Truthful Equilibria in Dynamic Bayesian Games

Johannes Hörner*, Satoru Takahashi† and Nicolas Vieille‡

November 16, 2014

Abstract

This paper characterizes an equilibrium payoff subset for Markovian games with private information as discounting vanishes. Monitoring is imperfect, transitions may depend on actions, types be correlated and values interdependent. The focus is on equilibria in which players report truthfully. The characterization generalizes that for repeated games, reducing the analysis to static Bayesian games with transfers. With correlated types, results from mechanism design apply, yielding a folk theorem. With independent private values, the restriction to truthful equilibria is without loss, except for the punishment level; if players withhold their information during punishment-like phases, a “folk” theorem obtains also.

**Keywords:** Bayesian games, repeated games, folk theorem.

**JEL codes:** C72, C73

1 Introduction

This paper studies the asymptotic equilibrium payoff set of repeated Bayesian games. In doing so, it generalizes methods that were developed for repeated games (Fudenberg and Levine, 1994; hereafter, FL) and later extended to stochastic games (Hörner, Sugaya, Takahashi and Vieille, 2011, hereafter HSTV).

Serial correlation in the payoff-relevant private information (or *type*) of a player makes the analysis of such repeated games difficult. Therefore, asymptotic results in this literature

---

*Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, johannes.horner@yale.edu.
†National University of Singapore, ecsst@nus.edu.sg.
‡HEC Paris, 78351 Jouy-en-Josas, France, vieille@hec.fr.
have been obtained by means of increasingly elaborate constructions, starting with Athey and Bagwell (2008) and culminating with Escobar and Toikka (2013). These constructions are difficult to extend beyond a certain point, however. Instead, our methods apply to games exhibiting:

- moral hazard (imperfect monitoring);
- endogenous serial correlation (actions affecting transitions);
- correlated types (across players) and interdependent values.

Allowing for such features is not merely of theoretical interest. There are many applications in which some if not all of them are relevant. In insurance markets, for instance, there is clearly persistent adverse selection (risk types), moral hazard (accidents and claims having a stochastic component), interdependent values, action-dependent transitions (risk-reducing behaviors) and, in the case of systemic risk, correlated types. The same holds true in financial asset management, and in many other applications of such models (taste or endowment shocks, etc.)

We assume that the state profile—each coordinate of which is private information to a player—follows a controlled autonomous irreducible Markov chain. (Irreducibility refers to its behavior under any fixed Markov strategy.) In the stage game, players privately take actions, and then a public signal realizes, whose distribution may depend both on the state and action profile, and the next round state profile is drawn. Cheap-talk communication is allowed, in the form of a public report at the beginning of each round.

Our analysis is about truthful equilibria. In a truthful equilibrium, players truthfully reveal their type at the beginning of each round, after every history. In addition, players’ action choices are public: they only depend on their current type and the public history. Our main result characterizes a subset of the limit set of equilibrium payoffs as the discount factor $\delta$ tends to one. While concentrating on truth-telling equilibria is with loss of generality given the absence of any commitment, it nevertheless turns out that this limit set includes the payoff sets obtained in all the special cases studied by the literature.\(^2\)

In Sections 2–5, we focus on the case of independent private values: payoffs only depend on a player’s own private information (and the action profile), and this information evolves

---

1 This is not to say that the recursive formulation of Abreu, Pearce and Stacchetti (1990, hereafter APS) cannot be adapted to such games, but these methods provide little guidance regarding qualitative properties of the equilibrium payoff set.

2 The one exception is the lowest equilibrium payoff in Renault, Solan and Vieille (2013), who also characterize Pareto-inferior “babbling” equilibria.
independently across players, conditional on the public information and one's own private action. We provide a family of one-shot games with transfers that reduce the analysis from a dynamic infinite-horizon game to a static game. Unlike the one-shot game of FL and HSTV (special cases of ours), this one-shot game is Bayesian. Each player makes a report, then takes an action; the transfer is then determined. This reduction provides a bridge between dynamic games and Bayesian mechanism design. As explained below, its payoff function is not entirely standard, raising interesting new issues for static mechanism design. Nonetheless, well-known results can be adapted for a wide class of dynamic games. This is our second contribution: under either independent private values, or correlated types, the analysis of the one-shot game yields an equilibrium payoff set that is best possible, except for the definition of individual rationality.

For such games, we prove a folk theorem: truthful equilibria might be restrictive in terms of individual rationality (lowest equilibrium payoff for a given player), but they do not restrict the set of equilibrium payoffs otherwise. Leaving aside individual rationality, we show that the payoff set attained by truthful equilibria is actually equal to the limit set of all Bayes-Nash equilibrium payoffs, whichever message sets one chooses. In other words, in the revelation game in which players commit to the map from reports to actions, but not to current or future reports, there is no loss of generality in restricting attention to truthful equilibria. In this sense, the revelation principle extends when players are patient enough. Beyond generalizing the results of Athey and Bagwell, as well as Escobar and Toikka, this characterization has some interesting consequences. For instance, when actions do not affect transitions, the invariant distribution of the Markov chain is a sufficient statistic for the Markov process, as far as this equilibrium payoff set is concerned, leaving individual rationality aside.

In Section 5, we further concentrate on games in which monitoring has a product structure. This is the class of games for which, absent any private information, existing “folk” theorems are actual characterizations of the set of (limit) sequential equilibrium payoffs. While insisting on truthfulness might be restrictive in terms of individual rationality (as mentioned above) we show that, for the case of product structure, a simple twist on such equilibria (in which players abstain from reporting their private information when punishing others) provides an exact characterization of all Bayesian Nash equilibrium payoffs.

In Section 6, we state a general version of our main theorem, which provides a subset of limit equilibrium payoffs, whether types are correlated and values are private, or not. When types are correlated, then all feasible and individually rational payoffs can be obtained in the limit (again, under suitable identifiability conditions). The “spanning” condition familiar
from mechanism design with correlated types must be stated in terms of *pairs* of states: more precisely, player \(-i\)'s current and *next* state must be sufficiently informative about player *i*’s current and *previous* state. Conclusive characterizations are obtained under independent private values as well as correlated types. This mirrors the state of affairs in static mechanism design. In fact, our results are obtained by applying familiar techniques to the one-shot game, developed by Arrow (1979) and d’Aspremont and Gérard-Varet (1979) for the independent case, and d’Aspremont, Crémer and Gérard-Varet (2003) in the correlated case.

Our approach stands in contrast with the techniques based on review strategies (see Escobar and Toikka for instance) whose adaptation to incomplete information is inspired by the linking mechanism described in Fang and Norman (2006) and Jackson and Sonnenschein (2007). Our results imply that, as is already the case for repeated games with public monitoring, transferring continuation payoffs across players is an instrument that is sufficiently powerful to dispense with explicit statistical tests. Of course, this instrument requires that deviations in the players’ reports can be statistically distinguished, a property that calls for assumptions closely related to those called for in static mechanism design. Here as well, we build on results from static mechanism design (in particular the weak identifiability condition introduced by Kosenok and Severinov (2008)) to ensure budget-balance in the dynamic game.

While the characterization turns out to be a natural generalization of the one from repeated games with public monitoring, it still has several unexpected features, reflecting difficulties in the proof that are not present either in stochastic games with observable states. These difficulties shift the emphasis of the program from payoffs to strategies.

