Repeated games with public information revisited

Marie Laclau, Tristan Tomala

Repeated games with public information revisited

Marie Laclau[*]and Tristan Tomala[†]

March 1, 2016

Abstract

We consider repeated games with compact actions sets and pure strategies in which players commonly observe a public signal which reveals imperfectly the action profile. We characterize the set of payoffs profiles that can be sustained by a perfect equilibrium, as players become increasingly patient. There are two conditions: admissibility and joint rationality. An admissibly feasible payoff can be achieved by an action profile that offers no unilateral deviation which is both undetectable and profitable. It is jointly rational if for all weights on players, the weighted payoff is greater than or equal to the minmax level of the weighted payoff function. This characterization is alternative to the one provided by the "score method" of Fudenberg and Levine (1994). We provide a simple construction of equilibrium strategies based on cooperation, punishments and rewards. Punishments rely on Blackwell's approachability algorithm.

[*] CNRS and Paris School of Economics, 75014 Paris, France, marie.laclau@psemail.eu.

[†] HEC Paris and GREGHEC, 78351 Jouy-en-Josas, France, tomala@hec.fr. The authors thank Ludovic Renou and Juuso Toikka for useful discussions and comments. Marie Laclau gratefully acknowledges the support of the French National Research Agency, under grant ANR CIGNE (ANR-15-CE38-0007-01) and through the program Investissements d'Avenir (ANR-10–LABX_93-0). Tristan Tomala gratefully acknowledges the support of the HEC foundation.

1 Introduction

Analyzing long-term relationship between rational agents is the task of the theory of repeated games with patient players. The existence of efficient equilibrium outcomes has been recognized quite a while ago, e.g. by Aumann and Shapley (1976, re-edited in 1994), and it is now common language to call "Folk Theorem" the statement asserting that in the long-term repeated game, all feasible and individual rational payoff profiles are sustained by equilibria. A seminal work is the one by Fudenberg and Maskin (1986) who proved the Folk Theorem for subgame perfect equilibria of discounted repeated games with high discount factor.

Those early works rested on the assumption that action profiles are publicly and perfectly observable. This restriction is quite demanding and imperfect observation (or monitoring) structures appear naturally in oligopoly or moral hazard models (see e.g. Stigler 1964, Green and Porter 1984). Most of the modern theory of repeated games with imperfect monitoring is built on the tryptic Abreu, Pearce and Stachetti (1990), Fudenberg and Levine (1994) and Fudenberg, Levine and Maskin (1994), who studied perfect public equilibria (henceforth PPE) of repeated games with public monitoring. Fudenberg and Levine (1994) defined the score in a direction of the payoff space, as the highest payoff in that direction that can be implemented by a Nash equilibrium of the one-shot game complemented with transfers contingent on public information. The score upper bounds the set of equilibrium payoffs in all directions, and using the dynamic programming methods of Abreu, Pearce and Stachetti (1990), it also characterizes the limit set of equilibrium payoffs as the discount factor grows. Fudenberg, Levine and Maskin (1994) obtain a Folk Theorem by finding conditions under which the set described by the score is the set of feasible and individually rational payoffs.

The goal of the present paper is to revisit repeated games with public monitoring using the approach of Fudenberg and Maskin (1986). That is, we give a characterization of limit equilibrium payoffs through two conditions which generalize feasibility and individual rationality. Then, we provide constructions of perfect equilibria made of cooperation, punishments and reward phases.

Preview of main results. More precisely, we consider PPEs of repeated games with pure strategies, compact actions sets and deterministic public signals. We first isolate two necessary conditions generalizing feasibility and individual rationality. An action profile is admissible if no player can profitably deviate by an undetectable deviation which leaves the public signal unchanged. A payoff profile is *admissibly feasible* if it is in the (closure of the) convex hull of payoffs generated by admissible action profiles. Next, for each positive weights system on players, we define the weighted minmax level as the minmax of the weighted sum of the players' payoff functions. The minimum is taken over admissible actions profiles, whereas the maximum is taken over the set of profiles of unilateral deviations which induce the same public signals across deviating players. A payoff profile is then said to be *jointly rational* if for any system of positive weights, the weighted payoff is no less that the weighted minmax level. We show that for each discount factor, any PPE payoff is admissibly feasible and jointly rational.

Second, under a non-empty interior condition, for any payoff profile v in the interior of the admissibly feasible and jointly rational set, we construct a PPE whose payoff is very close to v (for high enough discount factor). The construction can be sketched as follows. There is a (standard) main phase where the target payoff is approximated by a cycle of admissible action profiles. When the observed public signal is not the one prescribed, a punishment block starts. At the beginning of this phase, for each player i we compute the maximal payoff compatible with the observed signal and a unilateral deviation of this player. Each player for which there is such a unilateral deviation, is assigned a weight which is proportional to his maximal average gain from the deviation, relative to other players, and averaged over the past stages. The weighted minmax action profile is then played. The fact that this procedure actually punishes deviations is a consequence of Blackwell's (1956) approachability theorem. It remains to incentivize players to implement the punishment block. This is done through rewards as in Fudenberg and Maskin (1986). A punishment block during which a new deviation is recorded is followed by another one. When a punishment block where no deviation is recorded occurs, a compensation block is played where the payoffs are pushed upwards by an amount that is negatively proportional to the number of deviations recorded in the previous-to-last punishment block. This ensures that no player is willing to add an additional deviation within a punishment block. Finally, after the compensation block, a reward phase is played where the continuation payoffs of players who were not potential deviators is augmented by a bonus $\rho > 0$.

The nature of the contribution. In order to place our work within the literature, let us start by commenting on the main assumption of the paper: the consideration of pure strategies and deterministic signaling function. To a standard finite game with a stochastic signaling function, as in Fudenberg, Levine and Maskin (1994), we can associate the game whose compact actions sets are the mixed actions of the original game and the deterministic function maps a profile of mixed actions to the distributions of signals. Thus, our assumption is similar to the observability of mixed actions that was in Fudenberg and Maskin (1986) and Benoit and Krishna (1985). To motivate this assumption, think of the discounted game as a discretization of a game in continuous time over the time interval $[0, 1]$. The higher the discount factor, the quickest players change actions. In the continuous time limit, one can argue that it is impossible to observe which signal occurred at each time, but that the statistical distribution over a small time interval is accessible. On a technical side, the stochastic process of actions induced by a time dependent mixed action profile is ill-defined in continuous time. This problem disappears when one views mixed actions as deterministic pure strategies in a compact set. This is the approached followed in the recent literature on continuous time stochastic games (initiated by Neyman, 2012).

The equilibrium payoffs of the games with observable signal distributions or observable signal realizations are not comparable at the outset. Yet, we can show (see Section 4) that the set of admissibly feasible and jointly rational payoffs contains all the equilibrium (PPE) payoffs of the game with observable signal realizations. Thus, our conditions are necessary also without assuming observability of signal distributions. This latter assumption is important for the sufficiency part and in our equilibrium construction.

This being said, we view our contribution as two-fold. First, rather than insisting on finding sufficient conditions for the Folk Theorem, we display necessary conditions which show the limitations that an imperfect monitoring structure imposes on equilibrium payoffs. We attempt at finding conditions that are more transparently readable from the game and the monitoring structure, than those imposed by the score function. This effort is also found in Hörner et al. (2014) where an alternative definition of the score is given, through the dual of the optimization problem defining it. As a matter of fact, in a direction of the payoff space with all coefficients non-positive, one can see from Hörner et al. (2014) that the score corresponds to the weighted minmax level. In other directions, the score is more difficult to interpret.

Second, the sufficiency part relies on "simple" strategy constructions rather than on recursive methods. In a sense, our work revisits Fudenberg, Levine and Maskin (1994) with the lenses of Fudenberg and Maskin (1986). Notice that, even with our assumption of pure strategies and deterministic signals (i.e. observable signal distribution), the score method would apply routinely.

Our approach is similar in nature to the one used for time-average undiscounted games. Among others, Lehrer (1990) obtained a characterization of Nash equilibrium payoffs for signals with a product structure (semi-standard), Tomala (1998) considered undiscounted games with public signals and pure strategies, and Renault and Tomala (2004) characterized mediated communication equilibria. In these two latter papers were introduced the condition of joint rationality and punishments using Blackwell's approachability were designed. One should note that subgame perfectness is not an issue for undiscounted games, it is enough to restart the equilibrium strategies at very distant times to make sure that each player neglects the cost of punishing. The present paper aims at obtaining characterizations and equilibrium constructions that resemble those results, but which applies to perfect equilibria of discounted games. The difference can be seen in the construction using rewards, but also in the characterization. First, in equilibrium, action profiles have to be admissible, even off-path. Second, the notion of admissibility is more stringent for discounted games, than for undiscounted games (even in pure strategies). There, small gains from deviations can be tolerated, as long as they vanish to 0 in the long-run.

In a recent discussion paper, Sugaya (2016), obtains a characterization of perfect equilibria of repeated games with imperfect private monitoring and mediated communication. The project is similar to ours in motivation: characterize the limit set of equilibria and show the limitations imposed by a given monitoring structure. Sugaya shows how to amend the set obtained by Renault and Tomala (2004), in order to account for sequential rationality in the discounted game. The setting of Sugaya is more general than ours and does not assume observability of distributions of mixed actions. An important difference, leaving mediated communication aside, is that Sugaya's sufficiency part uses the score method and thus does not provide a strategy construction akin to the one of Fudenberg and Maskin (1986).

