internally generated by private monitoring; (ii) detecting profitable deviations from the equilibrium path; (iii) establishing punishments and continuation payoffs for the cases when a deviation is detected; and (iv) specifying a system of beliefs to accompany the strategies of the players in order to create sequential equilibrium. For a comprehensive discussion of these issues the reader is referred to Mailath and Samuelson (2006). In what follows we elaborate on these difficulties and specify which ones are particularly relevant to our model.

The first difficulty in studying equilibrium payoffs in repeated games with imperfect monitoring, is that private monitoring may actually serve as a correlation mechanism among the players. In Lehrer (1991) this effect is called *internal correlation* . The extent to which private signals may serve as an internal correlation device is still open. However, in the current context, internal correlation does not play a role because no correlation is needed to sustain efficient payoffs.

The common pattern of equilibrium strategies in repeated games is that players follow a play path unless a deviation occurs. Play paths are designed in a way that makes profitable deviations detectable. The second difficulty to establish an equilibrium with imperfect monitoring, is that profitable deviations might go unnoticed and therefore undetectable. However, when the payoffs are observable with a positive probability, as in this paper, all profitable deviations from efficient payoffs are detectable with positive probability. Thus, this difficulty too does not arise in our model.

The two main contributions of this paper lie in the methods it offers for handling difficulties (iii) and (iv). The punishment phase is designed using a new tool that we call the *information matrix* and the beliefs are devised using a combination of communication and a specification of off-equilibrium path beliefs.

Punishments and the information matrix When a deviation is detected during the play path, the players switch to a temporary punishment mode. While detecting profitable deviations from strategies leading to efficient payoffs is easy, dealing with the punishment phase and establishing proper continuation payoffs is not trivial. In some cases though, as when the desired equilibrium payoff strictly Pareto dominates a one-period Nash equilibrium, a punishment can be easily designed: when a deviation occurs, the players switch to playing the dominated one-period Nash equilibrium. In such cases (see Fudenberg and Levine (2007), for example) a Nash-threat folk theorem is established. However, when the target equilibrium payoff does not strictly Pareto dominate a one-period Nash equilibrium, when players wish to effectively punish the deviator, they typically do it by playing mixed (minmax) actions. These actions are often not stage-game best responses. In order to provide incentives for the players to nevertheless use these mixed actions, the equilibrium strategies should specify continuation payoffs that would make the punishing player indifferent to all pure actions used. This method was developed by Fudenberg and Maskin (1996). When the monitoring is full, designing continuation payoffs that would incentivize the players to follow the punishment scheme is not difficult. However, when players cannot fully monitor each other, coordinating the continuation payoffs becomes difficult. The challenge is to design an effective punishment scheme in which players can only observe their own payoffs (with positive probability) and not others'.

In order to better explain this challenge, imagine that Player 2 wishes to punish Player 1. The problem is that Player 1 observes only his own payoff; he does not know Player 2's actions nor her payoffs, and he does not know what she knows about him. Without proper (future) continuation payoffs (to be given after the punishment phase is over) Player 2 would have no incentive to abide by the punishment instructions; she could profitably deviate without being detected. How then can we make sure that Player 2 follows her strategy and keeps punishing Player 1? The way to do it is to increase her future payoff when, during the punishment, she uses a low-paying action, and to reduce her future payoff when she uses a high-paying action. This way we make her indifferent between all actions used. But how can the players agree on continuation payoffs if they do not share the same information?

One of the two main contributions of this paper is to develop a method that enables a construction of adequate continuation payoffs. Despite the fact that the punished player cannot observe the punishing player's payoffs, we design a scheme that enables the players to coordinate the continuation payoffs, a scheme which in turn renders the strategies incentive compatible. This means, in particular, that the information embedded in the punished player's own payoff, as reflected in the information matrix, is sufficiently rich to sustain efficient payoffs in equilibrium by designing an effective punishment phase. The method by which we design the punishment scheme is constructed by translating the private information available to the punished player to a $0, 1$ matrix, called the *information matrix.* This information matrix enables us to find proper continuation payoffs after every history of punishment.

Communication and beliefs. The fourth difficulty is rooted in the definition of sequential equilibrium (Kreps and Wilson (1982)). The definition requires that an elaborate system of beliefs accompanies the strategies and that players would always best-reply to these beliefs. The beliefs should be defined after any history following any number of deviations from the equilibrium path, regardless of the number of deviating players, their identities, and the actions played. However, when players do not share common information about histories, keeping track of all their possible beliefs, regardless of how far from the equilibrium path they had gone, is a demanding task.

There are three ways to prevent players' beliefs from drifting too far apart: introducing communication, introducing a mediator and assuming substantial assumptions on the signal structure, such as that the signals are highly correlated ('almost public') or almost accurate ('almost perfect'). These are discussed below.

Games with communication. When a communication device that generates public signals is present, its signals can be used to coordinate the beliefs of the players. However, the presence of a communication device introduces yet a new challenge: how to provide the players with proper incentives to signal honestly. Compte (1998) and Kandori and Matsushima (1998) analyse games with communication. In these papers the general results involve at least three players. The players have incentives for honest signalling regarding a deviation because a deviation of a player is detected by a subset of players whose members are not affected by the punishment of the deviator. Therefore, the players that detect a deviation are not hurt when triggering a punishment. When signals are highly correlated, Kandori and Matsushima (1998) provide conditions to guarantee strict truth-telling incentives. In addition, following Abreu et al. (1991) that investigated the effect of delayed revelation of public signals, Compte (1998) and Kandori and Matsushima (1998) employ a delayed communication for their two-players' models. The communication is conducted every T periods, and T increases as the players become more patient. The delay aims to increase efficiency by accumulating more information through communication (and thus making statistics-based decisions more accurate). Obara (2009) develops the ideas of delayed communication in games with more than two players when monitoring is almost public.

Fudenberg and Levine (2007) take a different approach, which does not require delayed communication. They obtain conditions under which a perfect public equilibrium is robust to small perturbations. They use a one-period Nash equilibrium for punishment, and thus the resulting set of payoffs supported in equilibrium is restricted to those payoffs that Pareto-dominate a one-period Nash equilibrium payoff.

Games with a mediator. Lehrer (1992a) added a mediator to two-player repeated games with private deterministic signals, and characterized the set of correlated equilibria payoffs. Hillas and Min (2016) generalized Lehrer's result to a model with stochastic signals. Recall that in games with imperfect monitoring, the histories may serve as a (internal) correlation device. However, in the presence of a mediator, this internal correlation plays no role. The correlation is already provided by the external mediation device. Renault and Tomala (2004) generalize Lehrer's result to an arbitrary number of players by adding the assumption that the players can also communicate with the mediator, thus using the solution concept of communication equilibrium (see, Forges (1986)).

In a recent work, Sugaya (2017) characterized a limit set of the communication equilibrium payoffs when the payoffs are random and one's own payoffs are observed by each player. Sugaya uses the randomization produced by a mediator in order to obtain signals with full support. The advantage of the full support is that it enables one to bypass the need to deal with beliefs off-equilibrium. The downside, however, is that due to this randomization, efficient payoffs may be only approximated in equilibrium, while in our model, they are obtained accurately.

Games with specific assumptions regarding the signal structure. Results that assume no form of communication or mediation assume strong assumptions regarding the monitoring structure. For example, Mailath and Morris (2002) obtain conditions for public perfect equilibrium to be robust under small perturbations in monitoring (perturbations that make the monitoring private). They obtain conditions for folk theorem when monitoring is almost perfect and almost public. Hörner and Olszewski (2006) obtain a more general result related to monitoring that is almost public. They obtain a Folk Theorem under a standard dimensionality condition.

Another type of monitoring assumptions is made by Sugaya (2015). He obtains a Folk Theorem for two-player games with no communication while assuming that the signals have a full-support, and that any deviation of a player against any pure action of the opponent changes the distribution of the opponent's signals. In our model, in contrast to this assumption, when a player receives the same payoff upon playing against two different actions of the opponent, the signals he receives might be the same as well.

The monitoring structure examined in this paper is neither almost-public nor almost-perfect. Yet, we show that the payoffs provide information that is detailed enough to allow for supporting all efficient payoffs and having a minmax scheme for cases in which deviations are detected. This is the first main contribution of this paper.

The treatment of beliefs. Players form beliefs regarding their opponents' private history with or without communication. Coordinating these beliefs is especially complicated when the private signals indicate that the game is off the equilibrium path. The equilibrium strategies that are common in the private monitoring literature have two ways with which to avoid dealing with off-equilibrium path beliefs. There are strategies, called *belief-free*, that each player plays optimally following every private history, independently of his beliefs regarding the opponents' private histories (see, for example, Piccione (2002), Ely and Välimäki (2002) and Ely et al. (2005)). Thus, there is no need to specify the beliefs of the players. Typically, the set of equilibrium payoffs that use belief-free strategies is limited, and frequently does not even contain all individually rational efficient payoffs. In fact, Kandori (2011) demonstrates that even strategies conditioned on beliefs regarding the last action alone (strategies called 'weakly belief-free') may improve efficiency over belief-free strategies. When playing *belief-based* strategies, on the other hand, each player plays a best reply to his beliefs. For example, in Bhaskar and Obara (2002) the prisoner's dilemma is analyzed, and the strategies induce two possible states for each player, 'cooperating' and 'deviating'. The entire belief system, in this case, boils down to beliefs about the state of the players. An initial randomization device chooses between the two states, while both states enjoy positive probability along the play. Therefore, any realization of private signals is obtained with a positive probability on the equilibrium path, and there are no off-equilibrium path beliefs. In this case, the efficient payoff, induced by pure strategies, can only be approximated. Other papers assume full-support of private signals following any pure-actions profile or, when full-support of signals is not assumed, as in Sugaya and Wolitzky (2016), the players are instructed to mix their actions in order to retain full-support of private signals. Here, as in results using belief-free strategy, the need to randomize often prevents exact efficient payoffs from being achieved.

Our model assumes a minimal form of communication: a single costly private signal is available to each player. This minimal communication is used only off-equilibrium path. In other words, in equilibrium, the communication channel will never be used. Communication takes place only when a player observes a signal that bluntly reveals a deviation. Such a situation is impossible when the signals always have full support. Our paper adds to the current literature the insight that the private information available to the players is rich enough to enable cooperation when (a) players observe their own payoffs with some positive probability, and this event becomes common knowledge; and (b) a costly communication channel is available. A special case of our model is when own payoffs are observed with probability 1. In this case, both conditions hold, in particular, any strictly efficient strictly individually rational payoff can be obtained as sequential equilibrium payoff.

In order to circumvent dealing with off-equilibrium path beliefs, most of the private monitoring literature either assumes or forces (through proper mixing) full support of the private signals observed, leaving no signal off-equilibrium path. Full-support strategies typically cannot precisely support efficient payoffs, but rather approximate them. In contrast, efficient payoffs in our model are obtained as equilibrium payoffs. Our model involves signals that are off-equilibrium path and the strategies instruct the players what to do after observing them. This is why we cannot shy away from specifying the beliefs held by players off-equilibrium path, following any private history. Moreover, we must show that the players indeed best reply to these beliefs. The challenge becomes even more complicated when simultaneous deviations of both players are considered. In equilibrium deviations are unprofitable and mutual deviations occur indeed with probability zero. However, the notion of sequential equilibrium still requires the existence of a consistent system of beliefs following every history, including after simultaneous deviations. Furthermore, the actions prescribed by each player's strategy must best reply to his respective beliefs. Finding such a system is a difficulty that never rises when signals have full support. Since the information structure explored in this paper typically does not provide a full support of signals, we have to explicitly construct proper beliefs after all histories, no matter how far from the equilibrium path these histories are. This is the second main contribution of this paper.

Relevant economic situations. The main concern of the paper is the ability to sustain cooperation through the ability to effectively sanction an opponent when deviating from this cooperation. Specifically, it is concerned with the ability to differentiate between effective sanctions (minmaxing actions) that might damage the sanctioning player as well as the sanctioned one, and non-effective sanctions (actions that reduce the sanctioned player's payoffs, but are not minmax, meaning, the sanctioned player may recover at least some of his losses when reacting accordingly).

The structure of the paper.

Section 2 presents motivating economic situations for the paper's topic. Section 3 presents an example that demonstrates some of the challenges, and ideas for coping with these challenges. Section 4 details the model and introduces the main result. Section 5 presents the information matrix, a key technical tool employed in the proof process. The proof of the main result appears in Section 6, including the description of the beliefs held by the players. Section 7 concludes with some final comments. All proofs appear in Appendix A, and a formal detailed treatment of off equilibrium path beliefs in Appendix B.

2 Motivating economic situations

2.1 Games with more than one sanction

Consider two rivaling firms. Both want to sustain a cooperative mode of operation. One way to enforce cooperation is by sanctioning a firm that deviates from it. The ability to impose effective sanctions is therefore crucial to support a cooperative mode of operation. The main problem addressed in this paper arises when there exist more than one way to sanction an opponent. One sanctioning action - the minmaxing one - might be more effective than others. This action, however, might be also harmful to the sanctioning player, and he might be reluctant to use it as a result.

When actions are not observed directly, an additional difficulty might arise: using a less effective sanction instead of a more effective one may go undetected, and contribute to the incentives to avoid using more effective sanctions. Using only the less effective sanctions reduces the set of efficient payoffs that can be supported in equilibrium. This is why it is important to introduce incentives for a firm to use the more effective sanctions. The problem is that effective and less effective sanctions might yield the same payoff to the punished player (if played against the best reply to the effective sanction), and thus become indistinguishable from the punished player's point of view. However, the latter may partially recover his loss when playing against the less effective sanction. For this purpose, he needs to be able to distinguish between more and less effective sanctions. This, however, requires the use of actions different from his simple best-reply.

Consider Google competing with Amazon over comparison shopping services. Each company might choose to effectively sanction its opponent, for instance by (secretly) offering significant discounts to some main vendors (i.e., secret price cuts), thus lowering the price for the consumers and increasing traffic. This sanction results in a lower demand for the rival's services, but it also damages the profits of the sanctioning firm. However, the firm may also find other ways to decrease its opponents' revenue, while causing less damage to itself. This can be done for example, by leveraging on other activities of the firm as a way to increase its own traffic (with no price reductions). Google, for instance, was found to promote its comparison shopping services through its search engine[2]. The latter sanction is not as effective, because the sanctioned firm may employ similar strategies to recover some of its losses in case it finds out the reasons for the losses. In case of secret price cuts, there is no way to recover some of the losses.

We model this situation as two-player strategic game. The row player represents Google

[2] This practice was eventually discovered and deemed as a breach of the EU antitrust rules, resulting in 2.42 billion euros fine by the European Commission.

while the column player represents Amazon. The actions available to both are cooperate (co), leverage (lv) and price cuts (pc). When Google employs pc, the minmaxing action, Amazon's profits cannot exceed 0, regardless of Amazon's actions. However, if Google chooses to play lv, Amazon can also play lv and thereby recover some its profits with a payoff of 2. When both play co the payoff is (5,5). However, by deviating to lv, Google increases its payoff while lowering the payoff of Amazon.

In order to prevent Google from deviating, Amazon can threaten to use lv in case of deviation. However, in order to support the entire set of Pareto efficient payoffs, (including those where Amazon obtains payoffs lower than 2) as sequential equilibrium payoffs, the threat should be able to reduce Amazon's payoff below 2. This can be done only by playing pc. However, when only the payoffs are observable, and Google is supposed to play pc, then the best reply of Amazon is to play co. Yet, if Google believes that Amazon is playing co, then Google may deviate to lv, the less effective sanction, without being detected.[3]

We show that when own payoffs are observed, pc can be made a credible threat, therefore enabling a richer set of efficient equilibrium payoffs.

	co	lv	pc
co	5,5	0,8	0,0
lv	8,0	6,2	-1,0
pc	0,0	0,-1	-1,-1

Table 1: Rivaling firms payoff matrix.