To bring these difficulties to light, consider the case of independent types. Together with the irreducibility of the Markov chain, this implies that the long-run (or asymptotic) payoff of a player is independent of his current state. To incentivize a player to disclose his private information, it no longer suffices to adjust his long-run payoff, as it affects the different types identically. Using solely the current (flow) payoff to elicit truth-telling is just as inadequate, when actions affect transitions. Player *i*’s incentives to disclose his information depends on the impact of his report on the *transient* component of his long-run payoff; that is, loosely speaking, on his flow payoffs until the effect of the initial state fades away. This transient component is bounded from above, even as \(\delta \to 1\): unlike in repeated games, future payoffs do not eclipse flow payoffs, as far as incentives to tell the truth are concerned. Furthermore, this transient component cannot be summarized by a single number: its value depends on the player’s initial state, according to the future actions played.

To resolve these difficulties, the proof adopts two time scales. Over the short run,
policy that players follow (the map from reports to actions) is fixed. The resulting transient component follows directly, and is treated as a flow payoff. In other words, in the short run, the flow payoff is computed as if strategies were Markov: the relative value that arises in (undiscounted) dynamic programming is precisely the right measure for this transient component. In the long run, play is decidedly non-Markovian. Play switches towards a new Markov strategy profile that metes out punishments and rewards according to the history of public signals.

The two time scales interact, however, leading to a characterization that intermingles both the relative value (treated as an adjustment to the flow payoff) and the changes in the long-run payoff (treated, as usual, as a transfer).

Games without commitment but with imperfectly persistent private types were introduced in Athey and Bagwell (2008) in the context of Bertrand oligopoly with privately observed cost. Athey and Segal (2013, hereafter AS) allow for transfers and prove an efficiency result for ergodic Markov games with independent types. Their team balanced mechanism is closely related to a normalization that is applied to the transfers in one of our proofs in the case of independent private values.

There is also a literature on undiscounted zero-sum games with such a Markovian structure, see Renault (2006), which builds on ideas introduced in Aumann and Maschler (1995). Not surprisingly, the average cost optimality equation plays an important role in this literature as well. Because of the importance of such games for applications in industrial organization and macroeconomics, there is an extensive literature on recursive formulations for fixed discount factors. In game theory, recent progress has been made in the case in which the state is observed, see Fudenberg and Yamamoto (2012) and HSTV for an asymptotic analysis, and Pęski and Wiseman (2013) for the case in which the time lag between consecutive moves goes to zero. There are some similarities in the techniques used, although incomplete information introduces significant complications.\footnote{Among others, HSTV (as before FLM) rely on the equilibrium payoff set being full-dimensional, an assumption that fails with independent private values: When the players' types follow independent Markov chains and values are private, the players' limit equilibrium payoff must be independent of their initial type, given irreducibility and incentive-compatibility.}

More related are the papers by Escobar and Toikka, already mentioned, Barron (2013) and Renault, Solan and Vieille. All three papers assume that types are independent across players. Barron introduces imperfect monitoring in Escobar and Toikka, but restricts attention to the case of one informed player only. This is also the case in Renault, Solan and Vieille. This is the only paper that allows for interdependent values, although in the context
of a very particular model, namely, a sender-receiver game with perfect monitoring. None of these papers allow transitions to depend on actions.

2 Model and Equilibrium

We consider dynamic games with imperfectly persistent incomplete information.

2.1 Extensive Form

The stage game is as follows. The finite set of players is denoted \( I \). We assume that there are at least two players. Each player \( i \in I \) has a finite set \( S^i \) of (private) states, or types, and a finite set \( A^i \) of actions. The state \( s^i \in S^i \) is private information to player \( i \). We denote by \( S := \times_{i \in I} S^i \) and \( A := \times_{i \in I} A^i \) the sets of state profiles and action profiles respectively.

In each round \( n \geq 1 \), timing is as follows:

1. Each player \( i \in I \) privately observes his own state \( s^i_n \in S^i \);
2. Players simultaneously make reports \( (m^n_i)_{i=1}^I \in \times_i M^i \), where \( M^i \) is a finite set. These reports are publicly observed;
3. The outcome of a public randomization device (prd) is observed. For concreteness, it is a draw from the uniform distribution on \([0, 1] \);
4. Players independently choose actions \( a^i_n \in A^i \). Actions taken are not observed;
5. A public signal \( y_n \in Y \), a finite set, and the next state profile \( s_{n+1} = (s^i_{n+1})_{i \in I} \) are drawn according to some joint distribution \( p(\cdot, \cdot | s_n, a_n) \in \Delta(S \times Y) \).

The stage-game payoff or reward of player \( i \) is a function \( r^i : S \times A \rightarrow \mathbb{R} \), whose domain is extended to mixed action profiles in \( \Delta(A) \). As is customary, we may interpret this reward as the expected value (with respect to the signal \( y \)) of some function \( g^i : S \times A^i \times Y \rightarrow \mathbb{R} \), \( r^i(s, a) = \mathbb{E}[g^i(s, a^i, y) | a] \). In that case, given \( (s, a^i, y) \), the realized reward does not convey additional information about \( a^{-i} \), so that whether this reward is observed or not is irrelevant (for the updating of beliefs over \( a^{-i} \), conditional on \( (s, a^i, y) \)). We do not make this assumption, but assume instead that realized rewards are not observed. Hence, we

\footnote{We do not know how to dispense with it. But given that public communication is allowed, such a public randomization device is innocuous, as it can be replaced by jointly controlled lotteries.}
assume that a player’s private action, private state, the public signal and report profile is all the information available to him.

Given the sequence of realized rewards \((r^i_n) = (r^i(s_n, a_n))\), player \(i\)’s payoff in the dynamic game is given by

\[
\sum_{n=1}^{+\infty} (1 - \delta)\delta^{n-1}r^i_n,
\]

where \(\delta \in [0, 1)\) is common to all players. (Short-run players can be accommodated for, as will be discussed.)

The dynamic game also specifies an initial distribution \(\pi_1 \in \Delta(S)\), which plays no role in the analysis, given the irreducibility assumption we will impose and the focus on equilibrium payoff vectors as elements of \(\mathbf{R}^I\) as \(\delta \to 1\).

Our focus will be on independent private values (hereafter, IPV). This is defined as the special case in which (i) transitions satisfy

\[
p(t, y \mid s, a) = p(y \mid a) \times \times_{i \in I} p^i(t^i \mid s^i, y),
\]

as well as

\[
\pi_1(s) = \times_{i \in I} \pi^i_1(s^i),
\]

for some transitions \(\{p^i(\cdot \mid s^i, y)\}_{s^i, y} \subseteq \Delta(S^i)\), and distributions \(\{p(\cdot \mid a)\}_a \subseteq \Delta(Y)\), \(\pi^i_1 \in \Delta(S^i)\), all \(i \in I\), and (ii) rewards satisfy, for all \(i \in I, s \in S, a \in A\), \(r^i(s, a) = r^i(s^i, a)\). The first assumption guarantees that beliefs over state profiles are common knowledge throughout the game, on and off path. We assume full support: \(\pi^i_1(s^i) > 0, p^i(t^i \mid s^i, y) > 0\) for all \(t^i, s^i\) and \(y\).

