Getting Efficient with Deterministic Observable Payoffs

preliminary

Abstract

In this paper we investigate two-players infinitely repeated discounted games where each player observes only his own payoffs. The payoffs for each pair of actions are deterministic and communication is not allowed. We prove that the efficient frontier can be obtained via sequential equilibrium in games where there exists an internal payoff that can be used as a continuation payoff. We show how to obtain such an internal payoff in all but a very degenerate set of games.

We show how this result can be applied to some economic models such as repeated duopoly game, when the firms can be free to choose from a set of actions that is as rich as we wish (as long as it is finite), including offering bundles, price-cuts, two-times tariff and so on, while requiring that the demand is known.

1 Introduction

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

Alice \ *Bob*	*Opera*	*Boxing*	*Home*
Opera	$-1, 4$	$3, 3$	$3, 0$
Boxing	$0, 2$	$-1, -1$	$0, 0$
Home	$0, 2$	$0, 3$	$0, 0$

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

Consider an infinitely repeated duopoly game. The effect of price-discrimination tools (such as bundling, prices that depend on past purchases, two-part tariffs etc.) was investigated mainly on the once or twice repeated duopoly game (see Armstrong (2006) for a survey).

The effect of such tools in an infinitely repeated framework is yet to be researched. Jin (1999) showed that under assumptions different than ours, collusion can be sustained in some games of oligopoly models even when actions are not directly observed. Jin (1999) restricts the actions available and fixes a linear demand function, while allowing for this demand to be stochastic. In addition, he assumes an outside source of aggregated information.

Our model assumes a finite action space for each firm, but no other assumption is imposed on the actions. Hence different pricing schemes, advertising tools and any other action that can be described in a finite manner ("the dollar amount spent on radio advertising“, for example) can be incorporated in our model.

In addition, we assume that those actions are not directly observed and we do not use any form of communication. We do assume, however, that the demand is known and constant, and that the firm's costs are known and constant[1]. What the firms do observe is their own profit following each period. Note that even if demand is known, and own profit is known, a firm cannot know for sure what action the opponent has taken to bring the profit for that level. For example, when profits fall, it can be because the rival firm used secret price-cuts, or maybe they offered a bundle-deal to a segment of the market, or maybe they spent a lot on advertising in local papers.

The last assumption we make is that the game always continues with both players, i.e. no firm is driven out of the market.

We make no other assumptions, so the demand and costs functions can take any form.

Our results suggest that both the information regarding the exact actions

[1]It seems like our model holds also when demand and costs are allowed to vary during time, but some assumptions need to be made about the amount by which they are allowed to vary, in order to keep the continuation payoffs obtainable. However, The assumption that all those parameters need to be known to both firms before each period cannot be relaxed in the framework of this paper.

of the opponent and the ability to communicate are not necessary in order to sustain collusion, if firms are sufficiently patient.

This paper belongs to the fast growing literature of repeated games with private monitoring. The model we study here is a model where following each period, each player observes only his own payoff. In addition, given the pair of actions taken by the players, the payoffs are deterministic. We show that when we consider only minimaxing in pure strategies, all payoffs on the strictly efficient frontier can be obtained as sequential equilibria payoffs, in games where one can have as a continuation payoff at least one payoff that is internal to the set of feasible payoffs. We also show how such a continuation payoff can be obtained in all but a set of degenerate games.

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

In order to resolve the problem of the possibility of the punishing player to profitably deviate from the punishment, we instruct the punishing player to do exactly that, meaning to play with some (positive) probability the profitable deviations from the minimaxing. The punishing player will randomize between the minimaxing action and some profitable dviations from it. Of course, playing the profitable deviation has a higher payoff, so we will need to balance it with appropriate continuation payoffs.

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

Thus the punished player might be asked to play an action that is not his original "best response against the minimaxing". We wouldn't want to enforce any further punishments for a player who is currently being punished, so we make sure that all the actions the punished player is being instructed to use during the punishment are indeed best responses to the (now mixed) action taken by the punisher.