2.2 Cournot game

In a repeated Cournot competition[4], the minmaxing action is to produce an output that equates the market clearing price to the marginal cost. Denote this quantity $\bar{q}$. The best reply against that minmax action is to produce nothing. When a firm does not produce, its profits are zero

[3] Another situation in which such a payoff matrix could be applicable is that of when two firms that compete repeatedly (on quality score and price) in auctions for projects. In sealed bid auction the actions are not observed while the outcome (including own payoffs) is. Cooperation in this scenario means tacit collusion. The effective sanction means bidding aggressively, in which case no firm can profit. The less effective sanction could be, for instance, investing in technology or knowledge, thus increasing the quality score. Increased quality score produces a loss to the opponent, as long as the opponent cooperates. However, if the opponent realizes that it is increased quality score that causes the losses, he may invest in increasing the quality score as well, and thus recover some of his losses. Our model enables threat that sustain a richer set of efficient equilibrium payoffs.

[4] We are grateful to an anonymous referee suggesting this example and the following one.

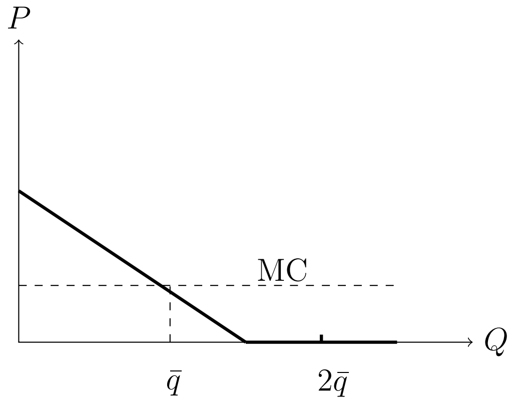

Figure 1: The Cournot game demand curve.

(assuming no fixed cost) regardless of the rivaling firm's actions. Thus, when a firm observes only its own payoffs, there are profitable deviations from the minmaxing action that are not observed by the opponent.

A similar deviation exists when playing the mutual minmax actions, namely both firms produce $\bar{q}$. If the demand curve is as in Figure 1, then when the total amount produced is $2\bar{q}$, the price is zero. Therefore, when one player slightly reduces the amount it produces, it is not going to be detectable by the opponent who observes only its own payoffs.

In a repeated Cournot game with observable payoffs, supporting the entire Pareto efficient frontier requires the kind of construction provided in this paper.

2.3 Games with a costly option to deter the opponent.

Consider a repeated game with positive payoffs, where at each period each opponent may choose to stay out or enter and take an action. Staying out yields a payoff of zero, regardless of the opponent's action. Among the actions available when a player enters there is a costly deterrence action, aimed to make the opponent opt to stay out. This is the minmax action, and the best response to which is to stay out. When a player plays 'out', which always yields a payoff of zero, a deviation of the opponent from playing 'deter' is not observable. Furthermore, when the mutual minmax is being played, a deviation from 'deter' to 'out' is also undetectable.

In the following table A_i stands for the set of available actions beyond 'deter' that are available to Player i, if she decides to enter. The payoffs related to A_i are not specified. Here and in all following examples, Player 1 is the rows player and Player 2 the columns player.

	out	A_2	deter
out	0,0		0,-c
A_1			
deter	-c,0		-c,-c

Table 2: The payoff matrix of a game with ‘deter’ and ‘out’ actions.

Suppose that a player can observe only its own payoffs. When a player is instructed to play the costly ‘deter’ option, a deviation from this action is undetectable. A main goal of this paper is to construct strategies in equilibrium that enable players to use actions like ‘deter’, despite the partial observability of opponents’ actions.

3 An elaborate example: a two-player repeated game with observable payoffs and unobservable actions

The purpose of this section is to demonstrate the main ideas of the paper using a specific game.

Example 1.

Consider the following infinitely repeated two-player game. The possible actions of Player 1 (henceforth, he – the rows player) are Top (T), Middle (M) and Bottom (B), while those of Player 2 (henceforth, she – the columns player) are Left (L), Center (C) and Right (R). After each period, the players privately observe their own payoff, here for simplicity, we assume that this occurs with probability 1. They cannot directly observe the other player’s action nor his or her payoff. In addition, during each period, the players may send a costly message.

	L	C	R
T	3,3	0,4	0,-2
M	-2,-1	4,0	-2,-2
B	5,-1	-2,0	-3,-2

Table 3: The payoff matrix

When Player 2 plays C and obtains a payoff of 0, for instance, she cannot distinguish between Player 1 playing M or B. In this game there are pure minmax actions, B and R. When these actions are played against the respective best-responses, they yield a payoff of 0

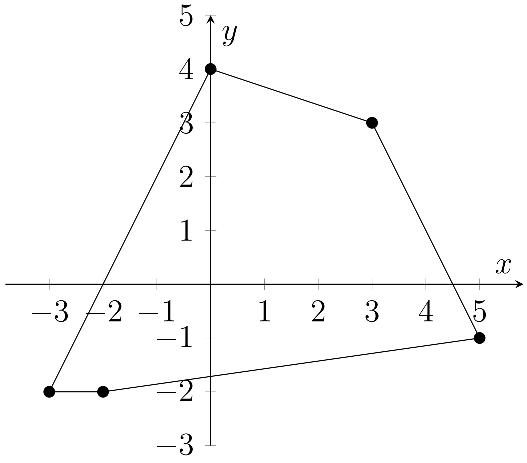

Figure 2: The set of feasible payoffs.

for the player being minmaxed. The efficient payoffs are in the segment connecting $(0, 4)$ and $(3, 3)$ and in the one connecting $(5, -1)$ and $(3, 3)$. The unique one-shot Nash Equilirium is (M, C) with the payoff of $(4, 0)$.

The Folk Theorem states that in an infinitely repeated game under full monitoring, when the players are sufficiently patient, any feasible and individually rational vector of payoffs is an equilibrium payoff. The question we are concerned with is, what payoffs other than the one-period Nash equilibrium can be obtained in equilibrium? For example, can $(3, 3)$ be a sequential equilibrium payoff?

Consider the 'classical' equilibrium structure: the players play a master plan until one player deviates, and then switch to a punishment phase, followed by a continuation game. At first glance, obtaining $(3, 3)$ in equilibrium does not appear too complicated. The pair (T, L) yields the payoff $(3, 3)$ and moreover, any profitable deviation from (T, L) is detectable. For instance, Player 1 can profitably deviate to B, and Player 2 to C, but these deviations are immediately detected. The difficulty in obtaining $(3, 3)$ as an equilibrium payoff is not in detecting deviations. However, the design of the punishment phase, and more importantly, the following continuation game, is not trivial. Designing an effective punishment for games with an information structure where players observe only their own payoffs, is one of two main contributions of this paper.

To better understand the difficulty in constructing punishment, suppose that Player 1 deviated, and that Player 2 should punish him by playing the minmaxing action R. Action R can be played, for example, against Player 1's minmax action, M (mutual minmax, as in Fudenberg and Maskin (1986)). If (M, R) is played, the payoff is $(-2, -2)$. However, both players have profitable deviations that are undetectable: Player 1 can deviated form M to T, and Player 2 from R to L. The mutual minmax is an effective punishment only if deviating from it makes the punishment duration longer. Facing undetectable deviations, this is a chalenge.

Another possibility is that against R Player 1 will play his only best response, T. Now there are no profitable deviations for Player 1. However, knowing that he plays T, Player 2 may profitably deviate to C without him noticing.

If C would also be minmax action, this problem could be solved easily: C could be played instead of R during the punishment of Player 1. In our case, however, C is not a minmax action, meaning that he has at least one action that when played against C yields him a payoff larger than the minmax payoff, 0. In our example, when M is played against C his payoff is 4.

If M would also be a best response to R, that is, if his payoff when (M, R) is played would be zero, we could instruct Player 1 to play M instead of T. In this case, the deviation to C would be immediately detected: he would notice getting 4 instead of 0. Moreover, when such a deviation is detected, it becomes common knowledge, and switching to punishing her, or merely reducing her continuation payoff in another manner, would make it unprofitable.

So far, we obtained that C is Player 2's undetectable deviation, because it gives Player 1 a payoff of 0 when played against T. Furthermore, all his actions that yield a non-zero payoff against C, and could therefore detect the deviation to C, are not best responses to R. Allocating some small probability to playing action C (instead of playing R with probability 1) will keep the payoff of Player 1 at 0 when playing T, and keep his payoff negative when playing M and B. When she minmaxes him, instead of playing R with probability 1, she can allocate some probability to playing C, and if the probability is small enough, it will still minmax him. We can increase the probability of C until the payoff of Player 1 in another row, in our case row M, becomes 0. We end up with the minmaxing mixed action of 1/3 to action C and 2/3 to action R. Both T and M are now Player 1's best replies .

The strategies constructed above are such that Player 2 minmaxes Player 1, Player 1 answers with a strategy composed of best-replies, and Player 2 cannot deviate to profitable actions outside the support of her strategies. Nonetheless, a new issue rises. When implementing this action, Player 2 is asked to randomize between C and R, while her payoff when playing C is higher. For the randomization to be a part of an equilibrium, playing C and R should have different continuation payoffs that would make C and R payoff equivalent in her eyes. For this to happen, he should be able to differentiate, at least with some probability, between periods when she plays C, and periods when she plays R. In this game, he can differentiate between them when playing M, which is also a best response.

During the punishment phase, Player 2 minmaxes Player 1 by playing C with probability 1/3 and R with probability 2/3, and Player 1 plays T with some probability $P(T)$ and M with the probability $P(M) = 1 - P(T)$. During the punishment phase, when Player 1 plays

M he updates her continuation payoff in a way that makes her indifferent between playing C and R. So, for example, following the first period of the punishment phase, if Player 1 played M and observed 4, it means that Player 2 played C, which yields her the expected payoff of $4 - 4P(M)$. When she plays R, however, she gets -2. The difference between these payoffs is $(1-\delta)[4P(M)-6]$, where δ denotes the (common) discount factor. This punishment phase lasts N periods and is followed by a continuation game. The continuation game begins with two periods of communication. In these periods a one period Nash equilibrium is played and messages are transferred. The change in the continuation payoffs that balances this difference happens only when Player 1 plays M and is able differentiate between Player 2's actions, hence it is divided by $P(M)$. In addition, the continuation payoff is realized in $N+2$ periods after the current one. Therefore the difference between the continuation payoffs that follow C and R should be $\frac{(1-\delta)[4P(M)-6]}{\delta^{N+2}P(M)}$.

The continuation payoff established during the punishment phase is obtained by playing one of two pre-specified strategies, one that supports Player 2's high payoff and one that supports a low payoff. Both strategies yield the same payoff for Player 1, and thus he is indifferent between the two. Which of the two strategies will be played is determined by a randomization solely controlled by Player 1. He randomizes between high and low payoffs in a way that induces the desired continuation payoff.

The First period of the communication phase is dedicated to a communication regarding the identity of the punished player. The need for this additional period of communication is explained in Section 6.1.3. The second communication period is used for conveying the message regarding the continuation payoff. In order to communicate the continuation payoff, only two signals are needed: one for the high and the other for the low continuation payoff.

We use Example 1 to clarify one more point. The case where Player 2 deviates is treated in an analogous way: two different continuation payoffs for Player 1 are needed, with identical payoffs for Player 2. This implies that in order to implement all possible continuation payoffs, a two-dimensional set of payoffs should be available. Specifically, an internal payoff in the set of feasible payoffs needs to be supported by an equilibrium.

A careful look at the payoff matrix reveals that from any pair of actions yielding a payoff internal to the set of feasible individually rational payoff, there is a profitable deviation that is not immediately detected. For example, from (B, R) Player 1 can deviate to M, and Player 2 will not notice, since she gets a payoff of -2 in both cases. In our example, the efficient frontier of the set of payoffs is not a straight line, and so a subset of the convex combinations of the payoffs of the strictly efficient frontier $(0, 4), (3, 3)$ and $(5, -1)$ is internal to the set of payoffs. Supporting a subset of these combinations of payoffs by a sequential equilibrium is done in the same way $(3, 3)$ is supported: at each period one of the pairs (T, C), (T, L) or (B, L)

is played. Any profitable deviation from these actions is immediately detected, and is followed by a punishment phase. Some convex combinations of $(4,0)$ and payoffs on the strictly Pareto efficient frontier are also internal to the set of feasible individually rational payoffs and can be obtained in a similar way. However, we cannot rely on having such internal payoffs available in the general case, so we describe another way to produce internal points as sequential equilibrium payoffs.

Example 1 demonstrates some of the problems of designing the punishment scheme, as well as a general direction of how to solve them.

4 The Model and Main Result

4.1 The model

We study two-player repeated game with imperfect monitoring. After each stage the players obtain a stochastic signal whose distribution depends on the pair of actions played. We assume that each player observes his payoff with a positive probability and that when he observes it, this event is common knowledge. In addition, we assume that players may communicate with each other using costly messages.

The base game[5]**:** The base game is defined by the following items.

- **Action sets:** Each player $i = 1, 2$ has a finite pure actions set A_i. Denote $A := A_1 \times A_2$ the set of pure action profiles.

- **Utility functions:** When the action profile $(a_1, a_2) \in A$ is played Player i obtains the payoff $U_i(a_1, a_2)$. For the sake of convenience we extend the domain of U_i in a linear fashion as follows. For every $(\lambda, \lambda') \in \mathbb{R}^{|A_1| \times |A_2|}$, define $U_i(\lambda, \lambda') = \Sigma_{a_j \in A_1} \Sigma_{b_k \in A_2} \lambda_j \lambda'_k U_i(a_j, b_k)$. Denote by $\Delta(A_i)$ the set of player i's mixed actions. Thus, when the players play the pair $(p, q) \in \Delta(A_1) \times \Delta(A_2)$, $U_i(p, q)$ is the expected payoff of Player i.

- **Monitoring:** let Θ_i be the set of possible signals of Player i. When the action profile $(a_1, a_2) \in A$ is played, a pair of signals $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ are randomized according to the (joint) distribution of the random variables $(\psi_1(a_1, a_2), \psi_2(a_1, a_2))$. The random variables satisfy that for any $(a_1, a_2) \in A$ there is a positive probability denoted by $I_i(a_1, a_2)$ that player i observes his own payoff.

  Formally, $\forall (a_1, a_2) \in A$, $\exists (\theta_1(a_1, a_2), \theta_2(a_1, a_2)) \in \Theta_1 \times \Theta_2$ such that $P(\psi_i(a_1, a_2) = \theta_i(a_1, a_2)) = I_i$ and $P(\psi_i(a'_1, a'_2) = \theta_i(a_1, a_2)) = 0$, $\forall (a'_1, a'_2) \in A$ such that $u_i(a'_1, a'_2) \neq u_i(a_1, a_2)$.

[5] We call the base game also one-shot game or stage game.

Moreover, the fact that Player i knows his own payoff is common knowledge (see comment on monitoring below).

Formally, denote $\Upsilon_i(a_1, a_2)$ the event that Player i knows his payoff when action profile $(a_1, a_2) \in A$ is played, that is, $\Upsilon_i(a_1, a_2) = \{\psi_i(a_1, a_2) = \theta_i(a_1, a_2)\}$. Then there exists $\theta_j \in \Theta_j$ such that $P(\psi_j(a_1, b_1) = \theta_j | \Upsilon_i(a_1, a_2)) = 1$, and $P(\Upsilon_i(a_1, a_2) | \psi_j(a_1, b_1) = \theta_j) = 1$.

- **Communication:** Each player has the possibility to convey a message to his opponent during each period. Player i can either send a signal, or keep silence. The set of available signals for player i is $\Xi = \{\xi, \phi\}$, where ξ is a costly signal and ϕ represents the option of conveying no signal. The cost of conveying a message for player i is $c_i > 0$.

The repeated game: In the repeated game, the private history of Player i at the end of period t is an element of $H_i^t = (A_i \times \Theta_i \times \Xi^2)^t$. For the formal definitions of strategies, assessments and sequential equilibrium the reader is referred to Maschler et al. (2013).

At period t, when the pair of actions played is $(a_1, a_2) \in A$, Player i receives a random signal $\psi_i^t(a_1, a_2)$, where $(\psi_1^t(a_1, a_2), \psi_2^t(a_1, a_2))$ has the same distribution as $(\psi_1(a_1, a_2), \psi_2(a_1, a_2))$. Finally, $(\psi_1(a_1, a_2), \psi_2(a_1, a_2))$, $t = 1, 2, 3...$ and $(a_1, a_2) \in A$ are all independent.