In Section 6, we extend our analysis to types that are not independent, and/or values that are not private. In the case of interdependent values, it matters whether players observe their payoffs or not. It is possible to accommodate privately observed payoffs: simply define a player’s private state as including his last realized payoff.\(^5\) As we shall see, the reports of a player’s opponents in the next round are taken into account when evaluating the truthfulness of a player’s report, so that one could build on the results of Mezzetti (2004) in static mechanism design with interdependent valuations. In fact, our main characterization result extends immediately to the case in which monitoring is private, rather than public; see Section 6 for a discussion.

\(^5\)With this interpretation, pointed out by AS, interdependent values with observable payoffs reduce to private values \(ex\ post\), as conditional on a player’s entire information, a player’s payoff does not depend on the other players’ types. It would then be natural to allow for a second round of messages at the end of each period.
Monetary transfers are not allowed. We view the stage game as capturing all possible interactions among players, and there is no difficulty in interpreting some actions as monetary transfers. In this sense, rather than ruling out monetary transfers, what is assumed here is limited liability.

The game defined above allows for public communication among players. In doing so, we follow most of the literature on Markovian games with private information, see Athey and Bagwell (2001, 2008), Escobar and Toikka, Renault, Solan and Vieille, etc. As in static Bayesian mechanism design, communication is required for coordination even in the absence of strategic motives; communication allows us to characterize what restrictions on payoffs, if any, are imposed by non-cooperative behavior.

2.2 Truthful Equilibria

2.2.1 Definition

We now define the class of Bayes Nash equilibria studied in this paper. This class coincides with perfect public equilibria (PPE) in repeated games with imperfect public monitoring. It follows that it is with loss of generality. As for PPE, the definition is motivated by tractability, with the hope that the resulting payoff characterization proves to be without loss under fairly weak conditions on the game.

The set of messages available to the players is an ingredient of the solution concept. Here and until Section 6, we assume that

\[ M^i = S^i. \]

This is restrictive. Because players cannot commit, the revelation principle does not apply (see Bester and Strausz, 2000), and richer message sets might lead to larger sets of equilibrium payoffs. Let \( M := \times_{i \in I} M^i \).

Furthermore, we focus on equilibria in which players truthfully reveal their private state in every period, on and off path. A priori, there is no reason to expect such equilibria to even exist.

Formal definitions require additional notation. A public history at the start of round \( n \geq 1 \) is a sequence \( h_{\text{pub},n} = (m_1, y_1, \ldots, m_{n-1}, y_{n-1}) \in H_{\text{pub},n} := (M \times Y)^{n-1} \). Player \( i \)'s private

\[^6\]This is not to say that introducing a mediator would be uninteresting. Following Myerson (1986), we could then appeal to a revelation principle, although without commitment from the players this would simply shift the inferential problem to the recommendation step of the mediator.

\[^7\]For clarity, we maintain the notational distinction.
history at the start of round $n$ is a sequence $h^i_n = (s^i_1, m_1, a^i_1, y_1, \ldots, s^i_{n-1}, m_{n-1}, a^i_{n-1}, y_{n-1}) \in H^i_n := (S^i \times M \times A^i \times Y)^{n-1}$. (Here, $H^i_1 = H_{\text{pub},1} := \{\emptyset\}$. ) A (behavior) strategy for player $i$ is a pair of sequences $(m^i, a^i) = (m^i_n, a^i_n)_{n \in \mathbb{N}}$ with $m^i_n : H^i_n \times S^i \rightarrow \Delta(M^i)$, and $a^i_n : H^i_n \times S^i \times M \rightarrow \Delta(A^i)$, which specify $i$’s report and action as a function of his private information, his current state and the report profile in the current round.\footnote{Recall however that a prd is assumed, although it is omitted from the notations.} A strategy profile $(m, a)$ defines a distribution over finite and infinite histories in the usual way.

**Definition 1** A strategy $(m^i, a^i)$ is truthful if $m^i_n(h^i_n, s^i_n) = s^i_n$ for all histories $h^i_n$, $n \geq 1$, and $a^i(h^i_n, s^i_n, m_n)$ depends on $(h_{\text{pub},n}, s^i_n, m_n)$ only.

Our analysis makes extensive use of the notion of a policy (or Markov strategy). This is simply a map $\rho : S \times \Delta(A)$, interpreted as a (possibly correlated) choice of action given the vector of states (or reports).

### 2.2.2 Limitations

To appreciate why truthful equilibria are restrictive, consider a two-player game with perfect monitoring in which player 1 has two possible states $\underline{s}, \bar{s}$, while player 2 has only one. Furthermore, suppose that player 1’s states are i.i.d. over time.\footnote{For the case of i.i.d. states, FL’s algorithm can be adapted, see Section 8 of FLM.} Suppose that one wishes to incentivize player 1 to randomize over two of his actions (say, $\underline{a}^1$ and $\bar{a}^1$) in both states, after some particular public history. Because states are i.i.d., player 1’s expected continuation payoff (from the following period onward) only depends on his report and the action profile played. Let $w^1(m, a^1)$ denote the expectation of this payoff over the action $a^2$. Similarly, with abuse of notation, let $r^1(s, m, a^1)$, denote player 1’s expected reward given state $s$, report $m$ (which might affect $a^2$) and $a^1$.

Because player 1 may always set $m = \underline{s}$ or $m = \bar{s}$ and play $\underline{a}^1$ (an action that is on path for both messages) whether his state is $\underline{s}$ or $\bar{s}$, truth-telling requires that his expected continuation payoff be independent of his report. Hence, it must be $w^1(s, \underline{a}^1) = w^1(s, \bar{a}^1)$. Similarly, $w^1(\underline{s}, \bar{a}^1) = w^1(\bar{s}, \underline{a}^1)$. Hence, $w^1(s, \underline{a}^1) - w^1(s, \bar{a}^1)$ is independent of $s$. At the same time, indifference over the two actions, given a report $s = \underline{s}, \bar{s}$, requires that

$$r^1(s, s, \bar{a}^1) - r^1(s, s, \underline{a}^1) = \frac{\delta}{1 - \delta} (w^1(s, \underline{a}^1) - w^1(s, \bar{a}^1)).$$

Because the right-hand side is independent of $s$, so must be the left-hand side. Yet there might be no action $\alpha^2 \in \Delta(A^2)$ for which $r^1(s, \bar{a}^1, \alpha^2) - r^1(s, \underline{a}^1, \alpha^2)$ is independent of $s$ –let alone for a specific choice of $\alpha^2$ that one might wish to implement.
It is clear that the argument is more general. Even with i.i.d. states, it is not usually possible to have a player be indifferent over several actions in more than one particular state in a truthful equilibrium.\footnote{In repeated games, players have a unique state, so this problem does not arise.}

Hence, asking for truth-telling rules out randomization (in all but at most one state). Yet randomization is helpful in achieving extremal payoffs in repeated games, for at least two reasons. First, it might be called upon by minmaxing. At the very least, truthful equilibrium curbs the ability to punish players. Second, it might help detection of deviations, when monitoring is imperfect, and the monitoring technology does not have the product structure: it might well be that, for each pure action of player 2, there are two actions of player 1 that are indistinguishable (in terms of public signals), yet none would be statistically indistinguishable if only player 2 were to randomize.