The paper is organized as follows. Section 2 describes the model, the main results are detailed in Section 3, and some concluding remarks are discussed in Section 4.

2 The model

Consider a stage game where the set of players is $N = \{1, \dots, n\}$, each player i has a set of actions X_i and a payoff function $u_i : X \to \mathbb{R}$ defined on the set of action profiles $X = \prod_j X_j$. All action sets are assumed to be non-empty and compact, each payoff function is assumed to be continuous. We are also given a continuous signaling function $f : X \to S$ which maps the set of action profiles to a compact set of signals S.

The repeated game is played as follows. At each stage $t = 1, 2, \dots,$ players choose actions simultaneously and if $x_t = (x_{i,t})_i$ is the action profile selected, the signal $s_t = f(x_t)$ is publicly announced. The game is repeated next period. Players discount payoffs at a common rate $\delta < 1$, so that if $\{x_t\}$ is the sequence of action profiles, player i's payoff is $\sum_t (1-\delta)\delta^{t-1} u_i(x_t)$.

The set of public histories at stage t is S^{t-1} and we denote $H = \cup_t S^{t-1}$ the set of all public histories. A *pure public strategy*, henceforth a strategy, for player i is a mapping $\sigma_i : H \to X_i$. A strategy profile $\sigma = (\sigma_i)_i$ induces a unique sequence of action profiles $\{x_t(\sigma)\}$ which in turns yields a payoff for player i denoted, $U_i^\delta(\sigma) = \sum_t (1-\delta)\delta^{t-1} u_i(x_t(\sigma))$. This δ-discounted game is denote Γ_δ. A *perfect public equilibrium* is a profile of strategies σ such that after every public history h, the profile of continuation strategies $\sigma(\cdot|h)$ is a Nash equilibrium of Γ_δ. We denote by $V(\delta)$ the set of perfect public equilibrium payoffs, that is the set of payoff vectors $(U_i^\delta(\sigma))_i$ with σ a perfect public equilibrium. This set is non-empty as soon as the stage game admits a Nash equilibrium, which we assume from now on. Our main interest is to characterize $V^* = \lim_{\delta \to 1} V(\delta)$.

This model is almost identical to the classical setup of Fudenberg et al. (1994) except that on one hand we allow for infinite action sets, and on the other hand we restrict our study to pure strategies. Note that, as far as pure strategies are concerned, the *public* qualification is superfluous. Indeed, when information is public, any pure strategy is equivalent to a public strategy, since actions are functions of the public history and of the strategy itself (see Mailath and Samuelson, 2006, Lemma 7.1.2., p. 229).

A particular case of interest is the one of a game with finite action sets, public signals, and observable signal distributions. Consider a repeated game with finite action sets A_i, finite set of signals R and endowed with a stochastic signaling function $\pi : \prod_i A_i \to \Delta(R)$ mapping action profiles to probability distributions over signals. In the corresponding "compact" game, the actions $X_i = \Delta(A_i)$ are the mixed actions in the underlying finite game, the payoffs are the expected payoffs, and the signaling function is the distribution of signals induced by a profile of mixed actions. Namely, $f(x) = s \in \Delta(R)$ with $s(r) = \sum_{r,a} \pi(r|a) \prod_i x_i(a_i)$.

3 The main results

We first describe two necessary conditions satisfied by all perfect public equilibrium payoffs.

3.1 Admissibility

A first necessary condition for a strategy profile to be a perfect public equilibrium is that no profitable and undetectable deviation is ever offered to players.

Definition 3.1. *An action profile x is i-*admissible *if for all $y_i \in X_i$,*

$$f(y_i, x_{-i}) = f(x) \Rightarrow u_i(y_i, x_{-i}) \leq u_i(x).$$

An action profile is admissible *if it is i-admissible for each player i in N.*

In particular, all Nash equilibria of the one-shot game are admissible. We denote $\mathcal{A}_i$ the set of i-admissible action profiles, $\mathcal{A} = \cap_i \mathcal{A}_i$ the set of admissible profiles and $\mathcal{V}$ the closure of the convex hull of payoffs associated to admissible action profiles, that is $\mathcal{V} = u(\overline{\text{co}}\,\mathcal{A})$. A payoff in $\mathcal{V}$ shall be referred to as an admissibly feasible payoff. The next lemma states that a perfect public equilibrium must induce admissible action profiles and a payoff vector in $\mathcal{V}$.

Lemma 3.2. *For each discount factor $\delta < 1$, if σ is a perfect public equilibrium, then for each public history h, $\sigma(h)$ is admissible. Therefore, $V(\delta) \subseteq \mathcal{V}$.*

Proof. Fix $\delta < 1$. If there exists a public history h, a player i and an action y_i such that $f(y_i, \sigma_{-i}(h)) = f(\sigma(h))$ and $u_i(y_i, \sigma_{-i}(h)) > u_i(\sigma(h))$, then player i can deviate at h and profit at the stage of deviation without affecting continuation play. This precludes σ from being a perfect public equilibrium. Moreover, given a perfect public equilibrium σ and an infinite history h, we then have $\overline{\sigma} = \sum_t (1-\delta)\delta^{t-1}\sigma(h_t) \in \overline{\mathrm{co}}\,\mathcal{A}$, and $U_i^\delta(\sigma) = u_i(\overline{\sigma})$ for each player $i \in N$. Hence, $V(\delta) \subseteq \mathcal{V}$. □

As a simple illustration, consider the following two-player prisoner's dilemma, with the signaling function given by the matrix on the right-hand side:

$$\begin{array}{c|c|c|} & C & D \\ \hline C & 3,3 & 0,4 \\ \hline D & 4,0 & 1,1 \\ \hline \end{array} \qquad\qquad \begin{array}{c|c|c|} & C & D \\ \hline C & s & s \\ \hline D & s & r \\ \hline \end{array}$$

with $s \neq r$. First, assume that this finite game is played in pure strategies. Clearly, (C, C) is not admissible, as player 1 (or player 2) can deviate to D and increase his payoff without affecting the public signal. As a consequence, $(3, 3) \notin \mathcal{V}$.

Second, consider the game played in mixed strategies with observable signal distribution: (C, C) is still not admissible. Now, for $x, y \in [0, 1]$, identify x (resp. y) with the mixed strategy of player 1 (resp. player 2) which plays C with probability x (resp. y). The distribution of signals, identified with the probability of r, is $f(x, y) = (1-x)(1-y)$. Thus, for any $y < 1$, $x \neq x' \Rightarrow f(x, y) \neq f(x', y)$ and similarly, for any $x < 1$, $y \neq y' \Rightarrow f(x, y) \neq f(x, y')$. If follows that any action profile (x, y) with $x < 1$ and $y < 1$ is admissible. The set of admissible action profiles is thus dense and the closure of the convex hull of the associated payoffs $\mathcal{V}$ is the convex hull of the four points $(3, 3), (0, 4), (4, 0), (1, 1)$.

We also gather from this example that the set of admissible action profiles need not be closed, which explains why $\mathcal{V}$ is defined as the *closure* of the convex hull of admissibly feasible payoffs.

3.2 Joint rationality

The second necessary condition resembles and generalizes the usual individual rationality condition. Clearly, an equilibrium payoff must satisfy individual rationality, since otherwise, a player would prefer to optimize myopically rather than abiding by the strategy profile. In absence of some identifiability condition granting that detecting a deviation entails identifying the deviator, individual rationality is not enough. If a detected deviation can be ascribed to several players and if it is not possible to punish those players simultaneously, then some feasible and individually rational payoffs, even efficient ones, cannot be obtained in any equilibrium. To see this, consider the following example.

Example 3.3. Consider a 3-player partnership game between two agents (players 1 and 2) and a principal (player 3). The principal gets a strictly positive payoff only if he hires the two agents and both of them work. He gets nothing if one agent shirks or if he hires only one of them. When both agents are hired, each of them prefers to shirk. An agent who is hired alone gets all the surplus to himself effortlessly. The payoff table is as follows:

	W	S
W	$w-c, w-c, B$	$w-c, w, 0$
S	$w, w-c, 0$	$w, w, 0$

Hire both (H_b)

	W	S
W	$b, 0, 0$	$b, 0, 0$
S	$b, 0, 0$	$b, 0, 0$

Hire 1 (H_1)

	W	S
W	$0, b, 0$	$0, b, 0$
S	$0, b, 0$	$0, b, 0$

Hire 2 (H_2)

where B is the profit of the principal when he hires the two agents and both work, w is the wage paid to each agent, c is the cost of effort and b is the net gain of either agent when he is hired alone. All parameters are positive. We assume $w > c$ and $B > \max\{2c, b\}$ so that (H_b, W, W) is the only surplus-efficient outcome. Observe that the set of feasible payoffs has non-empty interior and that the minmax level is 0 for each player, so that each feasible payoff is individually rational.

Assume that the public signal is made of the action and of the payoff of the principal. That is, in each period, it is announced who is hired, and when both are hired, it is known whether both worked or not.

We claim that any equilibrium payoff $v = (v_1, v_2, v_3)$ satisfies $v_1 + v_2 \geq \min\{2w, b\}$. As a consequence, if $2w - 2c < \min\{2w, b\}$, then the efficient outcome $(w-c, w-c, B)$ is not an equilibrium outcome of the repeated game, and equilibrium payoffs are bounded away from efficiency (for a numerical instance, take e.g. $w = 4$, $c = 3$, $B = 10$, $b = 3$).