A comment on monitoring: The monitoring item seems to require a little elaboration. Suppose, for instance, that with probability $\varepsilon > 0$ both players observe their own payoffs and with probability $1 - \varepsilon$ they get no information. In this case, when the players get to observe their own payoffs, this event becomes common knowledge. In particular, when a player receives information about his own payoff, he knows that his opponent observes her own payoff as well, and that she knows that he observes his own, etc.

To further explain, suppose that the payoff matrix is as in table 4. Suppose also that the signal that the players observe regarding own payoff is noisy in the following way: a player observes the sum of his own payoff and a noise that is either -1, 0 or 1 (with some known distribution over the noise). Then Player 2 always knows her payoff, and Player 1 knows his payoff when playing B. However, when Player 1 plays T, he knows his own payoff only when he observes a signal in the set $\{-1, 0, 2, 3\}$. The monitoring condition is fulfilled if when a signal from that set that was observed, it is common knowledge (that a signal from that set is observed). However, when observing a signal of 1, Player 1 cannot be certain about her payoff, which can be either 0 or 2, and observing such a signal needs not be common knowledge.

	L	R
T	0,1	2,0
B	3,1	-5,5

Table 4: A payoff matrix

4.2 The main result

Let F denote the convex hull of the set of feasible payoffs, and let EF denote the set of feasible payoffs that are strictly Pareto efficient. That is[6],

$$EF = \left\{(u_1, u_2) \in F;\ \text{for any } (u'_1, u'_2) \in F,\ u'_i > u_i \text{ implies } u'_{-i} < u_{-i}\right\}.$$

Let m_i be the minmax payoff of Player i.

$$m_i = \min_{q_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} U_i(a_i, q_{-i}).$$

Without loss of generality assume $m_i = 0, i = 1, 2$. A payoff is *strictly individually rational* if it is positive. Denote the set of positive payoffs of player i by IR_i, and let $IR = IR_1 \cap IR_2$. IR is the strictly positive orthant of $\mathbb{R}^2$.

The game is played repeatedly, and the stream of payoffs is evaluated using a common discount factor $0 < \delta < 1$: the total payoff of player i is $(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} U_i(a_1^t, a_2^t)$, where a_i^t denotes the action played by Player i at period t. The main result of the paper is that all strictly efficient payoffs that are strictly individually rational, and moreover, any convex combination of these payoffs and one-period Nash equilibria payoffs are sequential equilibrium payoffs when the players are sufficiently patient. Formally,

Theorem 1. *For every payoff* (u_1, u_2) *of the two following types:*

(a) (u_1, u_2) *is strictly efficient and strictly individually rational payoffs:* $\{(u_1, u_2) \in EF \cap IR\}$*;*

(b) (u_1, u_2) *is not efficient, but is in the convex hull of one-period Nash equilibrium payoffs and the payoffs of (a) above,*

there exists $0 < \delta' < 1$ *such that for every* $\delta > \delta'$, (u_1, u_2) *is a sequential equilibrium payoff when the discount factor is* δ*.*

In Nash-threat folk theorems, the set of payoffs achieved in equilibrium is the set of feasible payoffs that Pareto dominate a one-period Nash equilibrium. The set of payoffs described in

[6] We denote $-i = 3 - i$.

item (b) does not coincide with this set, it may include payoffs where one player obtains less than his one-period Nash equilibrium payoff.

For the sake of simplicity we prove the theorem assuming that each player observes his own payoff with probability 1. In Section 6.5 we explain why the proof, with minor modifications, actually shows the theorem with the general setup described above, where payoffs are observed only with some positive probability.

5 The Information Matrix

In this section we explore the structure of the information available to the players. To simplify the discussion, we consider only the information available to Player 1 (that of Player 2 is analogous). During each period the information Player 1 obtains consists of his own action and his own payoff.

For every pair of sets $A'_1 \times A'_2 \subseteq A_1 \times A_2$, we define a matrix $M(A'_1, A'_2)$ that consists of 0's and 1's. The number of the rows in $M(A'_1, A'_2)$ is the number of different action-payoff pairs (i.e, pairs of the type (pure action, realized payoff)) that are possible for Player 1 when using actions in A'_1 against any full-support distribution over A_2. The number of columns is the number of actions in A'_2. The cell of $M(A'_1, A'_2)$ that corresponds to the row $(a_1, U_1(a_1, a_2))$ and the column a_2 is 1, and otherwise, is 0. This matrix maps the ways by which the information of Player 1 depends on his and his opponent's actions. We refer to this matrix as *the information matrix* of Player 1 corresponding to (A'_1, A'_2). The matrix $M(A'_1, A'_2)$ is used to design the punishment phase, as well as future continuation payoffs of the punishing player.

For $\lambda \in \mathbb{R}^n$, let $\text{supp}(\lambda)$ be the set $\{i;\ \lambda_i \neq 0\}$. For any pair of mixed actions $(p, q) \in \Delta(A_1) \times \Delta(A_1)$ we replace $\text{supp}(p)$ by p and $\text{supp}(q)$ by q. For instance, $M(p, q) = M(\text{supp}(p), \text{supp}(q))$ and $M(A'_1, q) = M(A'_1, \text{supp}(q))$. Also denote the vector space spanned by the columns of $M(A'_1, A'_2)$ by $V(A'_1, A'_2)$.

5.1 Example 1 revisited: The information matrix

Consider the game described in Table 1. We illustrate the case where $A'_1 = A_1$ and $A'_2 = A_2$ in Table 5.

The pair $(T, 3)$, for instance, stands for playing T and receiving the payoff 3. Since this happens only when Player 2 plays L, the respective entry is 1, and it does not happen when Player 2 plays C or R, where the respective entries are 0. Note that in each row there is at least one '1' and the number of 1's in each column is equal to the number of actions in A_1.

	L	C	R
(T,3)	1	0	0
(T,0)	0	1	1
(M,-2)	1	0	1
(M,4)	0	1	0
(B,5)	1	0	0
(B,-2)	0	1	0
(B,-3)	0	0	1

Table 5: The information matrix $M(A_1, A_2)$.

5.2 Possible deviations

Suppose that q^* is the action by which Player 2 is minmaxing Player 1. Let A_1' be the set of pure best replies to q^*. Player 2 might have the following three types of deviations from q^*:

(i) A deviation to a (possibly mixed) action q' such that $\text{supp}(q') \nsubseteq \text{supp}(q^*)$ but $V(A_1', q') \subseteq V(A_1', q^*)$. Such deviations are made unprofitable by proper continuation payoffs that are constantly updated by Player 1, based on his information matrix.

(ii) A deviation to a (possibly mixed) action q' such that $\text{supp}(q') \nsubseteq \text{supp}(q^*)$, and $V(A_1', q') \nsubseteq V(A_1', q^*)$. These deviations are made unprofitable by using the continuation payoffs designed to make Player 2 indifferent between all actions in the support of q^*.

(iii) A deviation to a (possibly mixed) action q' such that $\text{supp}(q') \subseteq \text{supp}(q^*)$. Designing a minmaxing strategy such that there does not exist any strictly profitable deviation of this kind is the content of the Lemma 1.

In what follows we find a minmax action q^* (of Player 2) that enables one to design continuation payoffs that take care of deviations of types (i) and (ii), and is immunized against deviations of type (iii).

5.3 Finding a minmax action

The next lemma uses the information matrix. It guarantees that there is a minmax action, q^*, immunized against deviation of types (i) and (iii). (A deviation of type (ii) is handled later, in Lemma 2.) It requires a few pieces of notations. For $\lambda \in \mathbb{R}^{\text{supp}(q)}$ denote by $M(p, q)\lambda$ the product of the matrix $M(p, q)$ and the vector λ. The vector λ is in $\mathbb{R}^{|A_2|}$. It represents the difference between two mixed actions and it typically contains negative coordinates. Denote Δ_2 the set of all $q \in \Delta(A_2)$ such that $U_1(p, q) \leqslant 0,\ \forall p$. Since $m_1 = 0$, Δ_2 is not empty.

Lemma 1. *There exists a pair of distributions* $(p^*, q^*) \in \Delta(A_1) \times \Delta_2$ *such that*

(1) $U_1(p^*, q^*) = 0$*,*

(2) $U_1(a_1, q^*) = 0 \Rightarrow a_1 \in \mathrm{supp}(p^*)$*,*

(3) $\forall q' \in \Delta_2$*,* $M(p^*, A_2)q' = M(p^*, A_2)q^* \Rightarrow U_2(p^*, q') \leqslant U_2(p^*, q^*)$*,*

(4) $\forall q' \in \Delta(A_2)$ *and* $\lambda \in \mathbb{R}^{|A_2|}$*, if* $\mathrm{supp}(q'), \mathrm{supp}(\lambda) \subseteq \mathrm{supp}(q^*)$*, then*

$$M(p^*, A_2)q' = M(p^*, A_2)\lambda \Rightarrow U_2(p^*, q') = U_2(p^*, \lambda).$$

(5) $\forall q' \in \Delta(A_2)$*, and* $\lambda \in \mathbb{R}^{|A_2|}$*, if* $\mathrm{supp}(\lambda) \subseteq \mathrm{supp}(q^*)$ *and* $M(p^*, A_2)q' = M(p^*, A_2)\lambda$*, then* $U_2(p^*, q') \leqslant U_2(p^*, \lambda)$*.*

The proofs of all the lemmas including this one appear in the Appendix.

The mixed action q^* is, by requiring $q^* \in \Delta_2$, a minmax action, and by (1) and (2) of Lemma 1, p^* is a full support distribution over Player 1's best replies. When the players play the pair (p^*, q^*), the expected frequency of the signals obtained by Player 1 is in the space spanned (in the algebraic sense) by the columns of $M(p^*, q^*)$. From (3) there is no other minmax action, as profitable for Player 2 as q^*, that induces the same distribution of signals when played against p^*. From (4), for any linear combination λ and mixed action q' that induce the same signals (i.e., $M(p^*, A_2)q' = M(p^*, A_2)\lambda$) and whose support is included in $\mathrm{supp}(q^*)$, yield the same payoff for Player 2 (i.e., $U_2(p^*, q') = U_2(p^*, \lambda)$). This property renders any deviation of type (iii) (see above) unprofitable.

Let λ be a linear combination and q' be a mixed action whose support is not included in $\mathrm{supp}(q^*)$. Suppose that λ and q' induce the same signals (i.e., $M(p^*, A_2)q' = M(p^*, A_2)\lambda$). Then, by (5), q' yields a payoff for Player 2 that does not exceed that defined by λ (i.e., $U_2(p^*, q') \leqslant U_2(p^*, \lambda)$). This renders any deviation of type (i) unprofitable.

5.4 The use of the information matrix related to Example 1.

It turns out that $q^* = (0, \frac{1}{3}, \frac{2}{3})$ satisfies the properties of the previous lemma. As for the action of Player 1, p^* may be any full support distribution over the best replies: $(P(T), 1 - P(T), 0)$, $0 < P(T) < 1$. For the sake of simplicity, let p^* be $(\frac{1}{2}, \frac{1}{2}, 0)$. Consider the information matrix $M(p^*, A_2)$, described in Table 6.

The purpose of the information matrix is to construct the continuation payoff after the punishment phase. Player 1 updates the continuation payoffs of Player 2 based on the signals

	L	C	R
(T,3)	1	0	0
(T,0)	0	1	1
(M,-2)	1	0	1
(M,4)	0	1	0

Table 6: $M(p^*, A_2)$ - the information matrix when Player 1 plays p^*.

he observes during the punishment phase. At each period of the punishment his information could be one of the four pairs represented by the four rows of the matrix. For the sake of example let us enumerate these rows as $s = 1, ..., 4$. Thus, $s = 1$ stands for the signal $(T, 3)$, while $s = 2$ for $(T, 0)$, etc. The punishment phase is meant to last N stages and is followed by a continuation game that begins with two communication periods. After these $N + 2$ stages the game continues with a continuation that depends on what happened during these stages.

We define $4N$ variables X_s^t, $s = 1, ..., 4; t = 0, ..., N - 1$. When Player 1 observes the s-th signal at time t ($t = 0, ..., N - 1$) of the punishment phase, he adds the amount X_s^t to the continuation payoff of Player 2.

More specifically, suppose that Player 2 plays C at $t = 0$ (of the punishment phase). Player 1 plays T with probability $\frac{1}{2}$ and observes $(T, 0)$ (recall, he plays $(\frac{1}{2}, \frac{1}{2}, 0)$). In this case X_2^0 is added to the continuation payoff of Player 2. Likewise, with probability $\frac{1}{2}$ Player 1 plays M, observes $(M, 4)$ and X_4^0 is added to the continuation payoff. We obtain that when Player 2 plays C her expected payoff, taking into account the current payoff and the future expected change of continuation payoff (to be realized $N + 2$ periods in the future), is

$$(1 - \delta)\left[\frac{1}{2}4 + \frac{1}{2}0\right] + \delta^{N+2}\left[\frac{1}{2}X_2^0 + \frac{1}{2}X_4^0\right] \tag{1}$$

and when playing R her expected payoff is,

$$(1 - \delta)(-2) + \delta^{N+2}\left[\frac{1}{2}X_2^0 + \frac{1}{2}X_3^0\right]. \tag{2}$$

However, if Player 2 deviates and play L, her expected payoff is,

$$(1 - \delta)\left[\frac{1}{2}3 + \frac{1}{2}(-1)\right] + \delta^{N+2}\left[\frac{1}{2}X_1^0 + \frac{1}{2}X_3^0\right]. \tag{3}$$

Recall that Player 2 is prescribed to play $q^* = (0, \frac{1}{3}, \frac{2}{3})$. In order to make her indifferent between C and R and weakly preferring them over L, the figures in Eqs.(1) and (2) should coincide and they should be at least that of Eq.(3).

More generally, the variables X_r^t should solve the following system of linear equations, where $M_s(p^*, a_2)$ be the sth entry of the information column $M(p^*, a_2)$:

$$\begin{aligned} U_2(p^*, C)(1-\delta)\delta^t + \delta^{N+2}\sum_{s=1}^{4}\frac{1}{2}M_s(p^*, C)X_s^t &= U_2(p^*, R)(1-\delta)\delta^t + \delta^{N+2}\sum_{s=1}^{4}\frac{1}{2}M_s(p^*, R)X_s^t \\ U_2(p^*, L)(1-\delta)\delta^t + \delta^{N+2}\sum_{s=1}^{4}\frac{1}{2}M_s(p^*, L)X_s^t &\leqslant U_2(p^*, C)(1-\delta)\delta^t + \delta^{N+2}\sum_{s=1}^{4}\frac{1}{2}M_s(p^*, C)X_s^t, \\ & \quad t = 0, ..., N-1. \end{aligned}$$

It turns out that a certain condition related to the independence of columns of the information matrix (detailed in Lemma 6 in the appendix) enables one to solve this system, as stated in Lemma 2.

5.5 The general case

We turn to the design of the continuation payoffs in the general case. Let S be the number of rows in the information matrix $M(p^*, A_2)$. Let $a(s)$ be the action in A_1 associated with the s-th row of the information matrix $M(p^*, A_2)$. For every s ($s = 1, ..., S$), let X_s^t be the variable corresponding to the case where Player 1 plays $a(s)$ at time t and receives the signal corresponding to the s-th row. We refer to X_s^t as *stage-accumulated* payoffs. The signal obtained by Player 1 at time t when playing $a(s)$ will affect the continuation payoff of player 2 through X_s^t. Specifically, suppose that Player 2 plays a_2 at time t, $t = 0, ..., N-1$. Then, the expected changes in her continuation payoffs (evaluated from the beginning of the punishment phase) is $\delta^{N+2}[\sum_{s=1}^{S} p^*(a(s))M_s(p^*, a_2)X_s^t]$. This continuation payoff of Player 2 should make her indifferent between all the actions in the support of q^*: her stage-payoff plus the resulting continuation payoff should be constant across supp(q^*). That is,

$$(1-\delta)\delta^t U_2(p^*, a_2) + \delta^{N+2}[\sum_{s=1}^{S} p^*(a(s))M_s(p^*, a_2)X_s^t]$$

has to be constant across all $a_2 \in$ supp(q^*). Furthermore, actions out of supp(q^*) must be less profitable than those in supp(q^*). Lemma 2 guarantees the existence of continuation payoffs that satisfy these requirements.