Whether or not such randomization is easy to achieve when players do not reveal their type is irrelevant: What matters for minmaxing or statistical detection of deviations is that a player’s action be unpredictable, whether this is because he deliberately randomizes over actions, or because his type determining his pure action cannot be inferred from his report. Hence, mixed minmaxing is consistent with a player playing a pure action given his type for all of them but one, as long as he does not disclose his type.

Given these observations, the next two sections restrict attention to minmaxing strategies in pure strategies, and to monitoring structures for which randomization does not affect the scope for statistical detection. In Section 5, we weaken the solution concept to allow for mixed minmaxing.

2.3 The Revelation Game

The game described in Section 2.1 involves both a choice of report and action. To clarify the role of the assumptions that we will introduce, it is useful to consider an auxiliary game in which players make reports, but do not control actions. That is, we are given a map \( \rho = \{\rho_n\}_{n \in \mathbb{N}} \), \( \rho_n : (M \times Y)^n \rightarrow \Delta(A) \), and amend the timing above by replacing step 4 with:

4’ Given the public history \((m_1, y_1, \ldots, m_n, y_n)\), the action profile is drawn according to 
\[ \rho_n(m_1, y_1, \ldots, m_n, y_n). \]

The other steps are unchanged. Payoffs are defined as before. The definition of strategies and of equilibrium is as before, with the obvious restriction to reports. An equilibrium is \textit{truthful} if \( m^i_n(h_n^i, s_n^i) = s_n^i \) for all \( i \in I \), \( h_n^i \in H_n^i \), \( n \geq 1 \) and states \( s_n^i \in S^i \).
We will be interested in the set of equilibrium payoffs of the revelation game that can be achieved for some $\rho$. Because players only affect actions via messages, the revelation game dispenses with obedience—in particular, individual rationality. Hence, the set of truthful equilibrium payoffs of the original game is a subset of the set of truthful equilibrium payoffs of the revelation game. *A priori*, this is not obvious for the set of all equilibrium payoffs, because in non-truthful equilibria, actions may depend on states, and not just on reports. Nonetheless, our results below imply that this is case.

3 Perfect Monitoring, Action-Independent Transitions

This section introduces some of the main ideas within the context of perfect monitoring and action-independent transitions. This is the case considered by Athey and Bagwell (2008) and Escobar and Toikka (2013). Proofs for this section are in Appendix A.

We denote by $\mu \in \Delta(S \times S)$ the invariant distribution of two consecutive states $(s_n, s_{n+1})$. Marginals of $\mu$ will also be denoted by $\mu$. Our purpose is to describe explicitly the asymptotic equilibrium payoff set. The feasible (long-run) payoff set is defined as

$$F := \text{co} \{ v \in \mathbb{R}^I | v = E_\mu[r(s, a)], \text{ some policy } \rho : S \rightarrow A \}.$$

When defining feasible payoffs, the restriction to deterministic policies rather than arbitrary strategies is clearly without loss. Recall also that a public randomization device is assumed, so that $F$ is convex.

3.1 A Superset of Bayes Nash Equilibrium Payoffs

This section provides a benchmark to which the set of truthful equilibrium payoffs is compared. Namely, we define a set of payoffs that includes the (limit) set of Bayes Nash equilibrium payoffs both in the original game and in the revelation game.

Fix some direction $\lambda \in \Lambda$, where $\Lambda := \{ \lambda \in \mathbb{R}^I : \|\lambda\| = 1 \}$. What is the highest score $\lambda \cdot v$ that can be achieved over all Bayes Nash equilibrium payoff vectors $v$?

If actions can be dictated, knowing the state profile can only help. But if $\lambda^i < 0$, this information would be used against $i$’s interests. Not surprisingly, player $i$ is unlikely to be forthcoming about this. This suggests distinguishing players in the set $I_+(\lambda) := \{ i : \lambda^i > 0 \}$ from the others. Suppose that players in $I(\lambda)$ truthfully disclose their private state, while the remaining players choose a reporting strategy that is independent of their private state.
Define
\[ \bar{k}(\lambda) := \max_\rho \mathbb{E}_\mu [\lambda \cdot r(s,a)], \]
where the maximum is over all policies \( \rho : \times_{i \in I_+(\lambda)} S_i \to A \) (with the convention that \( \rho \in A \) for \( I_+(\lambda) = \emptyset \)). Note that \( \mathbb{E}_\mu [\lambda \cdot r(s, \rho(s))] \) is the long-run payoff vector when players report truthfully and use the policy \( \rho \). Furthermore, let
\[ V^* := \cap_{\lambda \in \Lambda} \{ v \in \mathbb{R}^I \mid \lambda \cdot v \leq \bar{k}(\lambda) \}. \]
We call \( V^* \) the set of incentive-compatible payoffs. Clearly, \( V^* \subseteq F \). Note also that \( V^* \) depends on the transition matrix only via the invariant distribution. It turns out that the set \( V^* \) is an upper bound on the set of all equilibrium payoff vectors.

Let \( NE_\delta \) (resp., \( NE^R_\delta \)) denote the equilibrium payoffs in the original (resp., revelation) game, given \( \delta \in [0,1) \).

**Proposition 1** The limit sets of Bayes Nash equilibrium payoffs are contained in \( V^* \):
\[ \limsup_{\delta \to 1} NE_\delta \subseteq V^*, \quad \limsup_{\delta \to 1} NE^R_\delta \subseteq V^*. \]

**Proof.** The proof of this lemma and the following results are gathered in Appendix. Here we provide a sketch. Fix \( \lambda \in \Lambda \). Fix also \( \delta < 1 \) (and recall the prior \( \pi_1 \) at time 1). Consider the Bayes Nash equilibrium \( \sigma \) of the game (with discount factor \( \delta \)) with payoff vector \( v \) that maximizes \( \lambda \cdot v \) among all equilibria (where \( v^i \) is the expected payoff of player \( i \) given \( p_1 \)). This equilibrium need not be truthful or in pure strategies. Consider \( i \notin I_+(\lambda) \). Along with \( \sigma^{-i} \) and \( \pi_1 \), player \( i \)'s equilibrium strategy \( \sigma^i \) defines a distribution over histories. Fixing \( \sigma^{-i} \), let us consider an alternative strategy \( \tilde{\sigma}^i \) where player \( i \)'s reports are replaced by realizations of the public randomization device with the same distribution (round by round, conditional on the realizations so far), and player \( i \)'s action is determined by the randomization device as well, with the same conditional distribution (given the simulated reports) as \( \sigma^i \) would specify if this had been \( i \)'s report.\(^{11}\) The new profile \( (\sigma^{-i}, \tilde{\sigma}^i) \) need no longer be an equilibrium of the game. Yet, thanks to the IPV assumption, it gives players \( {i} \) the same payoff as \( \sigma^i \) and, thanks to the equilibrium property, it gives player \( i \) a weakly lower payoff. Most importantly,

\(^{11}\)To be slightly more formal: in a given round, the randomization device selects a report for player \( i \) according to the conditional distribution induced by \( \sigma^i \), given the public history so far. At the same time, the device selects an action for player \( i \) according to the distribution induced by \( \sigma^i \), given the public history, including reports of players \( -i \) and the simulated report for player \( i \). The strategy \( \tilde{\sigma}^i \) plays the action recommended by the device.
the strategy profile \((\sigma^{-i}, \tilde{\sigma}^i)\) no longer depends on the history of types of player \(i\). Clearly, this argument can be applied to all players \(i \notin I_+(\lambda)\) simultaneously, so that \(\lambda \cdot v\) is lower than the maximum inner product achieved over strategies that only depend on the history of types in \(I_+(\lambda)\). Maximizing this inner product over such strategies is a standard partially observable Markov decision problem, which admits a solution within the class of deterministic policies (on the state space \(\times_{i \in I_+(\lambda)} S_i \times \times_{i \notin I_+(\lambda)} \Delta(S_i)\)).