To justify this claim, take σ a perfect public equilibrium of the repeated game with discount factor δ, and for $i = 1, 2$, define τ_i as the deviation of player i that always plays S. The sequence of public signals is the same under (τ_1, σ_{-1}) as under (τ_2, σ_{-2}). Thus, the sequences of actions of the principal are also the same under these two strategy profiles. For each action of the principal a_3, denote,

$$N_\sigma(a_3) = \sum\nolimits_t (1-\delta)\delta^{t-1}\mathbb{1}_{\{a_{3,t} = a_3\}},$$

the discounted average number of stages where a_3 is played under σ. We have that for any action of the principal a_3, $N_{\tau_1, \sigma_{-1}}(a_3) = N_{\tau_2, \sigma_{-2}}(a_3) := N(a_3)$. Now, $U_i^\delta(\tau_i, \sigma_{-i}) = wN(H_b) + bN(H_i)$, and from the equilibrium condition $v_i \geq U_i^\delta(\tau_i, \sigma_{-i})$. It follows that,

$$v_1 + v_2 \geq 2wN(H_b) + bN(H_1) + bN(H_2) \geq \min\{2w, b\},$$

since $N(H_b) + N(H_1) + N(H_2) = 1$. ◇

To pin down the relevant participation constraint, we need conditions ensuring that any detectable deviation be punishable, even when the deviator is not identifiable. To this end, we borrow tools from repeated games with incomplete information (Aumann and Maschler, 1995) and approachability theory (Blackwell, 1956). In two-player repeated games with lack of information on one-side, the uninformed party may want to punish the informed party in all possible states of the world. This can be done using an approachability strategy that pushes down the payoff of the informed player simultaneously in all states. The theorem of Blackwell (1956) characterizes the payoff vectors for which such a strategy exists. We are going to view the collectivity of players who design the public strategy profile as the uniformed party, and the deviating player as the informed party, the state being its identity. The derived approachability condition is precisely the participation constraint that we need.

We consider probability distributions $q \in \Delta(N)$ over the set of players, which can be seen as weights. For each such q and each action profile x, the set of *non-revealing action profiles at* (q, x) is:

$$NR(q,x) = \{y \in X : f(y_i, x_{-i}) = f(y_j, x_{-j}), \forall i, j \in \mathsf{supp}\, q\}.$$

This is the set of profiles of unilateral deviations that induce the same public signal, for each possible deviator i in the support of the distribution q. The *q-weighted minmax level* is defined by the following inf-max formula:[1]

$$\ell(q) = \inf_{x \in \mathcal{A}} \max_{y \in NR(q,x)} \sum\nolimits_i q_i u_i(y_i, x_{-i}).$$

Attach weight q_i to player i and consider the weighted payoff function $\sum_i q_i u_i(y_i, x_{-i})$ where y is a profile of "deviations" which does not reveal any information about the identity of the deviator besides that $q_i > 0$. The number $\ell(q)$ is the "minmax" of this weighted payoff function, where the infimum is over admissible profiles, and the maximum is over non-revealing action profiles.

Definition 3.4. *A payoff vector $v \in \mathbb{R}^I$ is* jointly rational *if for each $q \in \Delta(N)$:*

$$\sum\nolimits_i q_i v_i \geq \ell(q).$$

In other words, v is jointly rational if it is "individually rational" for each vector of weights q. We denote JR the set of jointly rational payoff vectors. The next lemma claims that every perfect public equilibrium payoff is jointly rational.

Lemma 3.5. *For each $\delta < 1$, $V(\delta) \subseteq JR$.*

[1] Since $\mathcal{A}$ may not be closed, the minimum may not be achieved. However, $NR(q, x)$ is compact, so the maximum is achieved.

Proof. Fix $\delta < 1$. Take a perfect public equilibrium σ and let v be the associated payoff. If $v \notin JR$, there exists $q \in \Delta(I)$ such that for every $x \in \mathcal{A}$, there exists $y(q,x) \in NR(q,x)$ with,

$$\sum_i q_i u_i(y_i(q,x), x_{-i}) > \sum_i q_i v_i.$$

For each player i in the support of q, let τ_i be the strategy such that for any public history h such that $\sigma(h) = x$, then $\tau_i(h) = y_i(q,x)$. Since for every public history h, $(\tau_i(h))_i \in NR(q, \sigma(h))$, the sequence of public signals induced by (τ_i, σ_{-i}) does not depend on the choice of i in the support of q. If we denote h^* this sequence, for each i in the support of q, (τ_i, σ_{-i}) induces the sequence of action profiles $\{(y_i(q, \sigma(h_t^*)), \sigma_{-i}(h_t^*))\}$. By construction, at each stage t, $\sum_i q_i u_i(y_i(q, \sigma(h_t^*)), \sigma_{-i}(h_t^*)) > \sum_i q_i v_i$, and averaging over time yields $\sum_i q_i U_i^\delta(\tau_i, \sigma_{-i}) > \sum_i q_i v_i$. Therefore, there exists i such that $U_i^\delta(\tau_i, \sigma_{-i}) > v_i$, which contradicts that σ is an equilibrium. □

Now, let us discuss the relationship between joint rationality and individual rationality. The minmax level of player i is $w_i = \min_{x_{-i}} \max_{x_i} u_i(x_i, x_{-i})$, and a payoff vector v is individually rational if $v_i \geq w_i$ for each player i.

Lemma 3.6. *Any jointly rational payoff is individually rational.*

Proof. Consider $q = \epsilon_i$ the dirac measure on player i. Clearly, $NR(\epsilon_i, x) = X_i$ and therefore, $\ell(\epsilon_i) = \inf_{x \in \mathcal{A}} \max_{y_i \in X_i} u_i(y_i, x_{-i})$. Thus, if v is jointly rational, then for each player i, $v_i \geq \inf_{x \in \mathcal{A}} \max_{y_i \in X_i} u_i(y_i, x_{-i}) \geq w_i$. □

The latter inequality says more. Due to imperfect monitoring, it may not be possible to punish player i down to his minmax level within a perfect public equilibrium, even when he is identified as a deviator. The reason is that, in a perfect public equilibrium, only admissible action profiles are used, even off the equilibrium path.

Example 3.7. Consider the following two-player game:

	L	M	R
T	$4,4$	$2,2$	$5,3$
M	$2,2$	$1,1$	$4,2$
B	$3,5$	$2,4$	$6,6$

with the signaling function,

	L	M	R
T	s	s	r
M	s	s	r
B	w	w	z

The action M of player 1 is strictly dominated by T, and T, M are *indistinguishable* in that $f(T, a_2) = f(M, a_2)$ for any action a_2 of player 2. By symmetry of the game, the only admissible action profiles are $\{T, B\} \times \{L, R\}$. It follows that $\ell(\epsilon_1) = \min_{x \in \mathcal{A}} \max_{y_1} u_1(y_1, x_2) = 4$, while $\min_{x_2} \max_{x_1} u_1(x_1, x_2) = 2$. $\diamond$

The previous example shows that the definition of individual rationality level has to be adapted to account for imperfect monitoring.

Definition 3.8. *The* admissible minmax level of player *i is:*

$$w_i^* = \inf_{x \in \mathcal{A}} \max_{y_i \in X_i} u_i(y_i, x_{-i}) \geq w_i.$$

A payoff vector v is admissibly individually rational *if for every player i,* $v_i \geq w_i^*$*.*

Note that the admissible minmax level is of different nature than the various *effective* minmax levels found in the literature (see e.g., Wen, 1994; Fudenberg et al. 2007). Effective minmax levels are adaptations required to tackle games with payoffs that do not satisfy full-dimensionality. In the example, full-dimensionality is satisfied, and the modification of minmax levels is required by purely informational considerations. Summarizing this discussion, we have the following lemma.

Lemma 3.9. *Any jointly rational payoff vector is admissibly individually rational. Therefore, any perfect public equilibrium payoff is admissibly individually rational.*

This lemma together with Example 3.7 provide a simple example of discontinuity of the equilibrium payoff set as δ reaches 1. For each $\delta < 1$, payoffs have to be admissibly individually rational, whereas with no discounting, all admissibly feasible and individually rational payoffs can be obtained.

Example 3.7 has another feature that is worth commenting: this very signaling function allows to identify the deviator. This is the case when the signaling function has the *product structure.* Precisely, this is when for each player i, there is a function $f_i : X_i \to S_i$ such that $f(x) = (f_i(x_i))_i$. In other words, some public information is released about the action of each player, and the public information about player i's action does not depend on the actions of other players. Under this assumption, when a deviation is detected, the deviator is identified. Indeed, if $i \neq j$, it is not possible to have $f(y_i, x_{-i}) = f(y_j, x_{-j}) \neq f(x)$. The set of jointly rational payoffs is then easily computed. Let $\mathcal{V}^* = \mathcal{V} \cap JR = u(\overline{\text{co}}\,\mathcal{A}) \cap JR$.