There is, though, another consideration that should be taken into account. The stage-accumulated payoffs are accumulated along the punishment phase. This accumulation should not be too large: it must fit within a certain range of possible continuation payoffs. This is the purpose of the last inequality in Lemma 2. We will elaborate more about it in Section 6.1.2.

Let Q^* be a subset of A_2 such that $V(p^*, Q^*) = V(p^*, q^*)$, and the columns of $M(p^*, Q^*)$ are independent.

Lemma 2. *For any $e \in (0,1)$, $D > 0$ and an integer N, there exists $\delta_2 \in (0,1)$ such that for any $\delta > \delta_2$ there are numbers $K^t \in \mathbb{R}$, $t = 0,...,N-1$, and a solution to the following system of linear inequalities (with X_s^t being the unknowns),*

$$\begin{cases} U_2(p^*, a_2)(1-\delta)\delta^t + e \sum_{s=1}^{S} p^*(a(s)) M_s(p^*, a_2) X_s^t = K^t & a_2 \in Q^*, t = 0,...,N-1 \\ \\ U_2(p^*, a_2)(1-\delta)\delta^t + e \sum_{s=1}^{S} p^*(a(s)) M_s(p^*, a_2) X_s^t \leqslant K^t & a_2 \notin Q^*, t = 0,...,N-1 \end{cases}$$

$$\sum_{t=0}^{N-1} \min_s \left\{X_s^t\right\} = 0,$$

and

$$\sum_{t=0}^{N-1} \max_s \left\{X_s^t\right\} \leqslant D.$$

The exact values of e and D will be determined in Section 6.1.2. Note that a deviation of type (ii) (see Section 5.2) is to an action whose support is not a subset of Q^*. Due to the inequalities in the second row of the linear system a deviation of this type is unprofitable.

6 Proof of the Main Result

The next lemma is essential for defining the equilibrium strategies. It guarantees that suitable continuation payoffs always exist, within a close proximity to the target payoff. Technically, it ensures that for any efficient payoff (u_1, u_2) there is an ε-environment so that any efficient payoff (v_1', v_2') in this environment has a rectangle of payoffs that satisfies (a) (v_1', v_2') Pareto dominates any payoff in the rectangle; (b) for any (v_1, v_2) in the rectangle there is a profile of strategies supporting it such that any profitable deviation is detected, and all future payoffs are always within an ε distance from (v_1, v_2).

Condition (b) implies, in particular, that (u_1, u_2) itself can be obtained using strategies such that any profitable deviation is detectable and any future payoff is within an ε distance from the target payoff (u_1, u_2). These strategies will be used in the play path, while the strategies leading to other payoffs in the rectangle will be used as continuation payoffs.

Let $\mathcal{B}((u_1, u_2), \varepsilon)$ be the ball of radius $\varepsilon > 0$ around (u_1, u_2).[7] For a strategy profile $\sigma = (\sigma_1, \sigma_2)$ define $U_i(\sigma_1, \sigma_2, \delta, h(t))$ to be the future payoff of Player i following a history $h(t)$ of t periods that have positive probability under σ and when the discount factor is δ.

[7] That is, $\mathcal{B}((u_1, u_2), \varepsilon) = \{(v_1, v_2);\ (v_1 - u_1)^2 + (v_2 - u_2)^2 \leqslant \varepsilon^2\}$.

Lemma 3. *For every $(u_1, u_2) \in EF$ and $\varepsilon > 0$, there exist $D_1, D_2 > 0$ and $\delta_3 \in (0,1)$, such that for any $(v'_1, v'_2) \in EF \cap \mathcal{B}((u_1, u_2), \varepsilon)$, $1 > \delta > \delta_3$ and any payoff (v_1, v_2) in the convex hull of $\{(v'_1, v'_2), (v'_1 - D_1, v'_2), (v'_1, v'_2 - D_2), (v'_1 - D_1, v'_2 - D_2)\}$ there exists a strategy profile $\sigma = (\sigma_1, \sigma_2)$ such that:*

(i) the payoff of the strategy profile (σ_1, σ_2) in the repeated game with discount δ is (v_1, v_2);

(ii) any profitable deviation from σ_i is detected with probability 1;

(iii) for every time t and a history $h(t)$ that has a positive probability under σ, $U_i(\sigma_1, \sigma_2, \delta, h(t)) \in \mathcal{B}((v_1, v_2), \varepsilon)$.

Note that by this lemma, for a given ε the same δ_3 applies for the entire convex hull mentioned in the lemma.

In order to illustrate the lemma we revisit Example 1. Suppose that the target payoff is $(4, 1)$. Figure 3 sketches the rectangles guaranteed by Lemma 3. On the play path that supports $(4, 1)$ all continuation payoffs are efficient and are located in $\mathcal{B}((4, 1), \varepsilon)$. Each such continuation payoff Pareto dominates a rectangle contained in the convex hull of payoffs that are strictly efficient, and the same Pareto efficient minus the communication costs.

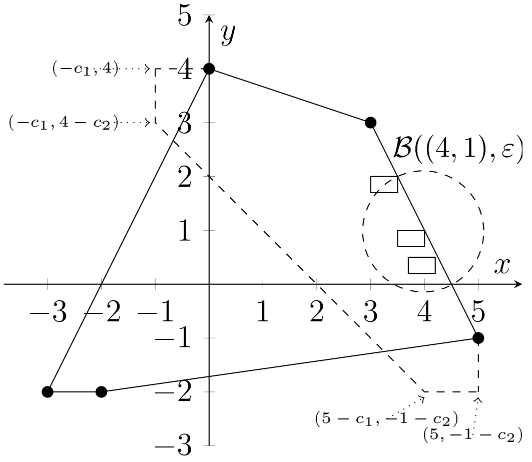

Figure 3: The rectangles referred to in Lemma 3 applied to Example 1.

We describe the equilibrium strategies that yield the payoffs of type (a) of Theorem 1. As for type (b), one may employ a standard convexifying argument. Indeed, suppose the target payoff, $u = (u_1, u_2)$, is a convex combination of a payoff on the strictly efficient frontier, $v^{EF} = (v_1^{EF}, v_2^{EF})$ and a one-period Nash equilibrium payoff, $v^{NE} = (v_1^{NE}, v_2^{NE})$. The payoff on the strictly efficient frontier is a convex combination of (at most) two pure actions' payoffs, both efficient, let them be u^{EF} and $\tilde{u}^{EF}$. There exist $\alpha, \tilde{\alpha}$, $(0 \leqslant \alpha, \tilde{\alpha}, 1-\alpha-\tilde{\alpha} \leqslant 1)$, such that $u = \alpha u^{EF} + \tilde{\alpha}\tilde{u}^{EF} + (1-\alpha-\tilde{\alpha})v^{NE}$. Then, by Sorin (1986), there is an infinite sequence of v^{NE}, u^{EF} and $\tilde{u}^{EF}$, whose discounted sum equals u. Moreover, the continuation payoffs, for sufficiently large discount factor, are always close to the target payoff. The players can switch between the actions according to the sequence, until a deviation from the strategies is observed (deviations from the one-period Nash equilibrium are, by definition, not profitable). If and when this happens, the players switch to playing the punishment phase.

6.1 The phases of the equilibrium strategies

The equilibrium strategies have three phases:

- The play path
- The punishment phase
- The continuation game

The players follow the play path forever or until a deviation is detected. If and when a deviation is detected, they switch to the punishment phase, followed by a continuation game until another deviation is detected, and if so, they play a punishment phase again, etc.

We describe in details the instructions to the players during the different phases. Let the payoff on the strictly efficient frontier we wish to support be (u_1, u_2).

6.1.1 The play path

The play path strategy coincides with the strategy profile (σ_1, σ_2) whose existence is guaranteed by Lemma 3. When (u_1, u_2) is strictly individually rational, from (iii) of the lemma so are all continuation payoffs. From (ii) any profitable deviation is both detected and common knowledge. Then, after a profitable deviation[8], the players know when the punishment phase begins.

6.1.2 The punishment phase

In case a deviation has been detected, this detection is common knowledge, and the players move to the punishment phase. In what follows, we assume that the deviating player is Player

[8] A non-profitable deviation may go unnoticed. Being non-profitable to begin with, the possible lack of punishment does not affect the incentives to follow the equilibrium path. Beliefs following such a deviation are detailed in Appendix B.

1. The case where Player 2 deviated is treated in a similar way. The case of simultaneous deviations is elaborated in Section 6.3. The punishment phase consists of N periods (N is to be determined in Lemma 4) while Player 1 and Player 2 play constantly the distributions p^* and q^* whose existence is guaranteed by Lemma 1.

During the punishment, Player 1 updates the continuation payoffs of Player 2. The stage-accumulated payoff, which depends on the actions taken at the t-th period of the punishment phase is X_s^t, defined in Lemma 2. Lemma 2 is employed with $e = \delta^{N+2}$ and $D = D_2$ (determined in Lemma 3), so that Player 2 has the proper incentives to follow the punishment randomizations, and the stage-accumulated payoffs fit within the range of available continuation payoffs included in the rectangle described in Lemma 3.

6.1.3 The continuation game

Fix one stage-game Nash equilibrium actions pair $\big(p^{NE}, q^{NE}\big)$, and denote $U(p^{NE}, q^{NE}) = (v_1^{NE}, v_2^{NE})$. The continuation game begins with two periods of communication. During these periods (p^{NE}, q^{NE}) is played, while the punished player, here Player 1, conveys messages.

At the first period, the punished player conveys a message stating that he was the punished player. This message is redundant in scenarios where a single player deviated, but is needed in the case of simultaneous deviations. If both players simultaneously deviate, their beliefs about the identity of the player being punished may not agree. This message is aimed at synchronizing the beliefs of the players. We elaborate more on this issue in Section 6.3.

At the second period the continuation payoff is communicated. The continuation of the game is determined by the public messages conveyed during both communication periods and depends on whether the punishment that has just ended followed a deviation from the play path or from a continuation play. Let (v_1', v_2') be the continuation payoff that would prevail without deviation at the time a deviation was detected the first time (i.e., a deviation from the play path). We divide the discussion regarding the first period of communication into two possible situations.

Situation 1: Continuation that follows a punishment due to a deviation from the play path

There are two cases to consider:

1. **Both players or none conveyed messages during the first period of the continuation game.**

   In case at the first communication period both players conveyed a message stating that they were punished, or if no player conveyed a message, no messages are needed during the second period. The continuation payoff in this case is $(v_1' - D_1, v_2' - D_2)$.

2. **A single player conveyed a message during the first period of the continuation game.**

   In case a single player, say Player 1, conveyed a message claiming that he was the one being punished, in the second communication period Player 1 conveys a message regarding the continuation payoffs. Player 1 randomizes between two possible continuation payoffs: the high, (v_1', v_2'), and the low, $(v_1' - \frac{1-\delta}{\delta}c_1, v_2' - D_2)$, both within the rectangle guaranteed by Lemma 3. A message sent in the second period indicates that the continuation payoff is the high payoff, while no message indicates that it is the low one. Note that Player 1 is indifferent between these payoffs, when accounting for the message cost.

   As for Player 2, the expected continuation payoff of Player 2 reflects the stage-accumulated payoffs and should be $v_2' - D_2 + \sum_{t=0}^{N-1} X_s^t$. This payoff is generated by randomizing with probability $\frac{\sum_{t=0}^{N-1} X_s^t}{D_2}$ for the payoff (v_1', v_2') and with the complement probability for $(v_1' - \frac{1-\delta}{\delta}c_1, v_2' - D_2)$. The reason is that $\left(\frac{\sum_{t=0}^{N-1} X_s^t}{D_2}\right) v_2' + \left(1 - \frac{\sum_{t=0}^{N-1} X_s^t}{D_2}\right)(v_2' - D_2) = v_2' - D_2 + \sum_{t=0}^{N-1} X_s^t$.

Situation 2: Continuation that follows a punishment for a deviation from the continuation play

Here as well, we use the same rectangle guaranteed by Lemma 3 and determined by (v_1', v_2') (recall, this is the the continuation payoff at the first time the deviation has been detected). However, the target payoff in situation 2 might be lower than the original equilibrium payoff, and the messages costs reduce it further. Some small adjustments, compared to situation 1, are needed in order to guarantee that the continuation payoff is always within the original rectangle. In case both players or non conveyed messages during the first period of the continuation game, the same rule holds: no messages are needed during the second period, after which the continuation payoff is $(v_1' - D_1, v_2' - D_2)$. Otherwise, assume that Player 1 was the only one conveying a message stating that he was the punished player. Let Player 1's continuation payoff at the time of the last deviation be u_1'. Player 2's post punishment payoff depends on the signal conveyed by Player 1. Player 1's post punishment payoff, denoted w_1', is defined as follows:

$$w_1' = \begin{cases} v_1' - D_1, & \text{if } u_1' < v_1' - D_1 \\ u_1', & \text{if } v_1' - D_1 \leqslant u_1' \leqslant v_1' \\ v_1' + D_1, & \text{if } u_1' > v_1' + D_1. \end{cases} \tag{4}$$

Among the three possible continuation payoffs, the first involves an increase in the continuation payoff, compared u_1'. The increase should be small enough to keep the total reduction of payoff (due to the punishment and the continuation game) large enough to cancel out a gain from a single period deviation. In this case, the post punishment payoff is higher than the future payoff at the time of deviation detection in order to remain within the rectangle mentioned in Lemma 3. All continuation payoffs are always within an ε distance from the target payoff, therefore at most ε is added to the future payoff.

In case $v'_1 - D_1 \leqslant w'_1 \leqslant v'_1 - \frac{D_1}{2}$, the two payoffs $(w'_1 + \frac{1-\delta^2}{\delta^2}c_1, v'_2)$ and $(w'_1 + \frac{1-\delta}{\delta}c_1, v'_2 - D_2)$ lie in this rectangle. These payoffs will serve as the continuation payoffs. Otherwise, the payoffs (w'_1, v'_2) and $(w'_1 - \frac{1-\delta}{\delta}c_1, v'_2 - D_2)$ will be the continuation payoffs. In both cases, when Player 1 is conveying a signal in the second period, he actually signals that the continuation payoff should be the high one. The first pair of continuation payoffs, designed for the lower w'_1, compensate for the cost of message at the first communication period. That is, the addition of $\frac{1-\delta}{\delta}c_1$ is for the sake of preserving the incentive of Player 1 to communicate the message stating he was the punished player, even when w'_1 is the lowest possible.

The punishment reduces the future payoff enough to make a deviation unprofitable, if it does so when future payoff is the lowest. When considering deviations from the play path, the lowest possible continuation payoff is $u_i - \varepsilon$. When addressing a deviation from the continuation play, the lowest possible continuation payoff is $u'_i = v'_i - \varepsilon$. Combining these two observations, we conclude that for the punishment to reduce future payoff enough, it suffices that:

$$u'_i \leqslant u_i - 2\varepsilon - D_i. \tag{5}$$

The strategies depend on the parameters N, ε and δ, whose existence is guaranteed by Lemma 4 below.