Because transitions do not depend on actions, the belief \(p_n \in \times_{i \notin I_+(\lambda)} \Delta(S_i)\) in round \(n\) about the states of players in \(I \setminus I_+(\lambda)\) converges to the ergodic distribution \((\mu^i)_{i \notin I_+(\lambda)}\) – an absorbing belief for the Markov chain. This defines a strategy that is only a function of the states \((s^i)_{i \in I_+(\lambda)}\) (the solution of the partially observable Markov decision problem evaluated at the belief \((\mu^i)_{i \notin I_+(\lambda)}\)).

Taking \(\delta \to 1\) yields that the limit set is included in \(\{v \in \mathbb{R}^I \mid \lambda \cdot v \leq \bar{k}(\lambda)\}\), and this is true for all \(\lambda \in \Lambda\). ■

As should be clear from the proof, Lemma 5 does not rely on \(M^i = S^i\) and holds for any message space.

The set \(V^*\) can be a strict subset of \(F\), as the following example illustrates.

**Example 1.** Each player \(i = 1, 2\) has two states \(s^i = \underline{s}^i, \overline{s}^i\). Rewards are given by Figure 1, with \(c(\underline{s}^i) = 2, c(\overline{s}^i) = 1\). (The interpretation is that a pie of total size 6 is obtained if at least one agent works; if both do only half the amount of work has to be put in by each worker. Their cost of working is fluctuating.) From one round to the next, a player’s state changes with probability \(p\), independently across players. Hence, the invariant distribution assigns equal weight to all four state profiles. Given that \(V^*\) only depends on the transition matrix via the invariant distribution, the specific value of \(p\) is irrelevant to compute \(V^*\) and \(F\), shown in Figure 2. Each player can secure \(3 - \frac{2 + 1}{2} = \frac{3}{2}\) by always working, so the actual equilibrium payoff set is smaller than \(V^*\).\(^{12}\)

\(^{12}\)In this particular example, the distinction between \(V^*\) and \(F\) turns out to be irrelevant once individual rationality is taken into account. Giving a third action to each player that yields both players a payoff of 0 independently of the state and the action of the opponent preserves the distinction.
A lower bound to $V^*$ is also readily obtained. Let $Ext^{po}$ denote the (weak) Pareto frontier of $F$. We write $Ext^{pu}$ for the set of payoff vectors obtained from pure state-independent action profiles, i.e. the set of vectors $v = E_{\mu/a}[r(s, a)]$ for some $\rho$ that takes a constant value in $A$. In their environment, Escobar and Toikka show that all individually rational (as defined below) payoffs in $co(Ext^{pu} \cup Ext^{po})$ are equilibrium payoffs (whenever this set has non-empty interior). Indeed, considering all directions $\lambda$ such that $I_+ (\lambda) \in \{\emptyset, I\}$, it follows that:

Lemma 1 It holds that $co (Ext^{pu} \cup Ext^{po}) \subset V^*$.

In Example 1, this lower bound is tight, but this is not always the case.

3.2 The Average Cost Optimality Equation

Our analysis makes use of the Average Cost Optimality Equation (ACOE) that plays an important role in dynamic programming. For completeness, we provide here an elementary statement, which is sufficient for our purpose and we refer to Puterman (1994) for details and additional properties.

Let be given an irreducible (or more generally unichain) transition function $q$ over the finite set $S$ with invariant measure $\mu$, and a payoff function $u : S \rightarrow \mathbb{R}$.\(^{13}\) Assume that

\(^{13}\)As is well known, the unichain assumption cannot be relaxed.
successive states \((s_n)\) follow a Markov chain with transition function \(q\) and that a decision-maker receives the reward \(u(s_n)\) in round \(n\). The long-run payoff of the decision-maker is \(v = \mathbb{E}_\mu[u(s)]\). While this long-run payoff is independent of the initial state, discounted payoffs are not. Lemma 2 below provides a normalized measure of the differences in discounted payoffs, for different initial states. Here and in what follows, \(t\) stands for the “next” state profile (“tomorrow’s” state), given the current state profile \(s\).\(^{14}\)

**Lemma 2** There is \(\theta : S \rightarrow \mathbb{R}\) such that

\[
v + \theta(s) = u(s) + \mathbb{E}_{t \sim p_s(\cdot)} \theta(t).
\]

The map \(\theta\) is unique, up to an additive constant. It admits an intuitive interpretation in terms of discounted payoffs. Indeed, the difference \(\theta(s) - \theta(s')\) is equal to \(\lim_{\delta \to 1} \frac{\gamma_\delta(s) - \gamma_\delta(s')}{1 - \delta}\), where \(\gamma_\delta(s)\) is the discounted payoff when starting for \(s\). For this reason, following standard terminology, call \(\theta\) the (vector of) relative values.

The map \(\theta\) provides a “one-shot” measure of the relative value of being in a given state; with persistent and possibly action-dependent transitions, the relative value is an essential ingredient in converting the dynamic game into a one-shot game, alongside the invariant measure \(\mu\). The former encapsulates the relevant information regarding future payoffs, while the latter is essential in aggregating the different one-shot games, parameterized by their states. Both \(\mu\) and \(\theta\) are usually defined as the solutions of a finite system of equations –the balance equations and the equations stated in Lemma 2. But in the ergodic case that we are concerned with, explicit formulas exist. (See, for instance, Iosifescu, 1980, p.123, for the invariant distribution; and Puterman, 1994, Appendix A for the relative values.)

### 3.3 Characterization

As mentioned, truthful equilibrium reduces to PPE in the case of repeated games with public monitoring. FL provide an algorithm to describe the limit set of PPE payoffs. Their characterization of the set of PPE payoff vectors, \(E_\delta\), as \(\delta \to 1\) relies on the notion of a score defined as follows. Recall that \(\Lambda\) denotes the unit sphere of \(\mathbb{R}^I\). We refer to \(\lambda \in \Lambda\) (or its coordinate \(\lambda^i\)) as weights, although the coordinates need not be nonnegative.

\(^{14}\)Lemma 2 defines the relative values for an exogenous Markov chain, or equivalently for a fixed policy. It is simply an “accounting” identity. The standard ACOE delivers more: given some Markov decision problem (MDP), a policy \(\rho\) is optimal if and only if, for all states \(s\), \(\rho(s)\) maximizes the right-hand side of the equations of Lemma 2. Both results will be invoked interchangeably.
Definition 2  Fix $\lambda \in \Lambda$. Let

$$k(\lambda) = \sup_{v, x, \alpha} \lambda \cdot v,$$

where the supremum is taken over all $v \in \mathbb{R}^I$, $x : Y \rightarrow \mathbb{R}^I$ and $\alpha \in \times_{i \in I} \Delta(A^i)$ such that

(i) $\alpha$ is a Nash equilibrium with payoff $v$ of the game with payoff function $r(a) + \sum y p_a(y) x(y)$;

(ii) For all $y \in Y$, it holds that $\lambda \cdot x(y) \leq 0$.