In order to obtain different continuation payoffs we shall need a two-dimensional set of continuation payoffs. In many games obtaining such a set is trivial. Such are games that have a one-period Nash-equilibrium whose payoff is an internal point of the set of payoffs; games where the efficient frontier is piecewise-linear (but not linear); and games where there exists a sub-matrix of payoffs and a distribution over that matrix such that any profitable deviation is detectable, and all actions in the support of the sub-matrix induce the same payoff to the player playing them. As for games where non of the three above hold, we show in the appendix how to obtain a two-dimensional set of payoffs for most of those remaining cases.

Having constructed the punishment phase and the continuation payoffs, we obtain the main result: all payoffs on the strictly efficient frontier can be achieved via sequential equilibria.

2 The Model

We consider a repeated discounted two-players game. Let $\{1,2\}$ be the set of players and A_1 and A_2 be the sets of actions available at each period to players 1 and player 2 respectively. Let $u_i(a_1, a_2)$ be the payoff for player i, $i \in \{1,2\}$ when player 1 plays action a_1 and player 2 action a_2. The payoffs are deterministic given the actions taken by the players. Let a_i^t denote the action taken by player i during period t. Let V be the convex hull of all feasible payoffs.

After each period, each player observes only his own payoff, so a player's private history consists of a stream of actions taken and payoffs observed. Formally, the private history of player i up to period t is:

$a_i^1, u_i(a_1^1, a_2^1), a_i^2, u_i(a_1^2, a_2^2), ..., a_i^t, u_i(a_1^t, a_2^t)$.

Normalize the payoffs of the game so that zero will be the minimax payoff of the players, when the minimaxing strategy is pure.

3 The Punishment Phase with Terminal Payoffs

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

We first consider a case where the punishing player can have different terminal payoffs following the punishment. We assume that the punishment lasts M periods after which the punisher receives his terminal payoff, and the game ends. The terminal payoffs will depend on the punished player, so that to obtain the different terminal payoffs for different actions, the punished player should be able to differentiate between those actions. In the sections that will follow, we show how such terminal payoffs are obtained as continuation payoffs in the infinitely repeated game, and how they are communicated.

Lemma 1: There exists a minimaxing for each player, such that the following conditions are fulfilled:

a. The player being punished is instructed to play only actions that are best-responses (and the payoff he obtains in all those best-responses is his minimax payoff, zero).

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

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

The proof of the lemma is constructive. The construction in some cases is rather technical, so we leave the full details to the appendix, and demonstrate the main ideas using Example 1.

Example 1

	a	b	c	d	e	f
A	$0, 0$	$0, 3$	$0, 2$	$0, 1$	$0, 3$	$0, 3$
B	$-1, 1$	$3, 2$	$2, 1$	$0, 1$	$-1, 0$	$3, 2$
C	$-3, 1$	$-3, 0$	$1, 3$	$1, 1$	$1, 1$	$-3, 1$
D	$0, -1$	$0, 0$	$0, 0$	$1, 0$	$0, 0$	$0, 0$

This example can be solved in four phases:

3.1 Phase 1 - The Original Minimaxing

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

We will instruct player 1 to play action A with a high probability $1 > P(A) \geq 1 - \varepsilon$, and action D with some small probability $0 < P(D) \leq \varepsilon$.

As we mentioned, these instructions make only a deviation of player 1 to action d detectable. What about the other deviations?

There are three profitable deviations in our example: b, e and f.

3.2 Phase 2: Adding a Second Column

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

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

Now, when the payoff of row B went up to zero, it became a best-response. When player 1 will play the new mixture of best-responses: A, B and D, he can differentiate between action action a and action b. In order to make payer 2 indifferent at period t of the punishment between playing a and b, player 1 will decrease the terminal payoff of player 2, whenever he plays B and observes 3, by:

$\frac{u(b)-u(a)}{P(B)\delta^{M-t}}$

where $u(a)$ and $u(b)$ denote the expected payoff for player 2 when he plays actions a and b respectively against the profile that player 1 is playing.

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

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

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

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

3.3 Phase 3: Adding a Third Column

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

Still, player 1 will prefer to play e over a.

Again, we can increase the probability of playing e, ζ, while decreasing the probability of playing action a until, in this case simultaneously, both the probability of action a decreases to zero, and the payoff of row C goes up to zero.