6.2 Deviations are not profitable

6.2.1 A deviation from the play path or from the continuation game is not profitable

Suppose that along the play path or during a continuation game Player i's payoff is u'_i. For a deviation of Player i to be unprofitable, it is sufficient that the punishment is severe enough so that the one period gain due to the deviation, plus the post deviation payoff (the payoffs during the punishment, plus the payoffs during the first two periods of the continuation game plus the continuation payoff) is lower than u'_i. At most ε is added to the future payoff after the punishment and the communication periods, and so the highest continuation payoff is $u'_i + \varepsilon$ (recall, $D_1, D_2 < \varepsilon$). Let $\bar{u}_i$ and $\underline{u}_i$ be the highest and lowest (respectively) feasible payoffs of Player i. The requirement, formally, is:

$$(1-\delta)\left(\bar{u}_i - \underline{u}_i\right) + \delta\frac{1-\delta^N}{1-\delta}0 + \delta^{N+1}(1-\delta^2)v_i^{NE} + \delta^{N+3}(u'_i + \varepsilon) < u'_i. \tag{6}$$

The last item is an upper bound of the continuation payoff. From (5) and (6) we conclude that a deviation is unprofitable if $\forall i \in \{1,2\}$:

$$1-\delta < \frac{(1-\delta^{N+3})(u_i - D_i - 2\varepsilon) - \delta^{N+3}\varepsilon}{\bar{u}_i - \underline{u}_i + \delta^{N+1}(1+\delta)v_i^{NE}}. \tag{7}$$

Ineq. (7) aids establishing a lower bound for the discount factor. In spirit, this is similar to Ineq. (5) in Fudenberg and Maskin (1986). Both aim to verify that the players are patient enough so that the future reduction of payoffs due to punishment shall cancel out the current gain from deviation.

There is yet another kind of deviation. In the first two periods of the continuation game the players play a one shot game Nash equilibrium and convey messages. Thus, a profitable deviation could only be either not conveying a message stating that one is the punished player when indeed one is, or conveying a message that one has been punished while one has not been. For the punishing player it is not profitable to convey a message because (a) communication is costly and (b) such a message is followed by the lowest continuation payoff, $v_i' - D_i$ (see 6.1.3).

For the player who was punished, it is sufficient to show that he prefers to convey the message even when the future payoff is the lowest possible, $u_i - \varepsilon - D_i$ (recall, the continuation payoff after the first two communication periods are always within the rectangle). In this case the payoffs used for continuation payoffs are $(v_1' + \frac{1-\delta^2}{\delta^2}c_1, v_2)$ and $(v_1' + \frac{1-\delta}{\delta}c_1, v_2 - D_2)$. It is then sufficient to guarantee that for i $\in \{1,2\}$:

$$(1-\delta)(-c_i) + (1-\delta)\delta v_i^{NE} + \delta^2(u_i - \varepsilon - D_i + \frac{1-\delta}{\delta^2}c_i) \geqslant (1-\delta)\delta v_i^{NE} + \delta^2(u_i - \varepsilon - D_i). \tag{8}$$

Basic algebra shows that it holds as an equality.

A profitable deviation at the second period of the communication phase could be only by altering the probability of sending a message (regarding the continuation payoff). However, the player conveying the message is indifferent between the two options (see 6.1.3). This shows that there is no profitable deviation from the second period either.

6.2.2 A Deviation from the punishment phase is not profitable

During the punishment phase, the punished player, Player 1, plays his best responses, therefore he has no profitable deviations. As for Player 2, due to Lemma 2 (applied, recall, with $e = \delta^{N+2}$ and $D = D_2$) she is indifferent between all actions in Q^*, and weakly preferring them over actions not in Q^*. Thus, all possible deviations are not profitable.

The following lemma proves that values of the parameters δ, ε and N can be found to satisfy all the equilibrium requirements simultaneously.

Lemma 4. *For every $(u_1, u_2) \in EF \cap IR$ there exists $\delta_4 \in (0,1)$ such that $\forall \delta > \delta_4$, parameters ε and N can be found so as to satisfy Ineq. (7) for D_1, D_2 derived from Lemma 3 and allow the linear system of Lemma 2 to have a solution.*

6.3 The case of simultaneous deviations

The solution concept of sequential equilibrium requires that the players have beliefs about the history of the game, and play a best response to other players' strategies based on these beliefs. The combination of strategies and beliefs is called assessment. During the game, the beliefs are updated in a Bayesian manner. In case the observed signal is inconsistent with following the instructions to the players the beliefs are derived from a converging system of assessments in which the strategies have full support (thus Bayesian updating of the beliefs is possible following any signal).

In our construction there are three phases of the game: the play path, the punishment phase and the continuation play. For each such phase there are specific instructions that are best replies when the belief is that the opponent is at the same phase. When players follow these instructions, the beliefs are easily constructed: each player believes with probability 1 that his opponent is conforming. When a single player deviates and the deviation is detected, the beliefs are simple as well: the conforming player detects the deviation, and the deviating player knows that his deviation was observed.

The case where beliefs should be explicitly designed is when both players simultaneously deviate. These beliefs require a delicate construction. In order to illustrate the idea of how to construct the beliefs, consider the following example.

Example 2.

	L	C	R
T	3,3	0,4	0,-2
M	-2,-1	4,0	-2,-2
B	5,-1	-2,4	-3,-2

Table 7: The payoff matrix of Example 2.

Suppose that on the equilibrium path the players are instructed to play (T, L) and they simultaneously deviate: Player 1 deviates from T to B, and Player 2 from L to C. Then, Player 1 gets -2 instead of 5*, thus knowing that Player 2 deviated as well, but Player 2 gets* 4*, which is the payoff she expects to obtain when deviating while Player 1 is conforming. In this* *case, Player 2, knowing that for Player 1 any deviation is not profitable (due to the subsequent punishment), will place zero probability on the event that Player 1 deviates, and will assume that she was the only one deviating. Thus, she believes that she is the one to be punished.*

Note that in some scenarios, the beliefs can be simply derived from Baysian updating. This happens when the signals are compatible with the assumption that the opponent is conforming. In this case, a player believes that the opponent is indeed conforming. So a player either believes with probability 1 that the opponent conforms, or knows the opponent deviated. In the later case, a player that deviated and observed a deviation of the opponent may not know whether his own deviation was observed by the opponent.

To further explain, consider a slight change:

Example 3.

	L	C	R
T	3,3	0,4	0,-2
M	-2,-1	4,0	-2,-2
B1	5,-1	-2,3	-3,-2
B2	5,-1	5,3	-3,-2

Table 8: The payoff matrix of Example 3.

Suppose that on the equilibrium path the players are instructed to play (T,L), yet they simultaneously deviate: Player 1 to B1 and Player 2 to C. Player 2 observes a payoff of 3, which indicates a deviation of Player 1. However, it does not indicate whether the deviation was to B1 or B2. A deviation to B1 allows Player 1 to detect Player 2's deviation, while a deviation to B2 does not. Both know that the opponent deviated. The belief of Player 1 is that Player 2 observed his deviation (he knows that she knows that he deviated). The belief of Player 2 needs to be defined. She does not know whether her deviation was observed. In fact, in this case, the beliefs of Player 2 may differ according to different limits of perturbations of beliefs.

In our equilibrium construction, following a single or simultaneous deviation, a player has a well-defined best reply if he places a mass-point probability on one of the following possibilities: no player is to be punished, Player 1 is to be punished or Player 2 is to be punished. We show how to obtain such beliefs when off-equilibrium they are derived from a converging sequence of perturbations in the following.

6.4 The beliefs off the equilibrium path

Appendix B elaborates on the formal definitions of the converging perturbations leading to the off equilibrium path systems of beliefs and their corresponding best replies. The general idea is that either (a) the mass-point limit leads a player to belief that a certain single deviation should be punished or (b) there is common knowledge that both players simultaneously deviated. The corresponding best replies follow from the instructions of the different phases. Appendix B also demonstrates the beliefs construction on the situations described in Examples 2 nd 3 above.

6.5 A Generalization of the information structures

6.5.1 Observing payoffs and an additional signal

The lemmas above referred to a monitoring structure where the only signal a player observes is his own payoff.

Consider an information structure where the players are informed of their own payoffs, and in addition, on some random signal that may depend on the actions taken. During all phases, any information other than the payoffs can be ignored without interfering with the equilibrium construction. Even signals that indicate deviations are ignored as long as they have no effect on observed payoffs. In the case where only payoffs are observable, a deviation is made unprofitable by punishments that follow detections, which are solely based on observed payoffs. This kind detection is effective also is cases where additional information is available and is being ignored.

In case of simultaneous deviations, the additional information may change the structure of the beliefs regarding the identity of the deviator. However, the general idea in Section 6.3 still holds: either there is common knowledge regarding the mutual deviation, or there is some order of ignorance, where the ‘assume ignorance’ rule applies just as well. Thus, the proof generalizes to the case of additional information with minor adjustments.

6.5.2 Observing own payoff with positive probability

Now consider an information structure where each player observes his own payoff only with some positive probability, and that this observation when occurs is common knowledge. The equilibrium we constructed relies on the detectability of profitable deviations, and on the ability of a punished player to properly define and communicate continuation payoffs. Only minor changes are needed in the equilibrium construction. In order to compensate for being detected only with some probability, the weight of the punishment phase should increase compared to a single period’s weight.

For simplicity, let I be the minimal probability of observing own payoff, across all pairs of actions. Ineq. (6) becomes,

$$(1-\delta)(\bar{u}_i - \underline{u}_i) + I\left[\delta^{N+1}(1-\delta^2)v_i^{NE} + \delta^{N+3}(u_i' + \varepsilon)\right] + \delta(1-I)u_i' < u_i'$$

or

$$(1-\delta)\left[\bar{u}_i - \underline{u}_i + I\delta^{N+1}(1+\delta)v_i^{NE} - u_i'\right] < \delta I u_i' - I\delta^{N+3}(u_i' + \varepsilon).$$

If the left hand side is negative, then for a small enough ε it trivially holds. Otherwise, Ineq. (7) becomes for the lowest possible continuation payoff,

$$1-\delta < \frac{\delta(u_i - D_i - 2\varepsilon) - I\delta^{N+3}(u_i - D_2 - \varepsilon)}{\bar{u}_i - \underline{u}_i + I\delta^{N+1}(1+\delta)v_i^{NE} - (u_u - D_i - 2\varepsilon)}.$$

This requires small enough ε and more patient players.

The linear systems of equations (Lemma 2) that defines the continuation payoffs should be slightly changed. Let $I(a(s), a_2')$ be the probability that Player 1 observes his payoff when the action profile $(a(s), a_2')$ is played. The updates of the continuation payoffs should be adjusted to the updates being made only with that probability. To be accurate, the linear system in Lemma 2 should be:

$$\begin{cases} U_2(p^*, a_2')(1-\delta)\delta^t + e\sum\limits_{s=1}^{S} p^*(a(s))M_s(p^*, a_2')X_s^t I(a(s), a_2') = K^t & \forall a_2' \in Q^*, t = 0, ..., N-1 \\ U_2(p^*, a_2')(1-\delta)\delta^t + e\sum\limits_{s=1}^{S} p^*(a(s))M_s(p^*, a_2')X_s^t I(a(s), a_2') \leqslant K^t & \forall a_2' \notin Q^*, t = 0, ..., N-1 \end{cases}$$

$$\sum_{t=0}^{N-1} \min_s \left\{X_s^t\right\} = 0,$$

and

$$\sum_{t=0}^{N-1} \max_s \left\{X_s^t\right\} \leqslant D.$$

This linear system has a solution for the same reason the former systems did (the linear independence of the columns of the information matrix).

The signals conveyed during the communication phase are still observed with probability 1, and thus Ineq. (8) is left unchanged. In addition, Lemma 3 (ii) should read: "any profitable deviation is detected with a positive probability".

The rest of the proof does not require any modification, and the loss in the accuracy of the signals mainly manifests itself in requiring extra patience on the part of the players.

7 Four final comments

7.1 Ties of payoffs

The difficulty of detecting deviations arises in our model only when there are ties between payoffs. Such ties appear in many classes of games, for example in auction games when the bidders are informed only of the outcome. More examples where such ties exist can be found in Section 2.

7.2 Supporting additional payoffs as sequential equilibrium payoffs

A full Folk Theorem would refer to the entire set of payoffs that could be obtained in a repeated game. Here we described only how to sustain Pareto efficient payoffs and the payoffs in the convex hull of the strictly Pareto efficient and one-period Nash equilibrium payoff. A natural question is what other payoffs could be supported.

Deviations from actions that produce payoffs that are not strictly efficient are not always detectable. The techniques presented in this paper are not easily generalized to these cases, as can be seen in the following example.

Example 4.

	L	C	R
T	3,3	3,0	0,1
M	1,2	1,1	0,2
B	2,0	2,-1	0,0

Table 9: Weak Pareto efficient cannot be a sequential equilibrium payoff

Here, the weakly Pareto efficient payoff $(3,0)$ *cannot be a sequential equilibrium payoff. The reason is that C and L yield the same payoff for Player 1, no matter what he plays. Therefore, based on the signals he receives, Player 1 cannot detect a deviation of Player 2 from C to L. Furthermore, due to such deviation Player 2 does not lose information (by playing L she can distinguish between any two actions that she can distinguish between when playing C).*

We conclude that when players observe their own payoff, typically, one cannot get a full Folk theorem.

7.3 Three players model

Extending the result presented here to games with three players or more is not straightforward. In a two-player game when a single player deviation takes place, there is common knowledge regarding the identity of the deviator. When three or more players are involved, either a mechanism that reveals the deviator's identity must be present or a punishment that simultaneously sanctions several players should be available. One situation where such a sanction is available is when attempting for a Nash-threat folk theorem.

7.4 Zero-cost communication

In this paper communication is costly. A careful reading of the proof reveals that communication takes place only off equilibrium. The construction used here relies on the positive cost of communication. Reducing the cost of communication to zero ('cheap-talk') gives rise to a whole different use of communication. When communication is free, it can be used in every period of the game without harming efficiency. In Ashkenazi-Golan and Lehrer (2019) we explore the free communication model and obtain a full characterization of the sequential equilibrium payoffs, using an equilibrium structure different from the one presented here.

References

Abreu, D., D. Pearce, and E. Stacchetti, "Optimal cartel equilibria with imperfect monitoring," *Journal of Economic Theory*, 1986, *39* (1), 251–269.

_ , _ , and _ , "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, 1990, *58* (5), 1041–1063.

_ , Milgrom P., and D. Pearce, "Information and timing in repeated partnerships," *Econometrica: Journal of the Econometric Society*, 1991, pp. 1713–1733.

Ashkenazi-Golan, G. and E. Lehrer, "Blackwell's comparison of experiments and discounted repeated games," *under revision*, 2019.

Aumann, R. J. and L. Shapley, *Long-term competition - a game-theoretic analysis*, Springer, 1994.

Bhaskar, V. and I. Obara, "Belief-based equilibria in the repeated prisoners' dilemma with private monitoring," *Journal of Economic Theory*, 2002, *102* (1), 40–69.

Compte, O., "Communication in repeated games with imperfect private monitoring," *Econometrica*, 1998, *66* (3), 597–626.

Ely, J. C. and J. Välimäki, "A robust folk theorem for the prisoner's dilemma," *Journal of Economic Theory*, 2002, *102* (1), 84–105.

Ely, J., J. Hörner, and W. Olszewski, "Belief-free equilibria in repeated games," *Econometrica*, 2005, *73* (2), 377–415.

Forges, F., "An approach to communication equilibria," *Econometrica: Journal of the Econometric Society*, 1986, pp. 1375–1385.

Fudenberg, D. and D. Levine, "The Nash-threats folk theorem with communication and approximate common knowledge in two player games," *Journal of Economic Theory*, 2007, *132* (1), 461–473.

_ and E. Maskin, "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, 1986, *54* (3), 533–554.

_ and _ , "On the dispensability of public randomization in discounted repeated games," *Journal of Economic Theory*, 1991, *53* (2), 428–438.

_ , D. Levine, and E. Maskin, "The folk theorem with imperfect public information," *Econometrica*, 1994, *62* (5), 997–1039.

Hillas, J. and L. Min, "Correlated equilibria of two person repeated games with random signals," *International Journal of Game Theory*, 2016, *45* (1-2), 137–153.

Hörner, J. and W. Olszewski, "The Folk Theorem for Games with Private Almost-Perfect Monitoring," *Econometrica*, 2006, *74* (6), 1499–1544.

Kandori, M., "Weakly Belief-Free Equilibria in Repeated Games With Private Monitoring," *Econometrica*, 2011, *79* (3), 877–892.

_ and H. Matsushima, "Private observation, communication and collusion," *Econometrica*, 1998, pp. 627–652.