Let $H := \bigcap_{\lambda \in \Lambda} \{ v \in \mathbb{R}^I \mid \lambda \cdot v \leq k(\lambda) \}$. FL prove the following.

**Theorem 1 (FL)** It holds that $E_\delta \subseteq H$ for any $\delta < 1$; moreover, if $H$ has non-empty interior, then $\lim_{\delta \rightarrow 1} E_\delta = H$.

This theorem is extended by HSTV (2011) to the case of stochastic games with observable states. Our purpose is to obtain an algorithm for truthful equilibrium payoffs for the broader class of games considered here.

Because we insist on truthful equilibria, and because we need to incorporate the dynamic effects of actions on states, we must consider instead policies $\rho : S \rightarrow \times_{i \in I} \Delta(A^i)$ and transfers, such that reporting truthfully and playing $\rho$ constitutes a stationary equilibrium of the dynamic two-step game augmented with transfers. While policies depend only on current states, transfers will depend on the previous state and current public outcome.

In what follows, the set of public outcomes in a given round is $\Omega_{\text{pub}} := S \times A$ (where the $S$-components stand for the reports). Let a policy $\rho : S \rightarrow \times_{i \in I} \Delta(A^i)$, and transfers $x : S \times \Omega_{\text{pub}} \rightarrow \mathbb{R}^I$ be given. The vector $x(\bar{s}, \omega_{\text{pub}})$ is to be interpreted as transfers, contingent on previous reports $\bar{s}$, and on the current public outcome $\omega_{\text{pub}}$. Assuming states are truthfully reported and actions chosen according to $\rho$, the sequence $(\omega_n)$ of outcomes is a unichain Markov chain, and so is the sequence of pairs of reports $(s_{n-1}, s_n)$. Let $\theta_{\rho, r+x} : S \times S \rightarrow \mathbb{R}^I$ denote the relative values of the players, obtained when applying Lemma 2 to the latter chain (and to all players).

As FL, we start with an auxiliary one-shot game. We define $\Gamma(\rho, x)$ to be the one-shot Bayesian game with communication where:

(i) first, $(\bar{s}, s) \in S \times S$ is drawn according to $\mu$; each player $i$ is publicly told $\bar{s}$ and privately $s^i$;

---

15Conceptually, it might make sense to condition transfers on previous actions as well. This extension is not needed when transitions are action-independent.
(ii) each player $i$ reports publicly some state $m^i \in S^i$, then chooses an action $a^i \in A^i$.

The payoff vector is $r(s, a) + x(\bar{s}, \omega_{pub}) + \theta_{\rho, r+x}(m, t)$, where $\omega_{pub} := (m, a)$ and $t \sim p(\cdot | s)$.

Given $\lambda \in \Lambda$, we denote by $P_0(\lambda)$ the optimization program $\sup \lambda \cdot v$, where the supremum is computed over all payoff vectors $v \in \mathbb{R}^I$, policies $\rho : S \rightarrow \times_{i \in I} \Delta(A^i)$ and transfers $x : S \times \Omega_{pub} \rightarrow \mathbb{R}^I$ such that

(a) truth-telling followed by $\rho$ is a PBE outcome of $\Gamma(\rho, x)$, with expected payoff $v$;

(b) $\lambda \cdot x(\cdot) \leq 0$.

Condition (a) implies that for all $\bar{s}, s \in S$, the mixed profile $\rho(s)$ is a Nash equilibrium in the (complete information) game with payoff function $r(s, a) + x(\bar{s}, (s, a)) + \theta(s, t)$. It puts no restriction on equilibrium behavior following a lie at the report step.

The condition that $v$ be the equilibrium payoff in $\Gamma(\rho, x)$ writes

$$v = \mathbb{E}_{(\bar{s}, s) \sim \mu, a \sim \rho(s)} [r(s, a) + x(\bar{s}, \omega_{pub})],$$

where $\omega_{pub} = (s, a)$.

We denote by $k_0(\lambda)$ the value of $P_0(\lambda)$, and let $\mathcal{H}_0 := \{v \in \mathbb{R}^I, \lambda \cdot v \leq k_0(\lambda) \text{ for all } \lambda \in \Lambda\}$ be the convex set with support function $k_0$.

Theorem 2 below is the exact analog of FLM and HSTV, yet requires a (rather innocuous) non-degeneracy assumption.

Two states $s^i$ and $\bar{s}^i$ of player $i$ are equivalent if $r^j(s^i, \cdot) = r^j(\bar{s}^i, \cdot) + c$ for some $c \in \mathbb{R}$. In this section, we maintain the assumption that there is no player with two distinct, equivalent states.

Let $TE_\delta$ ($TE^R_\delta$) denote the set of truthful equilibrium payoffs in the original (resp., revelation) game.

**Theorem 2** Assume that $\mathcal{H}_0$ has non-empty interior. Then $\mathcal{H}_0$ is included in the limit set of truthful payoffs:

$$\mathcal{H}_0 \subseteq \liminf_{\delta \rightarrow 1} TE_\delta.$$

The same result holds true for the revelation game, weakening condition (b) in the definition of $\mathcal{H}_0(\lambda)$ by dropping the requirement that playing $\rho$ be optimal. Let $k^R_0$ denote the corresponding score.

For $i \in I$, define $\underline{v}^i := \min_{a^{-i} \in A^{-i}} \max_{\rho^i : S^i \rightarrow A^i} \mathbb{E}_{\mu} [r^i(s^i, (a^{-i}, \rho^i(s^i)))]$. 

17
Proposition 2  For every \( \lambda \neq -e^i \), \( k_0(\lambda) = k^R_0(\lambda) = \bar{k}(\lambda) \).
For \( \lambda = -e^i \), \( k_0(-e^i) \geq -v_i \), and \( k^R_0(-e^i) = \bar{k}(-e^i) \).

Set \( V^{**} := \{ v \in V^*, v \geq \underline{v} \} \). By Proposition 2, \( V^{**} \subseteq H_0 \). Hence Theorem 2 implies the following.

**Corollary 3** Assume that \( V^* \) has non-empty interior. Then

\[
\lim_{\delta \to 1} T E^R_\delta = V^*.
\]

Assume that \( V^{**} \) has non-empty interior. Then

\[
\liminf_{\delta \to 1} T E_\delta \supseteq V^{**}.
\]

This corollary (or Proposition 2) makes clear that the only restriction imposed by truthfulness, if any, lies in the lowest equilibrium payoff that can be attained.

### 3.4 Proof overview

Theorem 2 is reminiscent of the characterization of PBE payoffs in FLM (see also HSTV), and the proof in the Appendix follows the logic of existing proofs to the extent possible. Yet, the combination of private information and of state persistence significantly complicates the analysis. To motivate and introduce our technical innovations, we transpose below the recursive proof of FLM, and point out where difficulties arise, and how to cope with them.