Lemma 3.10. *If the signaling function has the product property, then $\mathcal{V}^*$ is the set of payoffs in $\mathcal{V}$ which are admissibly individually rational.*

Proof. Take $v \in \mathcal{V}$ an admissibly individually rational payoff vector and consider $q \in \Delta(I)$ whose support contains at least two points. The product property implies that for each action profile x and each y in $NR(q,x)$, $f(y_i, x_{-i}) = f(y_j, x_{-j}) = f(x)$, for all i, j in the support of q. If x is admissible, then $u_i(y_i, x_{-i}) \leq u_i(x)$ and therefore,

$$\ell(q) = \inf_{x \in \mathcal{A}} \max_{y \in NR(q,x)} \sum\nolimits_i q_i u_i(y_i, x_{-i}) = \inf_{x \in \mathcal{A}} \sum\nolimits_i q_i u_i(x).$$

Now, each payoff vector $v \in \mathcal{V}$ is arbitrarily close to a convex combination of the form $\sum_k \lambda_k u(x_k)$ with x_k admissible. Thus, $\sum_i q_i v_i$ is arbitrarily close to $\sum_k \lambda_k \sum_i q_i u_i(x_k)$ which is no less than $\ell(q)$. As a consequence, $\sum_i q_i v_i \geq \ell(q)$. □

To conclude this set of remarks on joint rationality, let us say a few words about two-player games. A reasonable guess, albeit wrong, is that in a two-player game, $\mathcal{V}^*$ is the set of admissibly feasible payoffs which are admissibly individually rational. The intuition supporting this guess is the following: in the case where a deviation is detected, each player plays a minmax strategy against the other. The problem is that, although the mutual minmax is feasible in any two-player game, the admissible mutual minmax might not be. In other words, there may not exist a strategy profile which is admissible and such that both players receive at most their admissible mutual minmax. Another interpretation of this phenomenon pertains to the public nature of the strategies. Public signals might indicate a deviation but not the identity of the player. Suppose that player 1 deviates once and conforms afterwards. If the signal is compatible with a deviation of player 2, then in a perfect public equilibrium, after this history, both player act independently of the identity of the deviator. Yet, both players would be able to compute it, were they allowed to recall private past actions. This is illustrated by the following example.

Example 3.11. Consider the following two-player game:

	A_2	B_2	C_2
A_1	$4, 4$	$1, 5$	$6, 0$
B_1	$5, 1$	$0, 0$	$0, 0$
C_1	$0, 6$	$0, 0$	$2, 2$

with the signaling function:

	A_2	B_2	C_2
A_1	s	r	r
B_1	r	r	r
C_1	r	r	s

The set of admissible action profiles is $\mathcal{A} = \{(A_1, A_2); (C_1, C_2); (A_1, B_2); (A_2, B_1)\}$. On one hand, any unilateral deviation from either (A_1, A_2) or (C_1, C_2) changes the signal, on the other hand, (A_1, B_2) and (A_2, B_1) are Nash equilibria of the stage game. For any other action profile, the signal is r and there is a profitable deviation for either player that does not change the signal. Therefore, $\mathcal{V}$ is the convex hull of the payoff vectors $(4, 4)$, $(5, 1)$, $(1, 5)$ and $(2, 2)$. The minmax (either regular or admissible) is 1 for each player. We claim that any perfect public equilibrium payoff $v = (v_1, v_2)$ satisfies $v_1 + v_2 \geq 6$. This implies that the admissibly individually rational payoff $(2, 2)$ is not an equilibrium payoff and that joint rationality constraints have to be considered to characterize perfect public equilibrium payoffs.

To justify this claim, fix $\delta < 1$ and consider σ a perfect public equilibrium. Necessarily, $\sigma(h) \in \mathcal{A}$ for each history h. Let τ_i be the deviation of player i who plays B_i when $\sigma(h) = (A_1, A_2)$, A_i when $\sigma(h) = (C_1, C_2)$, and conforms with σ otherwise. The sequence of signals is $(r, r, \dots)$ both under (τ_1, σ_{-1}) and (τ_2, σ_{-2}). Denote x^*_t the action profile induced by σ after the history $r, \dots, r$ in which only r appears and $y^*_{i,t}$ the action of player i at stage t induced by (τ_i, σ_{-i}). The point is that the sequence $\{x^*_t\}$ is the same under both deviations and that for each t, $u_1(y^*_{1,t}, x^*_{-1,t}) + u_2(y^*_{2,t}, x^*_{-2,t}) \geq 6$. It follows that $U^\delta_1(\sigma) + U^\delta_2(\sigma) \geq U^\delta_1(\tau_1, \sigma_{-1}) + U^\delta_2(\tau_2, \sigma_{-2}) \geq 6$. $\diamond$

3.3 The characterizations

Our main result is the following.

Theorem 3.12. *Each payoff vector in the interior of $\mathcal{V}^*$ is arbitrarily close to a perfect public equilibrium payoff of the repeated game when the discount factor is high enough.*

The proof is constructive, we adapt the construction of Fudenberg and Maskin (1986) with a normal phase, punishments and rewards, the main differences being seen in the punishment phase. An informal description is as follows.

Take a payoff profile v in the interior of $\mathcal{V}^*$ such that $v' = v - 10\varepsilon$ is also in the interior of $\mathcal{V}^*$.

(i) The main phase consists in a cycle whose repetition achieves a payoff ε-close to v. This phase continues as long as the prescribed signals are observed. When an unexpected signal appears, the punishment phase starts.

(ii) The punishment phase consists of two parts.

    (a) First, the following approachability algorithm is used. If x_t is supposed to be played at stage t and s_t is the observed signal, then for each player i we let:

$$u^*_i(x_t, s_t) = \max\{u_i(y_i, x_{-i,t}) : f(y_i, x_{-i,t}) = s_t\}$$

be the maximal payoff of player i compatible with the observed signal and a unilateral deviation (it is set as v'_i if player i has no such unilateral deviation).

Denote $\bar{u}^*_{i,t}$ the average of the $u^*_i(x_t, s_t)$ from the start of the punishment phase up to stage t. We define a weight for each player i by,

$$q_{i,t} = \frac{\max\{\bar{u}^*_{i,t} - v'_i, 0\}}{\sum_{i \in N} \max\{\bar{u}^*_{j,t} - v'_j, 0\}}.$$

Then, x_{t+1} is chosen such that

$$\sum\nolimits_i q_{i,t} v'_i \geq \max_{y \in \mathcal{A}} \sum\nolimits_i q_{i,t} u_i(y_i, x_{-i,t+1}).$$

This choice is possible since v' is in the interior of JR.

This algorithm is derived from Blackwell's approachability strategy. It ensures that after a finite number T of stages, the payoff of each player is no more that $v_i - 9\varepsilon$. This algorithm is repeated identically for a number Q of blocks of T stages, even if there are additional deviations during its execution. If there are such deviations, after finishing the Q blocks of punishment, a new sequence of Q punishment blocks starts, where the set of punished players is initialized to be the set of the potential deviators at the *last* deviation.

The reason for repeating the algorithm identically Q times, is that when a player deviates once, he may profit from the deviation at the remaining stages of the current block only, and not at the stages of the subsequent blocks.

(b) When a block of punishment (containing Q blocks) is completed without recording any deviation, a compensation block is played. Its purpose is to link the payoff of each player who was punished in the last Q blocks, with the number of deviations recorded in the previous-to-last Q blocks (by construction, there was no deviation in the last Q blocks): the higher the number of deviations, the lower the payoff during the compensation block.

As for the punishment blocks, the compensation block is completed even if additional deviations are recorded. It also consists of a repetition of independent copies (R blocks of length T). Deviating in the compensation block triggers new punishments.

Intuitively, being the first to deviate during a punishment block will make the punishment longer (a new punishment block is added), and making an additional deviation will lower the payoff obtained in the compensation block.

(iii) After punishments and compensation, the continuation payoffs of players who were not potential deviators is augmented by a bonus $\rho > 0$.

This rewarding system is similar to the one of Fudenberg and Maskin (1986) and classically provides incentives to implement the punishments.

The proof consists simply in verifying that no player has a profitable one-shot deviation. The details are relegated to the appendix.

4 Concluding remarks

4.1 Folk Theorem

Our main result can be used to obtain sufficient conditions for a Folk Theorem. Suppose that no player has undetectable deviations:

$$\forall i, \forall x, \forall y_i, f(y_i, x_{-i}) = f(x_i, x_{-i}) \Rightarrow y_i = x_i.$$

Suppose also that deviators are always identified:

$$\forall i, j, \forall x, \forall y_i, y_j, f(y_i, x_{-i}) = f(y_j, x_{-j}) \Rightarrow f(y_i, x_{-i}) = f(y_j, x_{-j}) = f(x).$$

Then for any payoff function, $\mathcal{V}^*$ is the set of feasible and individually rational payoffs. This is a simple and straightforward consequence of the definitions of admissibility and joint rationality (the proof of Lemma 3.10 applies, with the additional property that all actions profiles are admissible). From the examples in Section 3, it is easily seen that these conditions are necessary. If they are not satisfied, then we can find a payoff function for which the Folk Theorem fails.

4.2 Finite games and mixed strategies

Consider a repeated game with finite action sets A_i, finite set of signals R and with a stochastic signaling function $\pi : \times_i A_i \to \Delta(R)$ as in Fudenberg et al. (1994). Let $E^*(\delta)$ be the set of payoffs associated to public perfect equilibria in mixed strategies of the δ-discounted game. Our approach gives an *upper bound* on this set. Precisely, let us associate to these data a "compact" game, in which the action sets are $X_i = \Delta(A_i)$, that is the mixed actions of the underlying finite game, payoffs are the expected payoffs, and the signaling function $f : X \to \Delta(R)$ is the distribution of signals induced by a profile of mixed actions. Then:

Claim 4.1. *For each δ, $E^*(\delta)$ is a subset of the admissible and jointly rational payoffs $\mathcal{V}^*$ of the associated compact game.*

The proof of this claim is a straightforward adaptation of Lemmas 3.2 and 3.5. For Lemma 3.2, if $\sigma(h)$ is not admissible, then there exists some player who can deviate profitably after history h, while inducing the same *distribution* of signals. Thus, continuation payoffs are unaffected and the deviation is profitable in the repeated game. For Lemma

3.5, we consider the same deviations τ_i for each player i in J, and argue that they all induce the same *distribution* of signals and thus the same responses of the public strategy profile. The rest of the proof goes through.