Kreps, D. and R. Wilson, "Sequential equilibria," *Econometrica*, 1982, *50* (4), 863–894.

Lehrer, E., "Lower equilibrium payoffs in two-player repeated games with non-observable actions," *International Journal of Game Theory*, 1989, *18* (1), 57–89.

_ , "Nash equilibria of n-player repeated games with semi-standard information," *International Journal of Game Theory*, 1990, *19* (2), 191–217.

_ , "Internal correlation in repeated games," *International Journal of Game Theory*, 1991, *19* (4), 431–456.

_ , "Correlated equilibria in two-player repeated games with nonobservable actions," *Mathematics of Operations Research*, 1992, *17* (1), 175–199.

_ , "On the equilibrium payoffs set of two player repeated games with imperfect monitoring," *International Journal of Game Theory*, 1992, *20* (3), 211–226.

— , “Two-player repeated games with nonobservable actions and observable payoffs,” *Mathematics of Operations Research*, 1992, *17* (1), 200–224.

Mailath, G. J. and L. Samuelson, *Repeated games and reputations: long-run relationships*, Oxford University Press, 2006.

— **and S. Morris**, “Repeated games with almost-public monitoring,” *Journal of Economic theory*, 2002, *102* (1), 189–228.

Maschler, M., E. Solan, and S. Zamir, *Game Theory (Translated from the Hebrew by Ziv Hellman and edited by Mike Borns)*, Cambridge University Press, New York, 2013.

Obara, I., “Folk theorem with communication,” *Journal of Economic Theory*, 2009, *144* (1), 120–134.

Piccione, M., “The repeated prisoner’s dilemma with imperfect private monitoring,” *Journal of Economic Theory*, 2002, *102* (1), 70–83.

Renault, J. and T. Tomala, “Communication equilibrium payoffs in repeated games with imperfect monitoring,” *Games and Economic Behavior*, 2004, *49* (2), 313–344.

Rubinstein, A., “Equilibrium in supergames,” in “Essays in Game Theory,” Springer, 1994, pp. 17–27.

Sorin, S., “On repeated games with complete information,” *Mathematics of Operations Research*, 1986, *11* (1), 147–160.

Sugaya, T., “Folk Theorem in Repeated Games with Private Monitoring,” *working paper*, 2015.

— , “The Characterization of the Limit Communication Equilibrium Payoff Set with General Monitoring: Observable Realized Own Payoff Case,” *working paper*, 2017.

— **and A. Wolitzky**, “Bounding equilibrium payoffs in repeated games with private monitoring,” *Theoretical Economics*, 2016.

8 Appendix A- Proofs

We start with a technical lemma which will be needed later.

Lemma 5. *For any $A_1' \subseteq A_1$ and $\lambda, \lambda' \in \mathbb{R}^{|A_2|}$, if $M(A_1', A_2)\lambda' = M(A_1', A_2)\lambda$, then*

$$\sum_{a_k \in A_2} \lambda_k' = \sum_{a_k \in A_2} \lambda_k.$$

Proof. The columns of the information matrix $M(p, A_2)$ consist of zeros and ones. The number of ones in each column equals the number of rows in the support of p. Denote it by ρ. Let $M_i(p, a_r)$ be the ith entry of the column $M(p, a_r)$. One obtains,

$$\begin{aligned}\sum_{a_k\in A_2}\lambda_k\rho &= \sum_{a_k\in A_2}\lambda_k\sum_{a_i\in A_1'}M_i(A_1',a_k)\\ &= \sum_{a_i\in A_1'}\sum_{a_k\in A_2}\lambda_k M_i(A_1',a_k) = \sum_{a_i\in A_1'}\sum_{a_k\in A_2}M_i(A_1',a_k)\lambda_k'\\ &= \sum_{a_k\in A_2}\lambda_k'\sum_{a_i\in A_1'}M_i(A_1',a_k) = \sum_{a_k\in A_2}\lambda_k'\rho.\end{aligned}$$

Thus, $\sum_{a_k\in A_2}\lambda_k' = \sum_{a_k\in A_2}\lambda_k$. ■

Lemma 1. *There exists a pair of distributions $(p^*,q^*)\in\Delta(A_1)\times\Delta_2$ such that*

(1) $U_1(p^*,q^*)=0$,

(2) $U_1(a_1,q^*)=0\Rightarrow a_1\in\operatorname{supp}(p^*)$,

(3) $\forall q'\in\Delta_2$, $M(p^*,A_2)q'=M(p^*,A_2)q^*\Rightarrow U_2(p^*,q')\leqslant U_2(p^*,q^*)$,

(4) $\forall q'\in\Delta(A_2)$ *and* $\lambda\in\mathbb{R}^{|A_2|}$, *if* $\operatorname{supp}(q'),\operatorname{supp}(\lambda)\subseteq\operatorname{supp}(q^*)$, *then*

$$M(p^*,A_2)q'=M(p^*,A_2)\lambda\Rightarrow U_2(p^*,q')=U_2(p^*,\lambda).$$

(5) For every $q'\in\Delta(A_2)$ *and* $\lambda\in\mathbb{R}^{|A_2|}$, *if* $\operatorname{supp}(\lambda)\subseteq Q^*$ *and* $M(p^*,A_2)q'=M(p^*,A_2)\lambda$, *then* $U_2(p^*,q')\leqslant U_2(p^*,\lambda)$.

Proof. Denote,

$$B:=\Bigg\{(p,q)\in\Delta(A_1)\times\Delta_2;\qquad U_1(p,q)=0\text{ and}\\ M(p,A_2)q=M(p,A_2)q'\Rightarrow U_2(p,q')\leqslant U_2(p,q)\Bigg\}.$$

The set B is not empty. To see this, consider $q'\in\Delta_2$, and let p be a distribution over the set of all (pure) best replies of Player 1. Thus, $U_1(p,q')=0$. The set of distributions $q\in\Delta_2$ satisfying the equality $M(p,A_2)q'=M(p,A_2)q$ is compact. Thus, $U_2(p,.)$ attains a maximum over this set, say at q.

The information matrix $M(p,A_2)$ describes the distribution over the different signals of Player 1. Since payoffs are observable, two Player 2's mixed actions, q and q', that satisfy $M(p,A_2)q=M(p,A_2)q'$ induce identical distributions over signals, and thus must induce identical distributions of Player 1's payoffs, against any action in the support of p. Formally, $M(p,A_2)q'=M(p,A_2)q$ implies $U_1(a,q')=U_1(a,q)$ for every $a\in\operatorname{supp}(p)$ and in particular, $0=U_1(p,q')=U_1(p,q)$. We therefore obtain that $(p,q)\in B$ and therefore B is not empty. Note that any $(p^*,q^*)\in B$ satisfies (1) and (3).

Let $(p^*,q^*)\in B$ be such that $\operatorname{supp}(p^*)$ is maximal: there does not exist $(p,q)\in B$ such that $\operatorname{supp}(p^*)\subsetneq\operatorname{supp}(p)$. We claim that (p^*,q^*) satisfies (2). Assume, by contradiction, that there exists

$a_1 \in A_1$ such that $U_1(a_1, q^*) = 0$, $a_1 \notin \text{supp}(p^*)$. Consider[9] $\overline{p} = (1-\varepsilon')p^* + \varepsilon'(a_1)$ and $\overline{q}$ such that: $\overline{q} \in \Delta_2$ maximizes $U_2(\overline{p}, .)$ among all $q \in \Delta_2$ that satisfy $M(\overline{p}, A_2)q^* = M(\overline{p}, A_2)q$. Since $M(\overline{p}, A_2)q^* = M(\overline{p}, A_2)\overline{q}$, for every $a \in \text{supp}(\overline{p})$, $0 = U_1(a, q^*) = U_1(a, \overline{q})$. Hence, $(\overline{p}, \overline{q})$ satisfies (1) and (3) while $\text{supp}(p^*) \subsetneq \text{supp}(\overline{p})$, a contradiction to $\text{supp}(p^*)$ being a maximal set.

To see that (p^*, q^*) satisfies (4) as well, let $q' \in \Delta(A_2)$ and $\lambda \in \mathbb{R}^{|A_2|}$ be such that $\text{supp}(q'), \text{supp}(\lambda) \subseteq \text{supp}(q^*)$ and $M(p^*, A_2)q' = M(p^*, A_2)\lambda$. Assume, by contradiction, that $U_2(p^*, q') \neq U_2(p^*, \lambda)$. From Lemma 5, $\sum_k \lambda_k = 1$. If $U_2(p^*, q') > U_2(p^*, \lambda)$, consider $\overline{q} = q^* + \varepsilon'(q' - \lambda)$, and if $U_2(p^*, q') < U_2(p^*, \lambda)$, consider $\overline{q} = q^* + \varepsilon'(\lambda - q')$. Due to $\text{supp}(q'), \text{supp}(\lambda) \subseteq \text{supp}(q^*)$, for $\varepsilon' > 0$ small enough $\overline{q}$ is in $\Delta(A_2)$. We obtain, $M(A_1, A_2)\overline{q} = M(A_1, A_2)q^*$. Furthermore, due to the definition of $\overline{q}$, $U_2(p^*, \overline{q}) > U_2(p^*, q^*)$, contradicting (3). We therefore conclude that (4) is also satisfied.

Recall, Q^* is a subset of A_2 such that $V(p^*, Q^*) = V(p^*, q^*)$, and the columns of $M(p^*, Q^*)$ are independent. As for (5), suppose that $q' \in \Delta(A_2), \lambda \in \mathbb{R}^{|A_2|}$, $\text{supp}(\lambda) \subseteq Q^*$, and $M(p^*, A_2)q' = M(p^*, A_2)\lambda$. Assume, in a way of contradiction, that $U_2(p^*, q') > U_2(p^*, \lambda)$. From Lemma 5,
$\sum_{a_k \in A_2} q'(a_k) = 1 = \sum_{a_k \in Q^*} \lambda_k$.

Since $\text{supp}(\lambda) \subseteq Q^*$, for $\varepsilon' > 0$ small enough, $\overline{q} = q^* + \varepsilon'(q' - \lambda)$ is a distribution as well. Also, $M(p^*, A_2)\overline{q} = M(p^*, A_2)q^*$ implies $\forall a_2 \in \text{supp}(p^*)$, $U_1(a_2, \overline{q}) = U_1(a_2, q^*) = 0$. Moreover, by (2), Player 1's actions out of $\text{supp}(p^*)$ are not best reply to q^* and so the payoff for Player 1 when playing them against q^* is negative, and for ε' small enough it is still negative when played against $\overline{q}$. Thus, $\overline{q}$ is a minmaxing strategy.

We obtained that $\overline{q}$ is a minmaxing strategy such that $U_2(p^*, \overline{q}) = U_2(p^*, q^*) + \varepsilon'(U_2(p^*, q') - U_2(p^*, \lambda)) > U_2(p^*, q^*)$ and $M(p^*, A_2)\overline{q} = M(p^*, A_2)q^*$, in contradiction to (3). ■

A set $Q \subseteq A_2 \backslash \text{supp}(q^*)$ is called *complete* if (i) The columns of $M(p^*, Q^* \cup Q)$ are independent, and (ii) $V(p^*, Q^* \cup Q) = V(p^*, A_2)$. In words, a subset Q of columns that is disjoint of $\text{supp}(q^*)$ is complete, if together with Q^* the corresponding columns of $M(p^*, Q^* \cup Q)$ are independent and algebraically span the entire space $V(p^*, A_2)$.

Before we get to the proof of Lemma 2 we prove two lemmas: Lemma 6 that provides a helpful property of the information matrix, and Lemma 7 which refers to the private case where $t = 0$.

Lemma 6. *There exists a complete set Q such that for any $a_2 \in A_2$, if $M(p^*, a_2) = M(p^*, A_2)\lambda$ and $\text{supp}(\lambda) \subseteq Q^* \cup Q$, then $U_2(p^*, a_2) \leqslant U_2(p^*, \lambda)$.*

The proof of Lemma 6 relies on the following claim, which uses two notations. For any vector $x \in R^n$ by $x \geqslant 0$ we mean that $x_i \geqslant 0$, $i = 1, ..., n$, and similarly $x > 0$ means $x_i > 0$, $i = 1, ..., n$.

For $q \in \Delta(A_2)$, and a complete set Q, let $\lambda^q_{Q^* \cup Q} \in \mathbb{R}^{|A_2|}$ be the unique vector that satisfies $M(p^*, A_2)\lambda^q_{Q^* \cup Q} = M(p^*, A_2)q$ and $\text{supp}(\lambda^q_{Q^* \cup Q}) \subseteq Q^* \cup Q$. With a slight abuse of notation let $\lambda^{a_2}_{Q^* \cup Q} \in \mathbb{R}^{|A_2|}$ be such that $M(p^*, A_2)\lambda^{a_2}_{Q^* \cup Q} = M(p^*, a_2)$.

Claim 1. *There exists a distribution q over A_2, such that*

(1) There is a non empty list of complete sets $Q_1, ..., Q_k$ such that $\lambda^q_{Q^* \cup Q_j} > 0, j = 1, ..., k$.

(2) For any complete set Q', $\lambda^q_{Q^* \cup Q'} \geqslant 0$ implies $Q' \in \{Q_1, ..., Q_k\}$ and thus, $\lambda^q_{Q^* \cup Q'} \geqslant 0$ implies $\lambda^q_{Q^* \cup Q'} > 0$.

[9] We use (a_1) to denote the distribution that assigns a_1 probability 1.

Proof. For any complete set Q' denote by $\text{cone}(Q^* \cup Q')$ the open cone generated by the columns of $M(p^*, Q^* \cup Q')$. That is, $\text{cone}(Q^* \cup Q') = \{M(p^*, A_2)\beta; \beta > 0, \text{supp}(\beta) = Q^* \cup Q\}$. Let $Q_1, ..., Q_k$ be a longest list (that is, k is maximal) of complete sets such that $\cap_{j=1}^k \text{cone}(Q^* \cup Q_j) \neq \emptyset$. This intersection is an open set (in $V(p^*, Q^* \cup Q')$) as an intersection of finitely many open sets.

Fix $q \in \Delta(A_2)$ such that $M(p^*, A_2)q \in \cap_{j=1}^k \text{cone}(Q^* \cup Q_j)$. Thus, (1) is satisfied. In order to show (2), assume that Q' is complete and $\lambda^q_{Q^* \cup Q'} \geqslant 0$. It means that $M(p^*, q)$ is at the same time in the closure of $\text{cone}(Q^* \cup Q')$ and in $\cap_{j=1}^k \text{cone}(Q^* \cup Q_j)$. Since $\text{cone}(Q^* \cup Q')$ is also open we obtain that the intersection of $\text{cone}(Q^* \cup Q')$ and $\cap_{j=1}^k \text{cone}(Q^* \cup Q_j)$ is not empty. If $Q \notin \{Q_1, ..., Q_k\}$ it would contradict the assumption about the maximality of k. We conclude that there is no complete Q' other than $Q_1, ..., Q_k$ such that $\lambda^q_{Q^* \cup Q'} \geqslant 0$. This completes the proof. ■

Proof of Lemma 6. Consider the distribution q and $Q_1, ..., Q_k$ guaranteed by Claim 1. Let j be such that $U_2(p^*, \lambda^q_{Q^* \cup Q_j})$ is the maximal across all $Q_1, ..., Q_k$. That is,

$$U_2(p^*, \lambda^q_{Q^* \cup Q_j}) \geqslant U_2(p^*, \lambda^q_{Q^* \cup Q_\ell}), \ell = 1, ..., k. \tag{9}$$

Set $Q = Q_j$. We claim that Q satisfies that for any $a_2 \in A_2$, if $M(p^*, a_2) = M(p^*, A_2)\lambda$, $\text{supp}(\lambda) \subseteq Q^* \cup Q$, then $U_2(p^*, a_2) \leqslant U_2(p^*, \lambda)$. Note that since Q is complete, $\lambda = \lambda^q_{Q^* \cup Q}$. Assume, by negation, that there exists $a_2 \in A_2$, such that $U_2(p^*, a_2) > U_2(p^*, \lambda^{a_2}_{Q^* \cup Q})$. It implies that $a_2 \notin Q^* \cup Q$, because otherwise $\lambda^{a_2}_{Q^* \cup Q} = 1_{a_2}$ implying $U_2(p^*, a_2) = U_2(p^*, \lambda^{a_2}_{Q^* \cup Q})$.