Let \( Z \) be a compact set with a smooth boundary contained in the interior of \( H_0 \), and a discount factor \( \delta < 1 \) be given. Given a target payoff \( z \in Z \), we construct recursively a truthful PBE candidate with payoff \( z \).

Given a target payoff \( z_n \in Z \) in round \( n \), we set a direction \( \lambda_n \in \mathbb{R}^I \) to be (heuristically) “the” normal vector to \( Z \) at \( z_n \), and pick a feasible triple \((v_n, \rho_n, x_n)\) in \( P_0(\lambda_n) \) such that \( \lambda_n \cdot z_n < \lambda_n \cdot v_n \). The target payoff is publicly updated in round \( n + 1 \) to

\[
z_{n+1} := \frac{1}{\delta} z_n - \frac{1 - \delta}{\delta} v_n + \frac{1 - \delta}{\delta} x_n(m_{n-1}, \omega_{\text{pub}, n}),
\]

where \( \omega_{\text{pub}, n} = (m_n, y_n) \) is the public outcome in round \( n \). The equilibrium candidate \( \sigma \) reports truthfully in round \( n \) and selects actions according to \( \rho_n \).

As in FLM (see also HSTV), the choice of \( \lambda_n \) and of \((v_n, \rho_n, x_n)\) given \( z_n \) ensure that \( z_{n+1} \in Z \), so that this recursive construction is well-defined. Moreover, the expected continuation payoff in round \( n \) (computed as of round 1) is \( E_\sigma[z_n] \). Fix now a history \( h_n \) up to round \( n \).
along which all reports are truthful, with public part \( h_{\text{pub},n} \). The choice of \( (\rho_n, x_n) \) and the updating formula (1) also ensure that truthful reporting followed by \( \rho_n \) is a truthful PBE outcome of the Bayesian game\(^{16} \) with payoff

\[
(1 - \delta) r(s_n, a_n) + \delta z_{n+1}(h_{\text{pub},n}, \omega_{\text{pub},n}) + \delta \theta_{\rho_n, r+x_n}(s_n, s_{n+1}),
\]

and the induced equilibrium payoff is \( z_n(h_{\text{pub},n}) \).\(^{17} \)

In the FLM setup of repeated games with public monitoring, and in stochastic games as well, this is sufficient to imply that \( \sigma \) is a PPE. For dynamic Bayesian games, it is not quite enough, even in the setup of this section, and the latter property does not rule profitable deviations in round \( n \) following \( h_n \).

Indeed, under this construction, the pair \( (\rho_n, x_n) \) is updated in round \( n + 1 \), and the actual continuation relative values need not coincide with \( \theta_{\rho_n, r+x_n} \). Whether a specific state reached in round \( n + 1 \) is “good” relative to some other state depends on \( \rho_{n+1} \), hence on \( \lambda_{n+1} \) and therefore on \( a_n \) through \( y_n \). That is, even though player \( i \)'s action choice has no influence on the distribution of \( s^i_{n+1} \), it does affect the relative values of the different states in round \( n + 1 \). These changes in relative values would cancel in expectation, if states in round \( n + 1 \) were drawn using the invariant measure \( \mu \). Yet, player \( i \) is computing expectations based on \( s^i_n \). Hence, (conditional on \( \omega_{\text{pub},n} \)), player \( i \)'s continuation payoff is not exactly equal to \( z^i_{n+1}(h_{\text{pub},n}, \omega_{\text{pub},n}) \) and state persistence thus affects the incentives faced by player \( i \).\(^{18} \)

In our construction, the above sketch is amended along the following lines. We first prove that any feasible triple \( (v, \rho, x) \) in \( P_0(\lambda) \) can be perturbed into some other triple, for which truth-telling incentives are strict. In other words, the value of \( P_0(\lambda) \) is unchanged when truth-telling incentives are required to be strict. We then divide the play into a sequence of phases of random duration. In effect, the p.r.d. chooses in each round with probability \( \xi \), whether to start a new phase. When a new phase starts, target payoffs and policies are updated according to formulas derived from the FLM ones.

The switching probability \( \xi \) is set to be large compared to \( (1 - \delta) \), so that the expected contribution of a single phase to the overall payoff is small. Yet, \( \xi \) is set to be small, so that the expected duration of each phase is large.\(^ {19} \) The former property ensures that the recursive procedure is well-defined. The latter one ensures that, in any phase \( k \) under the plan \( (v(k), \rho(k), x(k)) \), players perceive future payoffs as a small perturbation of the relative values \( \theta_{\rho(k), r+x(k)} \). Given that truth-telling incentives are strict in \( \Gamma(\rho(k), x(k)) \), it thus remains

\(^{16} \)The prior belief is unambiguously derived from the public history.

\(^{17} \)when computed under \( \mu \).

\(^{18} \)The issue does not arise only when successive states are i.i.d. But then the dynamic game is truly the repetition of a one-shot Bayesian game, to which the results of FLM apply.

\(^{19} \)To be clear, we pick \( \xi(\delta) \) as a function of \( \delta \) such that \( \lim_{\delta \to 1} \xi(\delta) = 0 \) and \( \lim_{\delta \to 1} \frac{\xi(\delta)}{1 - \delta} = +\infty \).
optimal to report truthfully in the dynamic game.

4 Action-dependent Transitions, Imperfect Monitoring

We now generalize these results to the case in which monitoring is no longer perfect, and actions affect transitions. The environment is still of independent private values (IPV), which (cf. Section 2.1) requires that

\[ p(t, y | s, a) = p(y | a) \times \times_{i \in I} p_i(t^i | s^i, y), \]

as well as \( \pi_1(s) = \times_{i \in I} \pi^i_1(s^i). \)

Proofs for this section are in Appendix B.

4.1 The Superset Revisited

Example 2. There are two players. Incomplete information is one-sided: player 2 might be in state \( s = 0, 1 \). Player 2 has a single action, while player 1 chooses action \( a = 0, 1 \). Transitions are given by \( p(s_{n+1} = a | s_n = s, a_n = a) = 1/3 \), for all \( s = 0, 1 \). That is, the state is twice as likely to differ from the previous action chosen by player 1 as it is to coincide with this choice. As for rewards, \( r^2(s, a) = -1 \) if \( s = a, = 0 \) otherwise. Suppose that the objective is to minimize player 2’s payoff. We note that any constant strategy (i.e., \( a = 0 \) or \( a = 1 \) in all periods) yields a payoff of \(-1/3\), while a strategy that alternates deterministically between actions has a payoff that tends to \(-2/3\) as \( \delta \to 1 \).