An important open issue is either to find conditions ensuring that $E^*(\delta)$ converges to $\mathcal{V}^*$ or to give an analogous characterization of the limit. Consider a repeated prisoner's dilemma with two possible public signals such that the probability of the high signal is $p < 1$ when both cooperate, and is $0 < q < p$ otherwise. It is known that perfect public equilibrium payoffs are bounded away from efficiency (see Radner et al., 1986; Mailath and Samuelson, 2006, section 7.2.3). Yet, if the distribution of signals is observable, cooperation is possible, that is, $\mathcal{V}^*$ is the set of feasible and individually rational payoffs. Thus, the observation of realized signals imposes further restrictions on equilibrium payoffs.

Of course, the correct restrictions are those imposed by the score. Our work, combined with the dual approach of Hörner et al. (2014), gives an interpretation of the score in negative directions. Finding similar interpretations in other directions is left for future research.

References

[1] D. Abreu, D. Pearce, and E. Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, **58**, 1041–1063, 1990.

[2] R.J. Aumann and L. S. Shapley. *Long-term competition—A game theoretic analysis.* 1976; reedited in: Aumann R.J., Collected Papers, Volume 1,*Essays on game theory*, pages 1–15, MIT Press, 1994.

[3] R.J. Aumann and M. B. Maschler. *Repeated games with incomplete information.* 1960's; reedited with Stearns, R, MIT Press, 1995.

[4] J.P. Benoît, and V. Krishna. Finitely repeated games. *Econometrica*, **53**, 905-922, 1985.

[5] D. Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, **6**, 1–8, 1956.

[6] D. Fudenberg and D. K. Levine. Efficiency and observability with long-run and short-run players. *Journal of Economic Theory*, **62**, 103–135, 1994.

[7] D. Fudenberg, D. K. Levine, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, **62**, 997–1039, 1994.

[8] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, **54**, 533–554, 1986.

[9] D. Fudenberg, D.K. Levine and S. Takahashi. Perfect public equilibrium when players are patient. *Games and Economic Behavior*, **61** (1), 27–49, 2007.

[10] E. J. Green and R. H. Porter. Noncooperative collusion under imperfect price information. *Econometrica*, **52**, 87–100, 1984.

[11] J. Hörner, S. Takahashi and N. Vieille. On the limit perfect public equilibrium payoff set in repeated and stochastic games. *Games and Economic Behavior.* **85**, 70–83. 2014

[12] E. Lehrer. Nash equilibria of n player repeated games with semi-standard information. *International Journal of Game Theory*, **19**, 191–217, 1990.

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

[14] A. Neyman. Continuous time stochastic games. Working paper, The Hebrew University of Jerusalem, 2012.

[15] R. Radner, R. Myerson and E. Maskin. An example of a repeated partnership game with discounting and with uniformly inefficient equilibria. *The Review of Economic Studies*, **53** (1), 59–69, 1986.

[16] J. Renault and T. Tomala. Communication equilibrium payoffs of repeated games with imperfect monitoring. *Games and Economic Behavior*, **49**, 313–344, 2004.

[17] G. Stigler. A theory of oligopoly. *Journal of Political Economy*, **72**, 44–61, 1964.

[18] T. Sugaya. The Characterization of the Limit Communication Equilibrium Payoffs Set with General Monitoring. Working paper, 2016.

[19] T. Tomala. Pure equilibria of repeated games with public observation. *International Journal of Game Theory*, **27**(1), 93–109, 1998.

[20] Q. Wen. The “folk theorem" for repeated games with complete information. *Econometrica*, **62**(4), 949-954, 1994.

A Proof of Theorem 3.12

Fix a payoff vector v in the interior of $\mathcal{V}^*$. For each subset of players $J \subseteq N$, denote v^J the payoff vector such that $v^J_i = v_i + \rho$ if $i \in J$, and $v^J_i = v_i$ otherwise, with $\rho > 0$ a positive reward such that each v^J is in the interior of $\mathcal{V}^*$. We choose an $\varepsilon > 0$ such that for each $J \subseteq I$, the ball with center v^J and radius 10ε is included in the interior of

$\mathcal{V}^*$. The vectors v, v^J can be approximated arbitrarily closely by convex combinations of admissible payoffs with rational coefficients. So without loss of generality, we assume that there exists an integer T such that $v = \sum_k \frac{T_k^0}{T} u(x_k^0)$, with $\sum_k T_k^0 = T$, $x_k^0 \in \mathcal{A}$ $(\forall k)$, and for each $J \subseteq N$, $v^J = \sum_k \frac{T_k^J}{T} u(x_k^J)$, with $\sum_k T_k^J = T$, $x_k^J \in \mathcal{A}$.

The following piece of notation is going to be useful. Given an action profile x and a signal s, denote $D(x,s) = \{i \in N : \exists y_i \in X_i, f(y_i, x_{-i}) = s\}$ the set of players that are able to induce the signal s by a unilateral deviation from x. When x is the action profile prescribed by the strategy and $s \neq f(x)$, then a deviation is detected and $D(x,s)$ is the set of *potential deviators*. We denote $J(x,s) = N \setminus D(x,s)$ the set of *innocents*.

We turn now to the construction of the strategy which has three kind of phases: Normal phases NORM(w) for any $w \in \{v, v^J : J \subseteq I\}$, a punishment phase and a reward phase.

Normal phase. For a vector $w = \sum_k \frac{T_k}{T} u(x_k)$, the normal phase NORM$(w)$ consists in playing a T-periodic sequence of action profiles where for each k, x_k is played T_k times within a period.

At stage 1, we start the main phase NORM(v), which defines a periodic sequence of action profiles x_t (and similarly for the phase NORM(w) which starts at some stage t). Denote $\underline{s}_t = f(x_t)$ the expected signal at stage t when x_t is played and s_t the signal actually observed. As long as $s_t = \underline{s}_t$, the main phase continues to the next period. If there is a first stage t^* such that $s_{t^*} \neq \underline{s}_{t^*}$, then the punishment phase starts at stage $t^* + 1$. The set of suspects is initialized to $D_{P_1} = D(x_{t^*}, s_{t^*})$.

Punishment phase. This phase is decomposed in two parts, first a sequence of one or several blocks P_1, P_2, etc, of stages of length P with P an integer; then, a compensation block which aims at giving each suspected player a payoff that depends on the number of deviations that occurred in the previous-to-last block.

(i) Punishment blocks. Each block P_k consists of Q blocks of stages of length T with $P = QT$, Q being an integer. We denote $\underline{x}_t$ the action profile prescribed at stage t during this phase and s_t the signal actually observed at stage t. The sequence $\{\underline{x}_t\}$ is defined recursively. The set of deviators of the first punishment block P_1 is $D_{P_1} = D(x_{t^*}, s_{t^*})$ (the set of deviators at blocks P_k for $k > 1$ are defined below). Take any block $q \in \{1, \ldots, Q\}$. The first action profile $\underline{x}_{t^*+1}$ is chosen arbitrarily in $\mathcal{A}$. Suppose that $\underline{x}_\ell$ has been defined for all $\ell = t^* + (q-1)T + 1, \ldots, t^* + (q-1)T + t$, with $t < T$, and that the signal s_ℓ has been observed at stage ℓ. For each player i and each stage $\ell = t^* + (q-1)T + 1, \ldots, t^* + (q-1)T + t$,

we define the *maximal payoff of player $i \in N$ at stage ℓ* as:

$$u_i^*(\underline{x}_\ell, s_\ell) := \begin{cases} \max_{y_i}\{u_i(y_i, \underline{x}_{-i,\ell}) : f(y_i, \underline{x}_{-i,\ell}) = s_\ell\}, & \text{if } i \in D_{P_1} \text{ and } \exists y_i \text{ s.t.} \\ & f(y_i, \underline{x}_{-i,\ell}) = s_\ell\,; \\ v_i - 10\varepsilon, & \text{otherwise.} \end{cases}$$

This is the maximal payoff of player i compatible with the observed signal s_ℓ and a unilateral deviation. Compute for each player i, the average maximal payoff $\bar{u}_{i,t}^* = \frac{1}{t}\sum_{\ell=1}^{t} u_i^*(\underline{x}_\ell, s_\ell)$ from the start of the punishment bock q up to stage t.

- If $\bar{u}_{i,t}^* \leq v_i - 10\varepsilon$, for each player i, then $\underline{x}_{t+1}$ is chosen arbitrarily in $\mathcal{A}$.

- Otherwise, define a vector of weights $q \in \Delta(I)$ by:

$$q_i = \frac{\max\{\bar{u}_{i,t}^* - (v_i - 10\varepsilon); 0\}}{\sum_{j \in J} \max\{\bar{u}_{j,t}^* - (v_j - 10\varepsilon); 0\}},$$

  and choose $\underline{x}_{t+1} \in \mathcal{A}$ such that:

$$\sum\nolimits_{i \in J} q_i(v_i - 10\varepsilon) \geq \max_{y \in NR(q, \underline{x}_{t+1})} \sum\nolimits_i q_i u_i(y_i, \underline{x}_{-i,t+1}).$$

Such an action profile exists from joint rationality and since the ball around v with radius 10ε lies in the interior of JR. This algorithm is Blackwell's approachability strategy.