Denote, $z(c) = c[1_{a_2} - \lambda^{a_2}_{Q^* \cup Q}]$, for every $c \geqslant 0$. Note that $M(p^*, A_2)z(c) = 0$ for any $c \geqslant 0$. Consider, $\lambda^q_{Q^* \cup Q} + z(c)$. Recall that $\lambda^q_{Q^* \cup Q}$ is strictly positive. Thus, when c is small enough, all the coordinates of $\lambda^q_{Q^* \cup Q}$ remain positive in $\lambda^q_{Q^* \cup Q} + z(c)$. We increase c gradually, until $c = c_0$, which is the first time one of the coordinates, say of a_2', becomes zero. At this point, since (as $a_2 \notin Q^* \cup Q$), $\lambda^q_{Q^* \cup Q}(a_2) = \lambda^{a_2}_{Q^* \cup Q}(a_2) = 0$, the coefficient of a_2 is positive (i.e., c_0), while that of a_2' is zero. Formally, $\lambda^q_{Q^* \cup Q}(a_2') + c_0[1_{a_2}(a_2') - \lambda^{a_2}_{Q^* \cup Q}(a_2')] = 0$. Since $a_2' \in Q$, by the choice of q, $\lambda^q_{Q^* \cup Q}(a_2') > 0$. Moreover, since $1_{a_2}(a_2') = 0$ and $c_0 > 0$, we obtain,

$$\lambda^{a_2}_{Q^* \cup Q}(a_2') > 0. \tag{10}$$

Set $Q' = Q \cup \{a_2\} \backslash \{a_2'\}$. We show first that Q' is complete. For this purpose we show that $M(p^*, A_2)1_{a_2'}$ is in the span of $V(Q^* \cup Q')$ and that $a_2' \notin Q^*$. Recall that, $M(p^*, A_2)1_{a_2} = M(p^*, A_2)\lambda^{a_2}_{Q^* \cup Q}$. Thus,

$$\begin{aligned} M(p^*, A_2)\left(1_{a_2} - \sum_{\tilde{a}_2 \in (Q^* \cup Q') \backslash \{a_2\}} 1_{\tilde{a}_2} \lambda^{a_2}_{Q^* \cup Q}(\tilde{a}_2)\right) &= M(p^*, A_2)\left(1_{a_2} - \sum_{\tilde{a}_2 \in (Q^* \cup Q) \backslash \{a_2'\}} 1_{\tilde{a}_2} \lambda^{a_2}_{Q^* \cup Q}(\tilde{a}_2)\right) \\ &= \lambda^{a_2}_{Q^* \cup Q}(a_2') M(p^*, A_2) 1_{a_2'}. \end{aligned}$$

Therefore,

$$M(p^*, A_2)\left(\frac{1_{a_2} - \sum_{\tilde{a}_2 \in (Q^* \cup Q') \backslash \{a_2\}} 1_{\tilde{a}_2} \lambda^{a_2}_{Q^* \cup Q}(\tilde{a}_2)}{\lambda^{a_2}_{Q^* \cup Q}(a_2')}\right) = M(p^*, A_2) 1_{a_2'}. \tag{11}$$

This implies that $M(p^*, A_2)1_{a_2'}$ is a linear combination of the column in $M(p^*, Q^* \cup Q')$ and thus in $V(Q^* \cup Q')$, as desired. Eq. (11) can be written as,

$$\lambda^{a_2'}_{Q^* \cup Q'} = \frac{1_{a_2} - \sum_{\tilde{a}_2 \in (Q^* \cup Q') \setminus \{a_2\}} 1_{\tilde{a}_2} \lambda^{a_2}_{Q^* \cup Q}(\tilde{a}_2)}{\lambda^{a_2}_{Q^* \cup Q}(a_2')}. \tag{12}$$

From the assumption regarding a_2:

$$U_2(p^*, a_2) > U_2(p^*, \lambda^{a_2}_{Q^* \cup Q}) = \lambda^{a_2}_{Q^* \cup Q}(a_2') U_2(p^*, a_2') + \sum_{\tilde{a}_2 \in (Q^* \cup Q') \setminus \{a_2\}} \lambda^{a_2}_{Q^* \cup Q}(\tilde{a}_2) U_2(p^*, \tilde{a}_2),$$

which implies,

$$U_2(p^*, a_2) - \sum_{\tilde{a}_2 \in (Q^* \cup Q') \setminus \{a_2\}} \lambda^{a_2}_{Q^* \cup Q}(\tilde{a}_2) U_2(p^*, \tilde{a}_2) > \lambda^{a_2}_{Q^* \cup Q}(a_2') U_2(p^*, a_2').$$

Due to Ineqs. (10) and (12) it implies,

$$U_2(p^*, \lambda^{a_2'}_{Q^* \cup Q'}) > U_2(p^*, a_2'). \tag{13}$$

This inequality has two consequences. First, by Lemma 1(4), $a_2' \notin Q^*$. This, in turn, implies that Q' is complete. Second, the facts that Q' is complete and that all the coefficients of $\lambda^q_{Q^* \cup Q'}$ are non-negative imply by Claim 1 that $Q' \in \{Q_1, ..., Q_k\}$. In order to finish the proof we show that $U_2(p^*, \lambda^q_{Q^* \cup Q'}) > U_2(p^*, \lambda^q_{Q^* \cup Q_j})$ which contradicts Eq. (9).

Note that $\lambda^q_{Q^* \cup Q'} = \lambda^q_{Q^* \cup Q} + \lambda^q_{Q^* \cup Q}(a_2')(\lambda^{a_2'}_{Q^* \cup Q} - 1_{a_2'})$. Therefore,

$$U_2(p^*, \lambda^q_{Q^* \cup Q'}) = U_2(p^*, \lambda^q_{Q^* \cup Q}) + \lambda^q_{Q^* \cup Q}(a_2') \left(U_2(p^*, \lambda^{a_2'}_{Q^* \cup Q}) - U_2(p^*, a_2') \right) > U_2(p^*, \lambda^q_{Q^* \cup Q}),$$

where the inequality is due to Eq. (13). Since $Q' \in \{Q_1, ..., Q_k\}$, this inequality indeed contradicts Eq. (9), and the proof is complete. ■

Lemma 7. *For any $e \in (0,1)$ and $D > 0$, there exists $\delta' \in (0,1)$ such that for any $\delta > \delta'$ there exists a solution, $\{X^t_s\} \in \mathbb{R}^S$ and $K \in \mathbb{R}$, to the following system of inequalities,*

$$\begin{cases} U_2(p^*, a_2')(1-\delta) + e \sum_{s=1}^{S} p^*(a(s)) M_s(p^*, a_2') X_s = K & \forall a_2' \in Q^*, \\ \\ U_2(p^*, a_2')(1-\delta) + e \sum_{s=1}^{S} p^*(a(s)) M_s(p^*, a_2') X_s \leqslant K & \forall a_2' \notin Q^*, \\ \\ \max_s \{X_s\} - \min_s \{X_s\} \leqslant D. \end{cases}$$

Proof. Let Q be any set of pure actions guaranteed by Lemma 6. The set of columns of $M(p^*, Q^* \cup Q)$ is a set of independent columns. Therefore, the following system has a solution.

$$(1-\delta) U_2(p^*, a_2') + e \sum_{s=1}^{S} M_s(p^*, a_2') Y_s = K, \quad \forall a_2' \in Q^* \cup Q.$$

The solution does not have to be unique. In order to obtain a unique solution, we may add linear equalities, so that the matrix $M(p^*, Q \cup Q^*)$ together with the additional rows is invertible, and let Y_s be the solution. For any $a(s) \in \text{supp}(p^*)$ let $X_s = \frac{Y_s}{p^*(a(s))}$.

We established that the first two lines of the system (equality in the first and inequality in the second) hold for all actions in $Q^* \cup Q$. We now show that it holds for any action in A_2. Let a_2 be any action in A_2. From Lemma 6 the set Q is complete, and so $V(p^*, Q^* \cup Q) = V(p^*, A_2)$. Hence, there exists $\lambda \in R^{|A_2|}$, $\text{supp}(\lambda) \subset Q \cup Q^*$, such that $M(p^*, a_2) = M(p^*, A_2)\lambda$.

$$\begin{aligned}
(1-\delta)U_2(p^*, a_2) + e\sum_{s=1}^{S} M_s(p^*, a_2)Y_s &= (1-\delta)U_2(p^*, a_2) + e\sum_{s=1}^{S} \left[M_s(p^*, A_2)\lambda\right] Y_s \leqslant \\
(1-\delta)U_2(p^*, \lambda) + e\sum_{s=1}^{S} \left[M_s(p^*, A_2)\lambda\right] Y_s &= \sum_{a_2' \in Q^* \cup Q} \lambda(a_2') \left((1-\delta)U_2(p^*, a_2') + e\sum_{s=1}^{S} M_s(p^*, a_2')Y_s \right) \\
\sum_{a_2' \in Q^* \cup Q} \lambda(a_2')K &= K.
\end{aligned}$$

The inequality is due to Lemma 6, and the last equality is due to Lemma 5.

As for the last inequality of the lemma, for any δ_1 there exists a bound D' such that

$$\max_{\delta_1 \leqslant \delta < 1} \left[\max_s X_s(\delta) - \min_s X_s(\delta) \right] \leqslant D'.$$

Multiplying by $\frac{D}{D'}$, we obtain,

$$\frac{D}{D'} U_2(p^*, a_2')(1-\delta) + e\frac{D}{D'} \sum_{s=1}^{S} M_s(p^*, a_2')eY_s = \frac{D}{D'}K, \quad \forall a_2' \in Q^* \cup Q.$$

For $X_s' = \frac{D}{D'}\frac{Y_s}{p^*(a(s))}$, and $K' = \frac{D}{D'}K$ the solution satisfies:

$$\max_{\delta_1 \leqslant \delta < 1} \left[\max_s X_s'(\delta) - \min_s X_s'(\delta) \right] \leqslant \frac{D}{D'}D' = D.$$

Thus, the last inequality is satisfied as well. ■

Lemma 2. *For any $e \in (0,1)$, $D > 0$ and an integer N, there exists $\delta_2 \in (0,1)$ such that for any $\delta > \delta_2$ there are numbers $K^t \in \mathbb{R}$, $t = 0, ..., N-1$, and a solution to the following system of linear inequalities (with X_s^t being the unknowns),*

$$\begin{cases}
U_2(p^*, a_2)(1-\delta)\delta^t + e\sum_{s=1}^{S} p^*(a(s))M_s(p^*, a_2)X_s^t = K^t & a_2 \in Q^*, t = 0, ..., N-1 \\
U_2(p^*, a_2)(1-\delta)\delta^t + e\sum_{s=1}^{S} p^*(a(s))M_s(p^*, a_2)X_s^t \leqslant K^t & a_2 \notin Q^*, t = 0, ..., N-1
\end{cases}$$

$$\sum_{t=0}^{N-1} \min_s \left\{ X_s^t \right\} = 0$$

and

$$\sum_{t=0}^{N-1} \max_{s} \left\{X_s^t\right\} \leqslant D.$$

Proof. All first expressions in the equalities above are a multiplication of the former equalities by δ^t, and so multiplying X and K by the value of δ^t and subtracting from each resulting K^t and X_s^t the value of $\sum_{t=0}^{N-1} \min_s \left\{X_s^t\right\} = 0$ give a solution.

The inequality follows from the identity $(1-\delta)\sum_{t=1}^{N-1} \delta^t = 1 - \delta^N$ and from the last inequality in Lemma 7. ■

Lemma 3. *For every $(u_1, u_2) \in EF$ and $\varepsilon > 0$, there exist $D_1, D_2 > 0$ and $\delta_3 \in (0,1)$, such that for any $(v_1', v_2') \in EF \cap \mathcal{B}((u_1, u_2), \varepsilon)$, $1 > \delta > \delta_3$ and any payoff (v_1, v_2) in the convex hull of $\{(v_1', v_2'), (v_1' - D_1, v_2'), (v_1', v_2' - D_2), (v_1' - D_1, v_2' - D_2)\}$ there exists a strategy profile $\sigma = (\sigma_1, \sigma_2)$ such that:*

(i) the payoff of the strategy profile (σ_1, σ_2) is (v_1, v_2);

(ii) any profitable deviation from σ_i is detected with probability 1;

(iii) for every $t \in N$ and history $h(t)$ that has a positive probability under σ, $U_i(\sigma_1, \sigma_2, \delta, h(t)) \in \mathcal{B}((v_1, v_2), \varepsilon)$.

Proof. For every $(u_1, u_2) \in FF$, for ε small enough, the set $(v_1', v_2') \in EF \cap \mathcal{B}((u_1, u_2), \varepsilon) \cap EF$ consists of payoffs that are each a convex combination of at most two strictly Pareto efficient payoffs out of at most three such payoffs. Denote the three payoffs by w^1, w^2 and w^3, where $w^k = (w_1^k, w_2^k)$. For ε small enough there exists $D_1, D_2 > 0$ such that for any $(v_1', v_2') \in EF \cap \mathcal{B}((u_1, u_2), \varepsilon)$, the set $\operatorname{conv}\{(v_1', v_2'),\ (v_1' - D_1, v_2'),\ (v_1', v_2' - D_2),\ (v_1' - D_1, v_2' - D_2)\}$ is included in $\operatorname{conv}\left\{(w_1^k, w_2^k),\ (w_1^k - D_1, w_2^k),\ (w_1^k, w_2^k - D_2),\ (w_1^k - D_1, w_2^k - D_2) | k = 1, 2, 3\right\}$. Any profitable deviation from the actions yielding a strictly Pareto efficient payoff is detected due to reducing the opponent's payoff. When using the cost of the messages, all extreme points of the above set are payoffs of action profiles such that any profitable deviation is detectable.

Sorin (1986) shows that for any $\varepsilon > 0$, there exists $\delta_3 \in (0,1)$, such that for all $\delta > \delta_3$, there exists a strategy profile which consists of a sequence of the pure action profiles, such that its discounted sum is (u_1, u_2). Moreover, Lemma 2 of Fudenberg and Maskin (1991) shows that for every payoff in the convex hull the sequence can be designed such that the continuation payoff is always within an ε distance from the target payoff for players patient enough. For a given ε, the same bound on the discount factor applies for the entire convex hull. ■

Lemma 4. *For every $(u_1, u_2) \in EF \cap IR$ there exists $\delta_4 \in (0,1)$ such that $\forall \delta > \delta_4$, parameters ε and N can be found so as to satisfy Ineq. (7) for D_1, D_2 derived from Lemma 3 and while enabling the linear system of Lemma 2.*

Proof. For ε small enough, both numerator and denominator of the right hand-side of Ineq. (7) are positive. Therefore, for δ close enough to 1 the inequality holds. Let δ_3 and ε_3 to be large enough and small enough, respectively, so that both Ineq. (7) holds and D_1 and D_2 exist as in Lemma 3.

Employ Lemma 2 with the D_i found above and with $e = \delta_3^N$. According to this lemma, δ_2 large enough enables one to satisfy the system. Observe that increasing e simply means multiplying all the vector of solutions X by a constant. Thus, when $\delta_2 > \delta_3$ one also has a solution for $e = \delta_2^N$.

For $\delta_4 = \max\{\delta_2, \delta_3\}$ all the conditions hold simultaneously. ■

9 Appendix B- Incidence Matrices

9.0.1 Defining a converging system of beliefs

A best response might be difficult to find for an arbitrary assignment of perturbation probabilities (probabilities for the different possible "trembles" of the opponent). However, when the perturbations converge to a mass-point distribution on one vertex, a best reply is easy to find. This is because all that is mot known to a player are the opponent's actions (our model is one of perfect recall and no chance moves). Thus, having a mass-point belief over the opponent's actions means a mass-point belief on the vertex reached within a player's information set.