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

$\frac{u(b)-u(e)}{P(C)\delta^{M-t}}$

This makes player 2 indifferent between all his actions, while player 1 is being minimaxed and is using only best responses.

3.4 Phase 4: Another Column

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

The general claim that we detail and prove in the appendix is that at each phase, when there exists a profitable undetectable deviation of player 2, we can either:

(a) design continuation payoffs for player 2, based on the observations of player 1, that will make player 2 indifferent between all the actions that he is instructed to play, and that will make the currently profitable deviation not profitable anymore;

or:

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

That way, given terminal payoffs, a player can punish his opponent using a mixed action that has no profitable deviation, while the opponent is both using only best-responses and having a set of signals that is rich enough to allow the player being punished to design the appropriate terminal payoffs.

4 From Terminal Payoffs to Continuation Payoffs

In this section we shall describe how the terminal payoffs, used after punishments, can be obtained as continuation payoffs in the infinitely repeated game. We do that in three steps: in the first step we obtain an initial two-dimensional set of payoffs, using strategies such that any profitable deviation is detected with a positive probability. In addition, this set would be such that any payoff on the strictly efficient frontier Pareto dominates a non-empty subset of it. Let this set be W. In step 2 we describe how we use the set W as continuation payoffs. Finally, in step 3 we demonstrate how the players communicate the continuation payoffs.

4.1 Step 1: A Two-Dimensional Set of Payoffs with Observable Deviations, W

In this section we shall first construct a two-dimensional set of payoffs that can be achieved through strategies such that every profitable deviation from the strategy is detected with a positive probability. Then we will show how to enlarge this set so that any payoff on the strictly efficient frontier Pareto dominates a non-empty subset of it.

4.1.1 An initial two-dimensional set of payoffs with detectable deviations

For the following cases obtaining that initial set, denote it by W^-, is trivial[2]:

1. There is a one-period Nash-equilibrium whose payoff is internal to the set of payoffs.

2. There is a piecewise-linear (and not linear) efficient frontier.

3. There is a sub-matrix (of the payoff matrix), and distributions over the actions of player A and player B that are in the sub-matrix such that every profitable deviation is observable and the payoff for the different actions of the submatrix are equal (In fact, one period Nash-equilibrium is a private case of this condition).

Note that cases in which non of the above hold have some unique features. For example, from any pure action whose payoff is not on the efficient frontier (that is a single line) there is a profitable undetectable deviation. Hence, if we start at some point and follow the "path of un-observable deviations", we will end up at some point on the efficient frontier. In addition, there is no one-period Nash equilibrium equilibrium whose payoff is internal to the set of payoffs. In cases where both extreme points of the efficient frontier are one-period Nash equilibria, it is trivial to support the entire efficient frontier as a sequential equilibria payoffs, so we are interested only in games where this is not the case. We use those features to show in the appendix how one can attain an internal payoff point in all those cases where obtaining that internal payoff is not trivial, all but a very degenerate subset of them.

4.1.2 Enlarging the set W^-

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

[2]One case for which supporting the efficient frontier via sequential equilibrium is trivial is when the efficient frontier is a line but both his extremum points are one-period Nash equilibria.

4.2 Step 2: Using the set W

As we saw in the previous section, the observations of the player who is being punished during the punishment, induce changes in the continuation payoffs if the punishing player. Let the change which induces the lowest and the highest continuation payoff be X_{low} and X_{high} respectively. In addition, as will be explained in the next section, there will be at most three periods during which the continuation payoff is communicated. When designing the continuation payoff we need to take into account those three periods as well. Let G be the highest absolute value of all payoffs that appear in the payoff-matrix.

Let
$Y_{low} = (1-\delta^N)X_{low} + \delta^N(1-\delta^3)(-G)$
$Y_{high} == (1-\delta^N)X_{high} + \delta^N(1-\delta^3)(G)$

The continuation payoff will range between Y_{low} and Y_{high}.