At the stage $t^* + T$, the first punishment block $q = 1$ ends. The same strategy is played for Q consecutive blocks and is restarted at the beginning of each block. At the stage $t^* + P$, that is after Q punishment blocks, the block P_1 ends. The strategies for blocks P_k, for $k > 1$, are defined similarly with an updated set of potential deviators as defined below. The following rule applies:

(i) If there was no deviation during the punishment block P_1, *i.e.* for every $t \in \{t^* + 1, \ldots, t^* + P\}$, $s_t = f(\underline{x}_t)$, then go to the compensation phase COMP$(0, D_{P_1})$ defined below: this is the compensation phase for the case there was 0 deviation during the initial punishment block P_1 in which the initial set of suspects was D_{P_1}.

(ii) If there was some deviation during the punishment block P_1, consider the last deviation $\hat{t} = \max\{t : t \in \{t^* + 1, \ldots, t^* + P\} \text{ and } s_t \neq f(\underline{x}_t)\}$. Then start a new punishment block P_2 at stage $t^* + P + 1$ where the set of suspects is initialized to $D_{P_2} = D(x_{\hat{t}}, s_{\hat{t}})$.

(iii) From (ii), this defines several punishments blocks denoted P_1 from stages $t^* + 1$ to $t^* + P$, and for each $k > 1$, P_k from stages $t^* + (k-1)P + 1$ to $t^* + kP$. For each $k > 1$, denote by D_{P_k} the initialized set of suspects at the punishment block P_k.

If there exists a first $k > 1$ such that there is no deviation in P_k, *i.e.* for every $t \in \{t^* + (k-1)P + 1, \ldots, t^* + kP\}$, $s_t = f(\underline{x}_t)$, then go to the compensation phase COMP(n_{k-1},D_{P_k}) with n_{k-1} the number of deviations at previous-to-last block P_{k-1}, that is $n_{k-1} = \sharp\{t : t \in \{t^* + (k-2)P + 1, \ldots, t^* + (k-1)P\}$ and $s_t \neq f(\underline{x}_t)\}$. This phase COMP($n_{k-1}$,$D_{P_k}$) described below is the compensation phase for the case there was k punishments blocks, there was n_{k-1} deviations in the previous-to-last punishment block, and the last set of potential suspects was D_{P_k}.

Notice that the play does not reach the compensation phase until a full punishment block P_k, for some k, is completed without any deviation. This ensures that suspected players are effectively punished.

(ii) Compensation block. The previous description implies that a punishment block has to be completed before starting a new one, even if there are some deviations in the current block (and similarly, within each punishment block P_k, each block $q \in \{1, \ldots, Q\}$ is completed before starting a new one). To prevent players from deviating after a deviation has been detected in a current punishment block P_k for some k, we add a compensation block such that the payoff of each player in the set of suspects of the last punishment block, depends on the number of deviations at the previous-to-last blocks (by construction, there was no deviation in the last block). The higher the number of deviations, the lower the payoff during the compensation block. Hence, intuitively, deviating during a punishment block either makes the punishment longer (a new punishment block is added) or lowers the payoff obtained in the compensation block.

More precisely, for each $k > 0$, we define the compensation block COMP(n_{k-1}, D_{P_k}), with $n_0 = 0$, which lasts C stages. For each player i in N, let $r_i(D_{P_k})$ be player i's realized average payoff during block P_k (this value can be computed by all players since there must have been no deviation in block P_k in order to reach the reward phase). Define now, for every player $i \in N$:

$$z_i(D_{P_k}) = r_i(D_{P_k}) \frac{1 - \delta^P}{\delta^P(1 - \delta^C)}.$$

For each subset of players $J \subseteq N$, denote c^J the payoff vector such that :

$$c_i^J = \begin{cases} \frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)} \left(v_i - \left(9 + \frac{n_{k-1}}{P}\right)\varepsilon\right) - z_i(D_{P_k}), & \text{if } i \in J \\ \frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)} v_i - z_i(D_{P_k}) & \text{otherwise.} \end{cases}$$

Since $\varepsilon > 0$ is such that for each $J \subseteq N$, the ball with center v^J and radius 10ε is included in the interior of $\mathcal{V}^*$, we have that c^J is also in the interior of $\mathcal{V}^*$. The vector c^J can be approximated arbitrarily closely by convex combinations of admissible payoffs with rational coefficients. So without loss of generality, we assume that for each $J \subseteq N$, $c^J = \sum_k \frac{T_k^J}{T} u(x_k^J)$, with $\sum_k T_k^J = T$, $x_k^J \in \mathcal{A}$, and T is the same as defined before (first paragraph of Appendix A).

The strategy of the players during phase COMP(n_{k-1}, D_{P_k}) is then to play cyclically the sequence of actions $x_k^{D_{P_k}}$ for $k \in \{1, \ldots, T\}$, which gives the vector payoff $c^{D_{P_k}}$ to the players until a new possible deviation. If there is a deviation at some stage, then players finish the compensation block COMP(n_{k-1}, D_{P_k}) and then start a new punishment phase. Otherwise, players go to the reward phase at the end of the compensation block. Let $C = RT$ with R an integer representing the number of cycles in the compensation phases.

Intuitively, this compensation block is such that suspected players in any punishing block have no incentive to deviate, otherwise they would only either lengthen their punishment (if this is the first deviation of the block) or increase n_{k-1} and therefore lower c^J by $\frac{\varepsilon}{P}$.

Reward phase. In order to provide each innocent player (who is not in the set of suspected players at some punishment block) with an incentive to play his Blackwell strategy during the punishment block, an additional bonus $\rho > 0$ is added to his average payoff. If the discount factor is large enough, the potential loss during the punishment is compensated by the future bonus. The possibility of defining such rewards relies on the fact that v is in the interior of $\mathcal{V}^*$ as in Fudenberg and Maskin (1986). The strategy of the players during this reward phase is then to play NORM$(v^{D_{P_k}})$ until a new possible deviation. If there is a deviation at some stage, then players start a new punishment phase.

The description of the strategy is now complete.[2] Next, we prove that it has the desired properties for appropriate choice of the parameters.

The equilibrium verification. We prove now that this strategy profile is a perfect public equilibrium for high discount factor and suitable choice of the parameters M, T, Q and R.

Remark first that the induced payoff is close to v when the discount factor is high. Indeed, since the main phase is cyclic, the discounted payoff under this profile is:

$$\sum\nolimits_{t=1}^{T} \frac{1-\delta}{1-\delta^T} \delta^{t-1} u(x_t),$$

which converges to $(1/T)\sum_{t=1}^{T} u(x_t) = v$ as $\delta \to 1$. More precisely, we let $M = \max_{x \in X}(\sum_i u_i(x)^2)^{\frac{1}{2}}$. For each $\alpha > 0$, we can choose δ high enough so that:

$$\sum\nolimits_{t=1}^{P} \left| \frac{1-\delta}{1-\delta^T} \delta^{t-1} - \frac{1}{T} \right| \leq \frac{\alpha}{M}.$$

[2] If at some stage, $D(x_t, s_t) = \emptyset$, that is, if signals indicate a multilateral deviation, the strategy prescribes a fixed Nash equilibrium of the one-shot game at all subsequent stages.

Thus, for any sequence $\{x_t\}$,

$$\left|\sum_{t=1}^{T} \frac{1-\delta}{1-\delta^T}\delta^{t-1}u(x_t) - \frac{1}{T}\sum_{t=1}^{T} u(x_t)\right| \leq \alpha.$$

Notice also that the same approximations hold for the punishment and compensations phases. Since P is a multiple of T, thus the discounted payoff over a block of length $P = QT$ can be written as,

$$\sum_{q=1}^{Q} \frac{1-\delta^T}{1-\delta^{TQ}}(\delta^T)^{q-1}\sum_{t=1}^{T} \frac{1-\delta}{1-\delta^T}\delta^{t-1}u(x_{t+(q-1)T}).$$

This is α close to the discounted sum of arithmetic averages,

$$\sum_{q=1}^{Q} \frac{1-\delta^T}{1-\delta^{TQ}}(\delta^T)^{q-1}\sum_{t=1}^{T} \frac{1}{T}u(x_{t+(q-1)T}).$$

The same holds for blocks of size C, since C is a multiple of T as well. We choose from now on an approximation error $\alpha << \min\{\varepsilon, \rho\}$, very small with respect to ε and ρ.

The next claim shows that each punishment block is effective and is based on approachability arguments.

Claim A.1. *Assume that a punishment block starts at stage $t^* + 1$. The average maximal payoff of each player j at stage $t^* + t \leq t^* + T$ satisfies,*

$$\bar{u}^*_{j,t} \leq (v_j - 10\varepsilon) + 2M/\sqrt{t}$$

with $M = \max_x \|u(x)\|$ and $\|.\|$ is the Euclidean norm on $\mathbb{R}^N$.

The proof is a straightforward adaptation of the one of Blackwell's (1956). Since it is quite simple, we provide it for the sake of completeness.