We design the perturbations to have yet another property. Following a deviation, once a player places probability 1 on whether or no a certain player should be punished, this belief does not change during the near future. In other words, the perturbations assigned to the initial deviations are more significant that those assigned to subsequent deviations.

In order to design beliefs where the initial update is the most significant one, we use powers of ε in a way similar to deconstructing a number by the digits: if the most significant digit of one number is larger than the other number's most significant digit, then the first number is larger, regardless of the remaining digits. An example of such beliefs can be the following. At each period, there is at least one action of Player i that is consistent with the equilibrium path, and at most $|A_i| - 1$ that are not. At the deviation period, assign probabilities $\varepsilon, \varepsilon^2, ...$ to all possible deviations. The highest power of ε is at most $|A_i| - 1$. In the first period after the deviation[10] we assign probabilities of $\varepsilon^{\frac{1}{|A_i|}}, \varepsilon^{\frac{2}{|A_i|}}, ...,$ where the maximal power is $\varepsilon^{\frac{|A_i|-1}{|A_i|}}$. In the k-th period after the deviation we assign probabilities of $\varepsilon^{\frac{1}{|A_i|^k}}, \varepsilon^{\frac{2}{|A_i|^k}}, ...$ It is as if number of "digits" here is $|A_i|$. That way, the belief regarding whether or not a punishment takes place and the identity of the player punished does not change during the punishment phase or in the periods afterwards, regardless of the observations. Note that in some cases the beliefs of the players may disagree, due to simultaneous deviations, and the disagreement may not be resolved. Yet, each one of the players has a crisp belief regarding the phase to be played and knowing the instructions is playing a best-reply to it.

9.0.2 Deriving the punished player's identity from the converging system of beliefs

The requirements from the beliefs in sequential equilibrium are the following:

(a) The beliefs given the private history of a player are derived as the limit of a converging system of full-support beliefs.

(b) Given the beliefs, each player is playing his best reply to the opponent's believed-to-be strategy.

[10] Note that these perturbations are relevant only following a deviation.

We establish the beliefs following a deviation, at the end of the period where a deviation took place. For the purpose of properly tracking down the players' system of beliefs (and their correspondingly best replies), we use an incidence matrix. This incidence matrix includes all possible pairs of action and signals of Player 1 (denote the set of these pairs Γ^1) as rows, and all possible pairs of action and signals of Player 2 (denoted Γ^2) as columns. Then 1's and 0's are placed in the cells according to whether the relevant combination of pairs of action-signal is possible or not.

For the sake of clarity we demonstrate the idea through Example 1. The incidence matrix of this example is given by:

	(L,3)	(L,-1)	(C,4)	(C,0)	(R,-2)
(T,3)	1	0	0	0	0
(T,0)	0	0	1	0	1
(M,-2)	0	1	0	0	1
(M,4)	0	0	0	1	0
(B,5)	0	1	0	0	0
(B,-2)	0	0	0	1	0
(B,-3)	0	0	0	0	1

Table 10: The incidence matrix of Example 1.

In Example 1, the combination (T,3) and (L,3) can occur, when Player 1 plays T and Player 2 L, Player 1 does observe 3 and Player 2 observes 3 as well. Hence, there is '1' in the respective entry. At the same time, the combination of (T,3) and (L,-1) cannot occur, so a '0' is placed in the corresponding entry.

Given a pair $\gamma \in \Gamma^i$, we design a belief over the opponent's information, which places a mass-point probability on an item of Γ^j, $j \neq i$. From this belief, a belief regarding the identity of the player to be punished is derived, namely, a belief placing unit-mass probability on an item from the set {Player 1, Player 2, none} (denoted, respectively $\{1, 2, none\}$). Denote by E_i, $i = 1, 2$ the function which assigns to any $\gamma \in \Gamma^i$ an item from Γ^j, $j \neq i$. Also denote by φ_i the function assigning to any $\gamma \in \Gamma^i$ an item from the set $\{1, 2, none\}$. Formally:

$E_1 : \Gamma^1 \to \Gamma^2$,
$E_2 : \Gamma^2 \to \Gamma^1$,
$\varphi_1 : \Gamma^1 \to \{none, 1, 2\}$,
$\varphi_2 : \Gamma^2 \to \{none, 1, 2\}$.

Translating requirements (a) and (b) above to these notations, we obtain:

(a) E_1 and E_2 are consistent with observations and with a converging sequence of off-equilibrium path perturbations.

(b) $\varphi_i(E_j(\gamma)) = \varphi_j(\gamma)$, $\forall \gamma \in \Gamma^i$, $i = 1, 2$, $j \neq i$.

The equilibrium instructs to play mixed strategy only during the punishment phase. Deviation during the punishment phase are not followed by further punishments[11]. Therefore, when we are establishing beliefs regarding whether or not a punishment should take place and regarding who the punished player should be, we are discussing deviations from pure strategy profiles.

In order to detail the functions E_i and φ_i, $i = 1, 2$, we divide the pairs of action and signal into the following four categories: the equilibrium action paired with the signal that should be observed if the opponent conforms (eq action eq signal); the equilibrium action paired with a signal that cannot be observed when the opponent conforms (eq action noneq signal); a deviating action paired with the signal that should be observed if the opponent conforms (noneq action eq signal); and a deviating action paired with a signal that cannot be observed when the opponent conforms. We denote these sets of pairs by Γ^i_{ee}, Γ^i_{en}, Γ^i_{ne} and Γ^i_{nn}, respectively. All the combinations of the above pairs are detailed in Table 11. Some entries hold zeros, due to the impossibility of the respective combination (for example, if a player conforms, the opponent cannot observe a message indicating a deviation). The remaining entries are labeled A to H, and they will be discussed below.

	eq action eq signal Γ^2_{ee}	eq action noneq signal Γ^2_{en}	noneq action eq signal Γ^2_{ne}	noneq action noneq signal Γ^2_{nn}
eq action eq signal Γ^1_{ee}	1	0	A	0
eq action noneq signal Γ^1_{en}	0	0	B	0
noneq action eq signal Γ^1_{ne}	C	D	E	F
noneq action noneq signal Γ^1_{nn}	0	0	G	H

Table 11: Combinations of actions and signals.

Consider the third column of the matrix. This column represents a situation where Player 2 deviated and observes a signal that is the expected one when Player 1 conforms. Given the Bayesian manner of the updates of the beliefs, it means that Player 2 places a probability of 1 on the event that Player 1 conformed. Knowing her deviating action, she knows for certain whether a conforming opponent observes this deviation (a situation represented in the cell labelled B) or does not observe it (cell A). Thus, whenever Player 2 is in an information pair belonging to the third column, she must either place probability 1 on A or probability 1 on B (and zero on all the remaining events). If it is A that has the unit probability, then Player 2 believes with probability 1 that nobody will be punished, and if it is B then the belief is that she is to be punished. Officially, for any $\gamma \in \Gamma^2_{ne}$, either $E_2(\gamma) \in \Gamma^1_{ee}$ and $\varphi_i(\gamma) = none$ or $E_2(\gamma) \in \Gamma^1_{en}$ and $\varphi_i(\gamma) = 2$. From Player 1's perspective, when the information pair belongs to Γ^1_{ee} with probability 1 he believes that no deviation occurred and when it belongs to

[11] The punished player has no profitable deviations from the minmax, and the punishing player simply has his continuation payoff updated.

Γ^1_{en}, he knows that Player 2's signal is in Γ^2_{ne} and she should be punished. Officially, for any $\gamma \in \Gamma^1_{ee}$, $E_1(\gamma) \in \Gamma^2_{ee}$ and $\varphi_1(\gamma) = none$, and for any $\gamma \in \Gamma^1_{en}$, $E_1(\gamma) \in \Gamma^2_{ne}$ and $\varphi_1(\gamma) = 2$. This implies that for cells A and B the beliefs satisfy the requirements. The same logic leads to the beliefs in cells C and D satisfying requirements (a) and (b).

Next, consider the combination in E. As the analysis above implies, when the realization is in E, Player 1 assigns probability 1 either to C (with none of the players being punished) or D (with Player 1 being punished). Similarly, Player 2 assigns probability 1 either to A (none punished) or B (Player 2 punished). For all these beliefs, we established above that the requirements hold. Appendix B demonstrates how the situation demonstrated by Example 2 above belongs to combination E.

When the combinations in G are considered, we note that Player 2, again, assigns probability 1 to either A or B. The considerations of Player 1 are derived from the limit of the perturbations. He knows that a deviation of Payer 2 occurred, but he might not know whether it was a deviation to an action which leads Player 1 to Γ^2_{ne} or one that leads to Γ^2_{nn}. However, when the limit of the perturbations is a mass-point probability on Γ^2_{ne} or probability 1 on Γ^2_{nn}. If it is Γ^2_{ne}, then Player 1 believes that Player 2 played a deviation that did not enable her to detect his own deviation. Moreover, Player 1 having this mass-point probability either on an event when Player 2 believes she was observed (B), and then she believes she should be punished) or on an event when she believes that she was not observed (A) thus she believes none of the players should be punished. In other words, if Player 1's believes are that Player 2's information is in γ^2_{ne}, then he has a well-defined best replies. The same analysis leads to the respective results for the combinations represented by entry F.

Finally, we analyze the combination represented by H. When the information of Player 1 is in Γ^1_{nn}, the limit of the perturbations either leads him to believe that Player 2 is in Γ^2_{ne} or in Γ^2_{nn}. The case Γ^2_{nn}. The case of Γ^2_{ne} was treated above. If Player 1 believes that Player 2's information is in Γ^2_{nn}, then his belief regarding the action she played, and his knowledge about the limit of the perturbations tells his whether she believes she is in F (and then he can further deduce her beliefs) or in H. If she believes indeed that his information is in Γ^1_{nn} then a symmetric argument holds. Appendix B demonstrates how the situation demonstrated be Example 3 above belongs to combination H.

Shortly, either (a) the chain of he-believes-that-she-believes-that... ends at some point where he believes to be in G or she in H (and then the best reply is well defined), or (b) he believes they are at H; she believes they are at H; he believes that she believes she is at H; she believes that he believes she is at H and so on. This is exactly the definition of common knowledge of being at H.

More formally, either there exists a finite expression such that $E_1(E_2(E_1(....) \in G$, or $E_2(E_1(E_2(...) \in F$, or that for any such finite sequence the result of the sequential application of E_i when i alternates between 1 and 2 is in H. In that common knowledge case, we defined Player 1 to be the punished one. This concludes the definition.

9.0.3 Example 2

We analyze the beliefs of the players and their corresponding best replies, following the scenario where the players are instructed to play (L,T), yet both deviate and the action profile actually played is (B,C).

For this scenario, the sets of action-signal are as follows:

$\Gamma^1_{ee} = \{(T,3)\}$; $\Gamma^1_{en} = \{(T,0)\}$; $\Gamma^1_{ne} = \{(M,-2),(B,5)\}$; $\Gamma^1_{ne} = \{(M,4),(B,-2),(B,-3)\}$;

$\Gamma^2_{ee} = \{(L,3)\}$; $\Gamma^2_{en} = \{(L,-1)\}$; $\Gamma^2_{ne} = \{(C,4),(R,-2)\}$; $\Gamma^2_{ne} = \{(C,0)\}$.

	(L,3)	(L,-1)	(C,4)	(R,-2)	(C,0)
(T,3)	1	0	0	0	0
(T,0)	0	0	1	1	0
(M,-2)	0	1	0	1	0
(B,5)	0	1	0	0	0
(M,4)	0	0	0	0	1
(B,-2)	0	0	**1**	0	0
(B,-3)	0	0	0	1	1

Table 12: The incidence marix for Example 2

The bold '1' is the realized combination after the simultaneous deviation. Player 1's information is $(B,-2)$ which is in Γ^1_{nn}, meaning he observes a signal indicating that Player 2 deviated. The bold '1' is the only one in that row, so Player 1 knows that player 2's information is $(C,4)$ (if there were several '1's in that row, meaning if several deviations of Player 2 could be associated with Player 1's information, then the limit of the perturbations is used to decide which one Player 1 believes that occurred). To conclude, $E_1((B,-2)) = (C,4)$. Player 2's information is $(C,4)$, and placing zero probability on deviations of Player 1, she deduces that Player 1's information is $(T,0)$. Formally, $E_2((C,4)) = (T,0)$. If Player 1 plays T and observes 0, then he knows that a deviation took place, and, since in this case all deviations from L are observable when Player 1 plays T, he believes that Player 2 knows she should be punished. That is, $\varphi_1(T,0) = 2$. To conclude, $\varphi_1(B,-2) = \varphi_2(E_1(B,-2)) = \varphi_2(C,4) = 2$ and $\varphi_2(C,4) = \varphi_1(E_2(C,4))) = \varphi_1(T,0) = 2$. Both players believe that Player 2 should be punished.

9.0.4 Example 3

We analyze the beliefs of the players and their corresponding best replies, following the scenario where the players are instructed to play (L,T), yet both deviate and the action profile actually played is (B_1,C).

For this scenario, the sets of action-signal are as follows:

$\Gamma^1_{ee} = \{(T,3)\}$; $\Gamma^1_{en} = \{(T,0)\}$; $\Gamma^1_{ne} = \{(M,-2),(B_1,5),(B_2,5)\}$;
$\Gamma^1_{ne} = \{(M,4),(B_1,-2),(B_1,-3),(B_2,-3)\}$;
$\Gamma^2_{ee} = \{(L,3)\}$; $\Gamma^2_{en} = \{(L,-1)\}$; $\Gamma^2_{ne} = \{(C,4),(R,-2)\}$; $\Gamma^2_{ne} = \{(C,0),(C,3)\}$.

The bold '1' is the realized combination after the simultaneous deviation. Player 1's information is $(B_1,-2)$. The bold '1' is the only one in that row, so $E_1(B_1,-2) = (C,3)$.

Player 2's information is $(C,3)$, but there are two '1's in the column of $(C,3)$, that is, there are two deviations of Player 1 corresponding to $(C,3)$. This is when the limit of the perturbations is needed. There are two possibilities regarding where the mass function is placed. We detail the two options below.

Option 1: the mass-point belief is on $(B_2,5)$

	(L,3)	(L,-1)	(C,4)	(R,-2)	(C,0)	(C,3)
(T,3)	1	0	0	0	0	0
(T,0)	0	0	1	1	0	0
(M,-2)	0	1	0	1	0	0
(B_1,5)	0	1	0	0	0	0
(B_2,5)	0	1	0	0	0	1
(M,4)	0	0	0	0	1	0
(B_1,-2)	0	0	0	0	0	**1**
(B_1,-3)	0	0	0	1	0	0
(B_2,-3)	0	0	0	1	0	0

Table 13: The incidence marix for Example 2

When Player 1's information is $(B_2, 5)$, he believes that no deviation of Player 1 took place, $E_1((B_2, 5)) = (L, -1)$. All the 1's in the column of $(L, -1)$ are in rows belonging to information pairs in Γ^1_{ne}, and in all of them, Player 1 believes that Player 2's belief is $(L, -1)$. Thus, these deviations of Player 1 are observed, he believes that they are observed, and should be the punished player. Formally: $\varphi_2(L, -1) = 1$. Thus $\varphi_1(B_1, -2) = \varphi_2(E_1(B_1, -2)) = \varphi_2((L, -1)) = 1$. Also, $\varphi_2(C, 3) = \varphi_1(B_2, 5) = 1$. The requirements are fulfilled.

Option 2: the mass-point belief is on $(B_1, -2)$

When Player 1's information is $(B_1, -2)$, he knows Player 1's information is $(C, 3)$. If when the information of Player 2 is $(C, 3)$ she places the mass-point belief on $(B_1, -2)$, then $E_1((B_1, -2)) = (C, 3)$ and $E_2((C, 3)) = (B_1, -2)$. In this case, there is common knowledge that the players are in H, that is, both place probability 1 on both deviating; both belief with probability 1 that the opponent believe that bpth deviate etc. We defined that in this case Player 1 is to be punished. Thus $\varphi_1(B_1, -2) = 1$ and $\varphi_2(C, 3) = 1$. The requirements are fulfilled.