We will use a subset of W in order to implement the continuation payoffs. Let the convex hull of $\{(v_1, v_2), (v_1', v_2), (v_1, v_2'), (v_1', v_2')\}$ be a subset of W such that $v_i' < v_i$ for $i \in \{1, 2\}$ and such that (v_1, v_2) is Pareto-dominated by the payoff on the efficient frontier for which we construct the equilibrium.

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

For reasons detailed in the next sub-section, we wish p to be bounded away both from 0 and 1. Hence there will be $\underline{p}$ and $\bar{p}$ such that $0 < \underline{p} < \bar{p} < 1$.

We will set the highest continuation, Y_{high} to be $\bar{v}_i = \bar{p}v_i + (1-\bar{p})v_i'$ and similarly Y_{low} will be $\underline{v}_i = \underline{p}v_i + (1-\underline{p})v_i'$.

There are three conditions that we will use later in our construction:

A player who deviates during the punishment of the opponent should be punished. In order to make such a punishment an effective one, we want the continuation payoff that follows being punished to be lower than the one following punishing. This is our first condition.

The second condition is that the weight of the minimaxing periods is small enough to enable balancing them using the set of the continuation payoffs mentioned above. If the weight of the punishment periods is too high, then the difference between the highest and the lowest continuation payoffs might not suffice for balancing the payoffs resulting from all possible randomizations during the punishment.

The third condition is that any profitable deviation from a strategy such that the continuation payoff Pareto-dominates W^- is not profitable if it is followed by a punishment. This means that the weight of the punishment should be high enough to eliminate the profit of one-period deviation.

As continuation payoffs following the punishment of player 1, we will use continuation payoffs that are close to the segment connecting $(v_1', \underline{v}_2)$ and $(v_1', \bar{v}_2)$. Respectively, as the continuation payoffs following the punishment of player 2 we will use continuation payoffs that are close to the segment connecting $(\underline{v}_1, v_1')$ and $(\bar{v}_1, v_2')$. This fulfils the first condition since $v_i' < \underline{v}_i$.

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

$(1-\delta^N)X_{high} + \delta^N(1-\delta^3)G + \delta^{N+3}\underline{v}_2 > (1-\delta^N)X_{low} + \delta^N(1-\delta^3)(-G) + \delta^{N+3}\overline{v}_2$

put in another way:

$(1-\delta^N)(X_{high} - X_{low}) + \delta^N(1-\delta^3)2G > \delta^{N+3}(\bar{v}_2 - \underline{v}_2)$

The same condition should hold for player 1's continuation payoffs. Those conditions hold for discount factor close enough to 1.

The punishment phase payoff and the continuation payoff for the player being punished, both are Pareto-dominated by the target payoff, that is on the strictly efficient frontier. We need to make sure that the continuation payoff of the equilibrium path for than target payoff Pareto-dominates the being punished as well. This is done by using an equilibrium path whose continuation payoffs are always within an epsilon distance from the target payoff, as done in Fudenberg and al (***).

4.3 Step 3: Communicating the continuation payoff

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

We will show how player 1 can transmit a message to player 2. Player 2 can convey a message to player 1 in an analogical way. We divide the discussion into the following three cases: there is at least one column such that there are two different efficient payoffs on that column; there is at least one row such that there are two different efficient payoffs on that row; and there are two different efficient payoffs in different columns and rows.[3]

4.3.1 case 1: two efficient payoffs in the same column

In this case, let that column be c'. Let the actions of player 1 which induce, combined with c', the first and the second efficient payoff to be $\hat{a}_1$ and $\hat{a}_2$ respectively. We will instruct player 2 to play column c', and player 1 to play either $\hat{a}_1$ or $\hat{a}_2$. The first payoff will convey the message that the continuation payoff for player 2 is v_2 and the second that it is v_2'. Since both actions induce an efficient payoff, non of the players can profit from deviating without being detected.[4]

The two messages induce different payoffs also for player 1. We want player 1 to be indifferent between the two messages. Let the first efficient payoff be $(u_1(\hat{a}^1, c'), u_2(\hat{a}^1, c'))$ and let the second be $(u_1(\hat{a}^2, c'), u_2(\hat{a}^2, c'))$ such that $u_1(\hat{a}^1, c') > u_1(\hat{a}^2, c')$. Then the continuation payoff in case player 1 played $\hat{a}^1$ will be $\left(v_1' - \frac{1-\delta}{\delta}(u(a_1 - u(a_2)), v_2\right)$ and in case he played $\hat{a}_2$ it is (v_1', v_2').