Proof. Let $\bar{u}^*_t$ be the vector $(\bar{u}^*_{j,t})_j \in \mathbb{R}^N$ and let $C = \{w \in \mathbb{R}^N : w_j \leq v_j - 10\varepsilon\}$. Endow $\mathbb{R}^N$ with the usual inner product and the corresponding Euclidean norm, and remark that the projection of $\bar{u}^*_t$ onto the convex and closed set C, is the vector $\pi(\bar{u}^*_t) = (\min\{\bar{u}^*_{j,t}, v_j - 10\varepsilon\})_j$. Thus, the vector of weights q is the normalization of the vector $\bar{u}^*_t - \pi(\bar{u}^*_t) = (\max\{\bar{u}^*_{j,t} - (v_j - 10\varepsilon), 0\})_j$. If $\bar{u}^*_t$ does not belong to C, then the hyperplane orthogonal to q which contains $\pi(\bar{u}^*_t)$, separates $\bar{u}^*_t$ from C. Now, $\underline{x}_{t+1}$ is such that for all $y \in NR(q, \underline{x}_{t+1})$, $\sum_j q_j(v_j - 10\varepsilon) \geq \sum_j q_j u_j(y_j, \underline{x}_{-j,t+1})$. In particular, since $x \in NR(q, x)$ for any x, the payoff vector $u(\underline{x}_{t+1})$ is separated from $\bar{u}^*_t$ by this hyperplane. Under a one-shot deviation, the next vector of maximal payoffs $u^*_{t+1} = u(\underline{x}_{t+1})$ is also separated from $\bar{u}^*_t$ by the hyperplane. Then,

$$\langle \bar{u}^*_t - \pi(\bar{u}^*_t), u^*_{t+1} - \pi(\bar{u}^*_t)\rangle \leq 0.$$

Define now D_t as the distance between $\bar{u}_t^*$ and C. We have, $D_{t+1}^2 \leq \left\| \bar{u}_{t+1}^* - \pi(\bar{u}_t^*) \right\|^2$, and since $\bar{u}_{t+1}^* = \frac{t}{t+1}\bar{u}_t^* + \frac{1}{t+1}u_{t+1}^*$, and $D_t^2 = \|\bar{u}_t^* - \pi(\bar{u}_t^*)\|^2$, we have,

$$D_{t+1}^2 \leq \left(\frac{t}{t+1}\right)^2 D_t^2 + \left(\frac{2M}{t+1}\right)^2 + 2\frac{t}{(t+1)^2}\left\langle \bar{u}_t^* - \pi(\bar{u}_t^*), u_{t+1}^* - \pi(\bar{u}_t^*) \right\rangle .$$

It follows that for each t, $D_{t+1}^2 \leq \left(\frac{t}{t+1}\right)^2 D_t^2 + \left(\frac{2M}{t+1}\right)^2$, the inequality being obvious when $D_{t+1} = 0$. A simple induction argument then gives, $D_t^2 \leq (2M)^2/t$, which implies that,

$$\max\{\bar{u}_{j,t}^* - (v_j - 10\varepsilon), 0\} \leq \left(\sum\nolimits_j \left(\max\{\bar{u}_{j,t}^* - (v_j - 10\varepsilon), 0\}\right)^2\right)^{1/2} \leq 2M/\sqrt{t},$$

as desired. □

A direct consequence is that we can choose T large enough so that the average payoff $\bar{u}_{j,T}^*$ is no more than $v_j - 9\varepsilon$. A one-shot deviation by player j thus entails an average payoff less than or equal to $v_j - 9\varepsilon$ over the punishment phase.

Thanks to the one-shot deviation principle (see e.g. Mailath and Samuelson (2006), Proposition 7.1.1., page 231), it is enough to check that no player has an incentive to deviate at a single stage and conform with the strategy afterwards. Since the strategy prescribes only admissible action profiles, a deviation that does not affect the public signal is not profitable. We may thus focus on detectable deviations.

Deviation from the main phase. Consider a one-shot deviation from some player j in the normal phase NORM(v_j) at some stage t^* and assume that player j deviates to y_j such that $f(y_j, x_{-j,t^*}(\sigma)) \neq f(x_{t^*}(\sigma))$. Recall that we consider one-shot deviations, so player j does not expect any further deviation, hence the punishment phase is composed of only one punishment block of length $P = QT$, and of the compensation block. Player j's discounted payoff normalized at the stage of the deviation is thus at most :

$$\begin{aligned} A &:= (1-\delta)M + \delta(1-\delta^P)r_j(D_{P_k}) \\ &\quad +\delta^{P+1}(1-\delta^C)\left[\frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)}(v_j - 9\varepsilon) - z_j(D_{P_k})\right] + \delta^{P+C+1}v_j \\ &= (1-\delta)M + \delta(1-\delta^{P+C})(v_j - 9\varepsilon) + \delta^{P+C+1}v_j. \end{aligned}$$

If player j conforms with the strategy, his payoff is at least $B := -(1-\delta^T)M + \delta^T v_j$, where the first term is there because the deviation may occur in the middle of a cycle of the normal phase. Now,

$$\begin{aligned}A-B &= (1-\delta)M+\delta(1-\delta^{P+C})(v_j-9\varepsilon)+(\delta^{P+C+1}-\delta^T)v_j+(1-\delta^T)M\\ &= (1-\delta)\left[M+\frac{\delta(1-\delta^{P+C})}{1-\delta}(v_j-9\varepsilon)+\frac{\delta^{P+C+1}-\delta^T}{1-\delta}v_j+\frac{(1-\delta^T)}{1-\delta}M\right]\\ &= (1-\delta)\left[M+\frac{\delta(1-\delta^{P+C})}{1-\delta}(-9\varepsilon)+\frac{\delta(1-\delta^{T-1})}{1-\delta}v_j+\frac{(1-\delta^T)}{1-\delta}M\right]\end{aligned}$$

The limit of $(A-B)/(1-\delta)$ as $\delta \to 1$ is then:

$$D := M+(P+C)(-9\varepsilon)+(T-1)v_j+TM.$$

Dividing by $P+C$, we get:

$$D/(P+C) = \frac{(T+1)M+(T-1)v_j}{P+C} - 9\varepsilon.$$

We may now tune the length of blocks T and of the punishment phase $P+C$, in such a way that this expression is negative. It is enough to choose $P+C$ large with respect to T and M, to ensure $D/(P+C) \leq -8\varepsilon$. For this choice of the parameters, δ can be chosen large enough so that $A-B < -\varepsilon$. The discounted payoff can then differ by at most α from the discounted sum of average payoffs over blocks. With $\alpha << \varepsilon$, the deviation is not profitable for high enough δ.

Deviation from a punishment block. Assume that a punishment phase is going with current punishment block P_k, $k > 0$ (recall that P_k consists of Q blocks of stages of length T). We consider two types of players: first, the players in D_{P_k} that will not be rewarded (*i.e.* the set of suspected players during block P_k), second the players in $J := N \setminus D_{P_k}$. The main difference is that for players $j \in D_{P_k}$, the maximal payoffs u_j^* considered in the punishment phase, upper bound their actual payoff. The approachability strategy ensures that all players that *could* be responsible for the deviation are actually punished. Since the signal do not discriminate between them, they all have to be treated the same way and in particular, not to be offered rewards (otherwise, that would created incentives to trigger the punishment). By contrast, for players j in J, the maximal payoffs u_j^* need not coincide with their actual payoff. Also, in each case, we need to focus on two subcases: a deviation might already have occurred in the current block, or not.

1. **First case, consider $j \in D_{P_k}$.**

   (a) *Suppose first that there has been no deviation in block P_k so far.*

   If player j deviates at block $q \in \{1, \ldots, Q\}$ in block P_k, then he may profit from the deviation for the remaining stages of block q. For the $Q-q$ remaining blocks of block P_k, he gets a payoff, denoted by W thereafter, which is the same as if he had not deviated. After block P_k, he has to go through an additional punishment block P_{k+1}, before starting the compensation block then the reward phase. His discounted payoff normalized at the beginning of block q (given that player j deviates) is thus at most:[3]

$$\begin{aligned} A &:= (1-\delta^T)M + \delta^T(1-\delta^{(Q-q)T})W \\ &\quad +\delta^{T+(Q-q)T}(1-\delta^{P+C})\left[v_j - \left(9+\frac{1}{P}\right)\varepsilon\right] \\ &\quad +\delta^{T+(Q-q)T+P+C}v_j. \end{aligned}$$

If player j conforms, his discounted payoff is at least:

$$\begin{aligned} B &:= -(1-\delta^T)M + \delta^T(1-\delta^{(Q-q)T})W \\ &\quad +\delta^{T+(Q-q)T}(1-\delta^{C})\left[\frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)}(v_j-9\varepsilon) - \frac{1-\delta^P}{\delta^P(1-\delta^C)}r_j(D_{P_k})\right] \\ &\quad +\delta^{T+(Q-q)T+C}v_j. \end{aligned}$$

Now,

$$\begin{aligned} A-B &= (1-\delta^T)2M - \delta^{T+(Q-q)T}\frac{1-\delta^P}{\delta^P}v_j + \delta^{T+(Q-q)T}(1-\delta^{P+C})(-9\varepsilon-\frac{\varepsilon}{P}) \\ &\quad -\delta^{T+(Q-q)T}(1-\delta^{C})\left[\frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)}(-9\varepsilon) - \frac{1-\delta^P}{\delta^P(1-\delta^C)}r_j(D_{P_k})\right]. \end{aligned}$$

The limit of $(A-B)/(1-\delta)$ as $\delta \to 1$ is then:

$$\begin{aligned} &= 2TM - Pv_j + (P+C)(-9\varepsilon-\frac{\varepsilon}{P}) - (P+C)(-9\varepsilon) \\ &\quad +Pr_j(D_{P_k}) \\ &= 2TM - Pv_j - (P+C)\frac{\varepsilon}{P} + Pr_j(D_{P_k}). \end{aligned}$$

Dividing by P, we get:

$$D/(P+C) = \frac{2TM}{P} - v_j - \frac{P+C}{P}\frac{\varepsilon}{P} + r_j(D_{P_k}).$$

By Claim A.1, $r_j(D_{P_k}) \leq v_j - 9\varepsilon$ so that $r_j(D_{P_k}) - v_j \leq -9\varepsilon$. Hence,

$$D/(P+C) \leq \frac{2TM}{P+C} - \frac{\varepsilon}{P} - \frac{P}{P+C}9\varepsilon.$$

[3] Given that player j deviates at block q, he should do so at the beginning of block q in order to maximize the potential gain of his deviation, since he may then profit from it for the remaining T stages of block q.

Similarly as in the previous case, choosing $P+C$ large with respect to T and M, ensures that the right-hand-side is negative, and therefore the deviation is not profitable for δ large enough.

(b) *Suppose now that there have been m deviations in block P_k so far, with $0 < m < P$, the last one at stage t_m.*

If player j deviates at block $q \in \{1, \dots, Q\}$ in block P_k, then he may profit from the deviation for the remaining stages of block q. As in the previous case, for the $Q-q$ remaining blocks of block P_k, he gets a payoff W which is the same as if he had not deviated. After block P_k, he has to go through an additional punishment block P_{k+1}, before starting the compensation block then the reward phase. His discounted payoff normalized at the beginning of block q is thus at most:

$$\begin{aligned} A &:= (1-\delta^T)M + \delta^T(1-\delta^{(Q-q)T})W \\ &\quad +\delta^{T+(Q-q)T}(1-\delta^{P+C})\left[v_j - \left(9+\frac{m}{P}\right)\varepsilon\right] \\ &\quad +\delta^{T+(Q-q)T+P+C}v_j. \end{aligned}$$

Regarding player j's payoff if he conforms, two cases are possible. First, $j \in D(\underline{x_{t_m}}, s_{t_m})$, hence player j is a potential deviator of the last deviation at stage t_m. Player j's payoff if he conforms is then at least:

$$\begin{aligned} B &:= -(1-\delta^T)M + \delta^T(1-\delta^{(Q-q)T})W \\ &\quad +\delta^T\delta^{(Q-q)T}(1-\delta^{P+C})\left[v_j - \left(9+\frac{m-1}{P}\right)\varepsilon\right] \\ &\quad +\delta^{T+(Q-q)T+P+C}v_j. \end{aligned}$$

Now,

$$A-B = (1-\delta^T)2M - \delta^{T+(Q-q)T}(1-\delta^{P+C})\left(-\frac{\varepsilon}{P}\right).$$

The limit of $(A-B)/(1-\delta)$ as $\delta \to 1$ is then:

$$D := 2TM - (P+C)\left(-\frac{\varepsilon}{P}\right).$$

Dividing by $P+C$, we get:

$$D/(P+C) = \frac{2TM}{P+C} - \frac{\varepsilon}{P}.$$

In that case, choosing C large with respect to P, T and M, ensures that the deviation is not profitable for δ large enough.

Second, suppose that player j is not in $D(\underline{x_{t_m}}, s_{t_m})$, that is player j is not suspected as a potentiel deviator at stage t_m. Player j's payoff if he conforms is then at least:

$$\begin{aligned} B' &:= -(1-\delta^T)M \\ &\quad +\delta^T\left[(1-\delta^{(Q-q)T})W + \delta^{(Q-q)T}(1-\delta^{P+C})v_j\right] \\ &\quad +\delta^{T+(Q-q)T+P+C}(v_j+\rho). \end{aligned}$$

Now,

$$\begin{aligned} A - B' &= (1-\delta^T)2M + \delta^{T+(Q-q)T}(1-\delta^{P+C})\left(-9\varepsilon - \frac{m}{P}\varepsilon\right) \\ &\quad -\delta^{T+(Q-q)T+P+C}\rho. \end{aligned}$$

The limit of $A - B'$ as $\delta \to 1$ is then $-\rho$, thus for δ large enough, the deviation is not profitable.

2. **Second case, player $j \in J = N \setminus D_{P_k}$.**

   (a) *Suppose first that there has been no deviation in block P_k so far.*

   If player j deviates at block $q \in \{1, \dots, Q\}$ in block P_k, then he may profit from the deviation for the remaining stages of block q. For the $Q-q$ remaining blocks of block P_k, he gets a payoff W (the same as if he had not deviated). After block P_k, he has to go through an additional punishment block P_{k+1}, before starting the compensation block then the reward phase. His discounted payoff normalized at the beginning of block q (given that player j deviates) is thus at most:

$$\begin{aligned} A &:= (1-\delta^T)M \\ &\quad +\delta^T\left[(1-\delta^{(Q-q)T})W + \delta^{(Q-q)T}(1-\delta^{P+C})\left(v_j - \left(9+\frac{1}{P}\right)\varepsilon\right)\right] \\ &\quad +\delta^{T+(Q-q)T+P+C}v_j. \end{aligned}$$

   If player j conforms, his discounted payoff is at least:

$$\begin{aligned} B &:= -(1-\delta^T)M + \delta^T(1-\delta^{(Q-q)T})W \\ &\quad +\delta^{T+(Q-q)T}(1-\delta^C)\left[\frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)}v_j - \frac{1-\delta^P}{\delta^P(1-\delta^C)}r_j(D_{P_k})\right] \\ &\quad +\delta^{T+(Q-q)T+C}(v_j+\rho). \end{aligned}$$

Now,

$$\begin{aligned}A-B &= (1-\delta^T)2M + \delta^{T+(Q-q)T}(1-\delta^{P+C})\left(v_j - \left(9+\frac{1}{P}\right)\varepsilon\right)\\ &\quad -\delta^{T+(Q-q)T}(1-\delta^C)\left[\frac{1-\delta^{P+C}}{\delta^P(1-\delta^C)}v_j - \frac{1-\delta^P}{\delta^P(1-\delta^C)}r_j(D_{P_k})\right]\\ &\quad -\delta^{T+(Q-q)T+C}(1-\delta^P)v_j - \delta^{T+(Q-q)T+C}\rho.\end{aligned}$$

As in the previous case, the limit of $A-B'$ as $\delta \to 1$ is $-\rho$ and for δ large enough, the deviation is not profitable.

(b) *Suppose now that there have been m deviations in block P_k so far, with $0 < m < P$, the last one at stage t_m.*

This is similar to case 1.(b) above. Indeed, whether player j is punished at block P_k or not, his potential gain from deviating only depends on the fact that he was suspected (or not) at the last deviation (which happened at stage t_m).

Deviation from the compensation phase. If player j deviates at block $r \in \{1,\ldots,R\}$ of the compensation block COMP(n_{k-1}, D_{P_k}), then he may profit from the deviation for the remaining stages of block r. For the $R-r$ remaining blocks of the compensation block, he gets a payoff W (as before, the same as if he had not deviated). After the compensation block, he has to go through a new punishment block P_1, before starting a new compensation block then the reward phase. His discounted payoff normalized at the beginning of block r is thus at most:

$$\begin{aligned}A &:= (1-\delta^T)M\\ &\quad +\delta^T\left[(1-\delta^{(R-r)T})W + \delta^{(R-r)T}(1-\delta^{P+C})(v_j - 9\varepsilon)\right]\\ &\quad +\delta^{T+(R-r)T+P+C}v_j.\end{aligned}$$

Regarding player j's payoff if he conforms, two cases are possible. First, suppose that $j \in D_{P_k}$, his discounted payoff is then:

$$B := -(1-\delta^T)M + \delta^T(1-\delta^{(R-r)T})W + \delta^{T+(R-r)T}v_j.$$

Now,

$$A - B = (1-\delta^T)2M - \delta^{T+(R-r)T}(1-\delta^{P+C})(-9\varepsilon).$$

The limit of $(A-B)/(1-\delta)$ as $\delta \to 1$ is then:

$$D \;:=\; 2TM + (P+C)(-9\varepsilon).$$

Dividing by $P+C$, we get:

$$D/(P+C) \;=\; \frac{2TM}{P+C} - 9\varepsilon.$$

As in previous cases, choosing $P+C$ large with respect to T and M ensures that the deviation is not profitable for large δ.

Second, suppose that player $j \in J = N \setminus D_{P_k}$. His discounted payoff if he conforms is then:

$$B \;:=\; -(1-\delta^T)M + \delta^T(1-\delta^{(R-r)T})W + \delta^{T+(R-r)T}(v_j+\rho),$$

which is strictly more than the previous case, hence the deviation is not profitable either.

Deviation from the reward phase. Suppose that player j deviates during a reward phase. Two cases are possible. First, take player $j \in D_{P_k}$ with P_k the last punishment block. Then, player j's discounted payoff if he conforms during the reward phase is v_j. The situation is similar the main phase, and and deviation is not profitable for choices of parameters and discount factors as in previous cases. Second, if player $j \notin D_{P_k}$, then his discounted payoff is $v_j + \rho$, which is more than in the previous case, so for the same choice of parameters and a high discount factor, the deviation is not profitable either.

To conclude, this strategy is an equilibrium for the following choice of the parameters: Q and R large enough so that $P+C$ large with respect to both T and M, and R large enough so that C is large with respect to P. This ends the proof of Theorem 3.12. □