4.3.2 case 2: two efficient payoffs in the same row

In this case, transmitting a message will have two stages: first player 1 will communicate a message, and then player 2 will "echo" back the message to player 1. Let the set of columns where the highest payoff of payer 2 appears be $\{C_{max}\}$. We shall look, within the payoffs of the columns $\{C_{max}\}$ for the column with the highest payoff of player 2 that is not that maximal payoff (the "second best" within those columns). Let that column be c''. (Note that at least one payoff that is different from the maximal payoff appears at each column in the row that minimaxes player 2, so such a "second best" payoff must exist). Player 1 may have more than one action which gives

[3]obtaining the efficient frontier as sequential equilibrium payoff in the case where there is only one efficient payoff is trivial.

[4]even if the belief of player 2 is that with a probability 1 one of the messages will be transmitted, he still would have no incentive to deviate.

player 2, against c'', his highest payoff. We will allow player 1 to choose the action which gives him (player 1) the highest payoffs among those actions. Let the row that gives player 1 the highest payoff against c'' while inducing the highest payoff to player 2, and the one that gives player 1 the highest payoff against c'' while inducing that "second best" payoff to player 2, be r' and r'' respectively.

We instruct player 1 to play r' when he wants to convey the message that player 2's continuation payoff is v_2. We should keep $\underline{p}$, the lowest probability for v_2, high enough such that the following hold:

$\forall a_2' \in A_2$ such that $u_2(r'', a_2') < u_2(r', a_2')$,
$\underline{p}u_2(r', c'') + (1-\underline{p})u_2(r'', c'') > \underline{p}u_2(r', a_2') + (1-\underline{p})u_2(r'', a_2')$

Let that minimal $\underline{p}$ needed be $\underline{p}^*$.

A loss of information regarding the message may be revealed at the second stage, when player 2 is required to "echo" the message. However, in order to keep him from deviating and loosing that information, his belief regarding his continuation payoff should not be "too firm", that is, he shouldn't believe given any of his histories that the probability to observe one of the messages is "too close" to 1. This is because if he does believe that the message is that the continuation payoff is, for example, v_2, then he may deviate during the period when the message is being conveyed to an action that does not allow him to differentiate between the two messages, and then he will "echo" the message "v_2", being almost certain that he is correct. In order to avoid such a scenario we will keep the probabilities of the two continuation payoffs bounded away from 1.

Formally, $\overline{p}$ will need to be low enough such that:

$$(1-\delta)u_2(c'', r') + \delta\,(1-\overline{p})\left[0\delta^N + (1-\delta^N)v_2'\right] + \overline{p}v_2 <$$
$$(1-\delta)\left[\overline{p}u_2(c'', r') + (1-\overline{p})u_2(c', r'') + \overline{p}v_i + (1-\overline{p})v_2'\right]$$

If player 2 deviated and chose during the first period of the communication a strategy which did not allow him to differentiate between the two actions of player 1 (he "did not listen"), then only player 1 knows whether the message that is echoed back is the correct one. If it is the wrong message, then player 1 should punish player 2. Player 2 does not know whether his deviation was detected, and hence he does not know whether he is to be punished. We cannot instruct player 1 to simply begin the punishment, because if player

2 does not know that he is being punished, then he will not necessarily play the set of best-responses, and he will not be able to design the continuation payoffs for player 1. Therefore, player 1 does not have the proper incentives to randomize during the punishment. In order to solve this, we will have the one period following the communication to consist of player 2 obtaining his maximal payoff. In case player 2 did not echo the right message then player 1 should play for one period the (pure) action that minimaxes player 1, so that player 1 will learn that he should be punished. Note that since "not listening" is not profitable, then player 2 can have any belief about what will player 1 do in the next period. For player 1 to have an incentive to indeed play the minimaxing strategy, we need the continuation payoff that follows being punished to be strictly less than the one induced by punishing.

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

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

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

Example 2

	L	R
T	$\alpha, e(\alpha)$	$\alpha e(\alpha)^-$
B	$\beta, e(\beta)^-$	$\beta, e(\beta)$

We will show how player 1 can transmit a message to player 2 when the sub-matrix in example 2 appears.

We assume $\alpha > \beta$, hence $e(\alpha) < e(\beta)$ and $e(\alpha)^- < e(\beta)$. If player 2 will play R, then player 1 will be able to convey a message by playing either T or B. If the probability that the message will be such that player 1 will play B is kept high enough (regardless of history) then profitable undetectable deviations of player 2 will be to an action that gives him a payoff of $e(\beta)$ against B, and a payoff higher than $e(\alpha)^-$, say $e(\alpha)^{-+}$, paired with a payoff of α against action T[5].

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

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

Note that when player 2 deviates during the iterations above, he raises his own payoffs while keeping the payoffs of player 1 the same as before. The same goes for when we allow player 1 to choose the actions that give himself the highest payoffs while keeping the payoffs of player 2 the same. Hence, each iteration leads to a payoff that is the same as before for one player and strictly higher for the opponent. The payoff matrix is finite so the iterations have to end.

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

[5]player 2 cannot improve his payoff against action B without being detected. An improvement that is not detectable will have to be such that it gives a higher payoff against T. If the probability of B is high enough, then for a deviation to be profitable it has to first not decrease the payoff against B, and only second to improve against T.

5 References

Armstrong, M. (2005): "Recent Developments in the Economics of Price Discrimination", University of London, London.
Jin, Y.J. (1999): "Collusion with Private and Aggregate Information", working paper.

6 Appendix A - The Minimaxing Constructions

Here are the details of the algorithm which either makes profitable deviations of the punisher un-profitable or incorporates those deviations into the punishment. This is done while keeping the payoff of the player being punished at most zero, and while allowing the player being punished enough information based on his observations (while playing best-responses). The information is rich enough to enable the player being punished to construct continuation payoffs for the punishing player that will make him indifferent between all the actions that he is instructed to play.

6.1 The Algorithm

Assume that there is a set of columns that were already added to the support of the punishment, and there is a new column that is currently a profitable deviation, call it u'.

Basically, for each profitable deviation of the punisher, player 2, the algorithm decides on one of the two following possibilities. The first possibility is to create such continuation payoffs that will make that deviation not profitable. If creating such continuation payoffs is not possible, then the second possibility is at add the deviation to the support of the punishment, while either enlarging the set of best-responses or decreasing the number of actions player 2 is instructed to play during the punishment until the observations of player 1 will be informative enough to design continuation payoffs that will make the profitable deviation just as profitable to play just as profitable to play as the rest of the actions of player 2 in the support of the punishment.

If possible to build continuation payoffs - build them.

If not - remove columns or add rows as explained until there are appropriate payoffs.

If two rows are added simultaneously for the same column - run the algorithm twice - with one row and with the other. Keep the result of the iteration with the row that keeps getting to zero first, even with the perturbation of the column after.

Adjust continuation payoffs.

6.2 when are different continuation payoffs possible

Note that the continuation payoffs can be changed whenever player 1 plays a row and observes something, that is, they change on the form: "If I play action X and obtain payoff Y then I change the continuation payoff of player 2 by Δ_{XY}".

Let's enumerate all the different possible combinations of row and observation as 1,2,3,... And let $c(k)$ be the continuation payoff change when player 1 played+observed k. Then we can find appropriate continuation payoffs to make player 2 indifferent between all the columns he is supposed to play, iff we the following equalities can be solved $\sum I_{k\in j} c(k) = u(1) - u(j)$. That can be solved if the matrix $I_{k\in j}$ is regular[6].

6.3 how to remove a column in case different continuation payoffs are not possible

In case the above matrix is not regular, then there exist two sums with positive coefficients: $\sum_{k\in\subset N} \lambda_k I_{k\in j} = \sum_{k\in N\setminus K\subset N} \lambda_k I_{k\in j}$, such that both sides have the same continuation payoffs.

Then for each row, when we sum over the columns:

$\sum_{column j\in K} \lambda_j = \sum_{column j\in N\setminus K} \gamma_j$,

We sum over all rows:

$\sum_{rows} \sum_{column j\in K} \lambda_j = \sum_{rows} \sum_{column j\in N\setminus K} \gamma_j$

Change the order of the summation:

$\sum_{column j\in K} \sum_{rows} \lambda_j = \sum_{column j\in N\setminus K} \sum_{rows} \gamma_j$

In each column there are n different combinations of row+observation, so at each column in the matrix there are n 1's.

[6]In fact, we can make the profitable deviation strictly less profitable than the other actions in the support of the punishment by using in the equality of that column, call it column n, a value less than $u(1) - u(n)$

$n\sum_{column j\in K}\lambda_j = n\sum_{column j\in N\setminus K}\gamma_j$
$\sum_{column j\in K}\lambda_j = \sum_{column j\in N\setminus K}\gamma_j$

therefore, if we divide the coefficients of the two combinations by $\sum_{column j\in K}\lambda_j = \sum_{column j\in N\setminus K}\gamma_j$ we get two convex combinations of columns. Note that before the last column was added, we could find continuation payoffs that make player 2 indifferent between all the actions he is supposed to play. Given those continuation payoffs, the last column added is more profitable than all former columns. Hence, the convex combination which induces a positive probability to the new column is more profitable than the one that gives that column probability zero.

One side has current payoffs that are higher than those of the other side. Fix the continuation payoffs of the "better" side, and move, given the weights, the probabilities from the "bad" side to the "good" side, until at least one of the following occur:

(a) some row's payoff (for player 1) goes up to zero, and we add this row to the support of the punishment phase.

(b) some column's probability goes down to zero, and it is taken out of the support of the punishment phase.

If there is still linear dependency in the matrix that does not enable the design of the required continuation payoffs, then we will repeat the procedure of shifting probabilities from one side to another, until such continuation payoffs can be found.

Once they are found, a column that was removed from the support cannot become a profitable undetectable deviation (while keeping the probabilities of the rows fixed), since we can, instead of increasing the probability of the column that was taken out, assume without loss of generality that it is column 1, increase the probability of the following combination: $\frac{\sum\gamma_j u_j - \sum_{k\neq 1}\lambda_k u_k}{\lambda_1}$. In terms of continuation payoffs, it is equivalent to column 1, but it has better current payoffs.

The algorithm is finite since we may only add rows to the support, and for a given set of rows in the support, once we removed a column we don't add it back (no circles and a finite number of rows).

7 Appendix B - Obtaining an Internal Point in Other Matrices

7.1 A Simple Randomization with Continuation Payoffs

Consider the following matrix

	a_1	a_2	a_3
A_1	$-1,-1$	$-1,0$	$-1,0$
A_2	$0,-1$	$0,0$	$0,6$
A_3	$0,-1$	$0,6$	$6,0$

It has only one-period Nash-equilibrium, (A_3, a_2), with the payoff $(0,6)$, it's strictly efficient frontier is the line connecting $(0,6)$ and $(6,0)$, and from each common action which induces a payoff that is not on the efficient frontier there is an un-detectable profitable deviation (for example, from the common action (A_1, a_1) there is a profitable un-detectable deviation for both players - for player 1 to either A_2 or A_3, and for player 2 from a_1 to either a_2 or a_3). As for other sub-matrices, A_1 dominated by A_2 and A_3 and a_1 by a_2 and a_3. As for sub-matrices of the sub-matrix $\{A_2, A_3\} \otimes \{a_2, a_3\}$, within this sub-matrix, A_3 dominates A_2.

However, there is a way to obtain a payoff that is not on the strictly efficient frontier. We will instruct player 2 to play a_2, and player 1 to randomize between A_2 and A_3, all this for M periods. Player 2 cannot deviate without being detected, and for player 1 deviating to A_1 is not profitable. The payoffs for player 1 when he plays actions A_2 and A_3 are different. In order to make player 1 indifferent between playing those actions, they need to have different continuation payoffs.

The different continuation payoffs will take place after those M periods, and will be on the strictly efficient frontier. Since player 2, when playing a_2, can tell whether player 1 is playing A_2 or A_3, both players know what should the continuation payoffs be, and such continuation payoffs can be designed in a way that will make player 1 indifferent between A_2 and A_3. Of course, for every M there is a $\delta(M)$ such that for all $\delta \geq \delta(M)$ one can find proper continuation payoffs on the efficient line that will take place after M periods and that will make player 1 indifferent.

7.2 A Complicated Randomization with Continuation Payoffs

Consider the following matrix

	a_1	a_2	a_3
A_1	$-1,-1$	$-1,0$	$-1,0$
A_2	$0,-1$	$6,0$	$0,6$
A_3	$0,-1$	$0,6$	$6,0$

In this matrix as well, the only one-period Nash-equilibrium is on the efficient frontier, and there is a profitable un-detectable deviation from every sub-matrix that contains an internal point (from a distribution such that the players are indifferent between the actions in that sub-matrix). However, in this matrix, from every observable randomization over two actions for one player and one action of the opponent that induces a positive probability for a payoff not on the efficient frontier, there is a profitable undetectable deviation. For example, if we instruct player 2 to play a_2, and player 1 to randomize between A_1 and A_2, player 1 can deviate from A_1 to A_3.

There is still a way to obtain a payoff not on the efficient frontier in this matrix. We will use randomization and continuation payoffs on the efficient frontier. We will ask player 1 to randomize between all his actions (A_1, A_2 and A_3) and player 2 to randomize between a_2 and a_3. The players should have common knowledge regarding the continuation payoffs, so the different continuation payoffs are by the following division:

	a_2	a_3
A_1	x	x
A_2	x	y
A_3	z	x

The division is the "common knowledge" division. For example, when played a common action that falls under part x, both players know that this is the case (and both players know that both players know etc.).

We will show that proper continuation payoffs exist.

Let the probability for player 1 to play A_1, A_2 and A_3 to be p_1, p_2 and $1-p_1-p_2$ respectively, and the probability for player 2 to play a_1 and a_2 to be q and $1-q$ respectively.

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

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

(i) $(1-\delta)(-1) + \delta^M x_1 = (1-\delta)q6 + \delta^M [qx_1 + (1-q)y_1]$
or:
(i) $(1-q)\delta^M(x_1 - y_1) = (1-\delta)(1+q6)$
and:
(ii) $(1-\delta)(-1) + \delta^M x_1 = (1-\delta)(1-q)6 + \delta^M [qz_1 + (1-q)x_1]$
or:
(ii) $q\delta^M(x_1 - z_1) = (1-\delta)[1+(1-q)6]$

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

(iii) $(1-\delta)(1-p_1-p_2)6 + \delta^M [(p_1+p_2)x_2 + (1-p_1-p_2)z_2] =$
$(1-\delta)p_2 6 + \delta^M [p_2 y_2 + (1-p_2)x_2]$
or:
(iii)$(1-\delta)(1-p_1-2p_2)6 =$
$\delta^M [p_2 y_2 + (1-p_1-2p_2)x_2 - (1-p_1-p_2)z_2]$

put differently:
(i) $\Delta^1_{xy} = \frac{(1-\delta)}{\delta^M}\frac{[1+q6]}{1-q}$
(ii) $\Delta^1_{xz} = \frac{(1-\delta)}{\delta^M}\frac{[1+(1-q)6]}{q}$
(iii)$\frac{(1-a)}{\delta^M}(1-p_1-2p_2)6 = p_2\Delta^2_{yx} + (1-p_1-p_2)\Delta^2_{xz}$

Now, given that $\Delta^2_{yx} = -\Delta^1_{yx} = \Delta^1_{xy}$ and $\Delta^2_{xz} = -\Delta^1_{xz}$, we obtain:
(iii)$(1-p_1-2p_2)6 = p_2\frac{[1+q6]}{1-q} + (1-p_1-p_2)\left[-\frac{1+(1-q)6}{q}\right]$

For any given p_1 and p_2, when we take q to 1, the right hand side approaches infinity, while when we take it to zero, it approaches minus infinity, so far any given p_1 and p_2 we can find a q that solves this equation.