This example demonstrates that constant action choices no longer suffice to minimize or maximize a player’s payoff, when his state is unknown to others and he fails to reveal it, even as \( \delta \to 1 \). Plainly, in the example, player 1’s belief about the state of player 2 matters for the choice of an optimal action, and the chosen action matters for player 1’s next belief. Hence, if we wish to describe player 1’s choice as a Markov policy, we must augment the state space to account for player 1’s belief. In the previous example, there is a binary sufficient statistic for this belief, namely, the last action chosen by player 1. Yet in general, the role of the belief is not summarized by such a simple statistic. It is necessary to augment the state space.
space by (at least) an arbitrary summary statistic, which follows a Markov chain as well. The next result establishes that finite representations suffice, under our assumptions.\textsuperscript{21}

We need to generalize the notion of a policy. Let a finite set $K$, and a map $\phi : K \times Y \rightarrow K$ be given. Together with $\phi$, any map $\rho : S \times K \rightarrow \Delta(A)$ induces a Markov chain $(s_n, k_n, a_n, y_n)$ over $S \times K \times A \times Y$. We refer to such a triple $\rho_{\text{ext}} = (\rho, K, \phi)$ as an extended policy. An extended policy is thus a policy that is possibly contingent on a public, extraneous and payoff irrelevant variable $k$ whose evolution is dictated by $y$. The extended policy $\rho_{\text{ext}}$ is unichain/irreducible if the latter chain is unichain/irreducible. We then denote by $\mu_{\rho_{\text{ext}}} \in \Delta((S \times K \times A \times Y)^2)$ the invariant distribution of successive states, actions and signals. Again, we will still denote by $\mu_{\rho_{\text{ext}}}$ various marginals of $\mu_{\rho_{\text{ext}}}$.

Given a direction $\lambda \in \Lambda$, let as before $I_+(\lambda) = \{i \in I, \lambda^i > 0\}$. We then set $\bar{k}_1(\lambda) := \sup_{\rho_{\text{ext}}} \mathbb{E}_{\mu_{\rho_{\text{ext}}}}[\lambda \cdot r(s, a)]$, where the supremum is taken over all pure unichain extended policies $\rho_{\text{ext}} = (\rho, K, \phi)$ such that $\rho : S \times K \rightarrow A$ depends on $s$ only through its components $s^i$, $i \in I_+(\lambda)$.

Let then $V_1^* := \{v \in \mathbb{R}^I, \lambda \cdot v \leq \bar{k}_1(\lambda) \text{ for all } \lambda \in \Lambda\}$, and denote by $NE_\delta(\pi_1)$ the set of Nash equilibrium equilibrium payoffs of the game with discount factor $\delta$, as a function of the initial distribution $\pi_1$.

**Proposition 3** Assume IPV. Then $\lim \sup_{\delta \rightarrow 1} NE_\delta(\pi_1) \subseteq V_1^*$, for all $\pi_1$.\textsuperscript{22}

### 4.2 Characterization

Given a unichain extended policy $\rho_{\text{ext}} = (\rho, K, \phi)$, the relevant set of public outcomes is $\Omega_{\text{pub}} = S \times K \times Y$. Let a map $x_{\text{ext}} : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \rightarrow \mathbb{R}^I$ be given. The vector $x(\tilde{\omega}_{\text{pub}}, \omega_{\text{pub}})$ is interpreted as transfers, contingent on the public outcomes in the previous and current rounds. Relative values associated with the pair $(\rho_{\text{ext}}, x_{\text{ext}})$ are thus maps $\theta_{\rho_{\text{ext}}, x_{\text{ext}}} : \Omega_{\text{pub}} \times S \times K \rightarrow \mathbb{R}^I$.

We then define $\Gamma(\rho_{\text{ext}}, x_{\text{ext}})$ to be the one-shot Bayesian game with communication where (i) $(\tilde{\omega}_{\text{pub}}, s, k) \in \Omega_{\text{pub}} \times S \times K$ is first drawn according to $\mu_{\rho_{\text{ext}}}$, (ii) each player $i$ is publicly told $\tilde{\omega}_{\text{pub}}$ (from which he deduces $k = \phi(\tilde{k}, \tilde{y})$) and privately told $s^i$, publicly reports some state $m^i \in S^i$, then chooses an action $a^i \in A^i$, and the payoff vector is

$$r(s, a) + x_{\text{ext}}(\tilde{\omega}_{\text{pub}}, \omega_{\text{pub}}) + \mathbb{E}_{(y,t) \sim p(\cdot | s,a)}[\theta_{\rho_{\text{ext}}, x_{\text{ext}}}](\omega_{\text{pub}}, t),$$

\textsuperscript{21}This is closely related to the literature on finite-state controllers in partially observable MDP, see Yu and Bertsekas, 2006.

\textsuperscript{22}A more precise statement holds. For each $\eta > 0$, there is $\bar{\delta} < 1$ such that, for each discount factor $\delta \geq \bar{\delta}$ and each initial distribution $\pi_1 \in \Delta(S^i)$, $NE_\delta(\pi_1)$ is included in the $\eta$-neighborhood $V_{1,\eta}^*$ of $V_1^*$.
with $\omega_{\text{pub}} = (m, k, y)$.

Given $\lambda \in \Lambda$, we denote by $\mathcal{P}_1(\lambda)$ the optimization program $\sup \lambda \cdot v$, where the supremum is over payoffs $v \in \mathbb{R}^I$, extended policies $\rho_{\text{ext}} = (\rho, K, \phi)$ and transfers $x_{\text{ext}} : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \to \mathbb{R}^I$, such that

(a) truth-telling followed by $\rho$ is a perfect Bayesian outcome of $\Gamma(\rho_{\text{ext}}, x_{\text{ext}})$ with expected payoff $v$;

(b) $\lambda \cdot x_{\text{ext}}(\cdot) \leq 0$.

We denote by $k_1(\lambda)$ the value of $\mathcal{P}_1(\lambda)$, and by $k_1^R(\lambda)$ the corresponding value when the requirement that obedience (namely, following $\rho$) be optimal is dropped.

As in the case of action-independent transitions and perfect monitoring, we prove our characterization result, Theorem 4 below, under a non-degeneracy assumption on payoffs, which we now introduce.

Given an action profile $a \in A$, let $\bar{a}$ be the policy which plays $a$ in each state profile $s \in S$. Observe that for $i \in I$ and $s \in S$, the relative value $\theta_{\bar{a},r}^i(s)$ is independent of $s^{\sim i}$ under IPV.

**A1** For all $i \in I$, $s^i \neq \tilde{s}^i \in S^i$, there exist action profiles $a, b \in A$, such that

$$
\theta_{\bar{a},r}^i(s^i) - \theta_{\bar{b},r}^i(s^i) \neq \theta_{\bar{a},r}^i(\tilde{s}^i) - \theta_{\bar{b},r}^i(\tilde{s}^i).
$$

(2)

When successive states are i.i.d., **A1** is equivalent to the assumption of no-two-equivalent states made in Section 3.3. However, when **A1** is specialized to the case of action-independent states, it neither implies nor is implied by this assumption.

In addition, we require the usual identifiability condition. In **A2**, $p$ refers to the marginal distribution over signals $y \in Y$ only. Let $Q^i(a) := \{p(\cdot | \hat{a}^i, a^{-i}) : \hat{a}^i \neq a^i\}$ be the distributions over signals $y$ induced by a unilateral deviation by $i$ at the action step, whether or not the reported state $s^i$ corresponds to the true state $\hat{s}^i$ or not. For simplicity, we make the assumption on all action profiles, rather than on the relevant subset.

**A2** For all $a \in A$,

\footnote{Yet, all results below still hold when **A1** is weakened and it is only required that (2) holds for some sequences $\bar{a} = (a_n)_n$ and $\tilde{b} = (b_n)_n$ in $A$ – at the cost of a slight extension of the notion of relative value, and of added notational complexity. The weakened assumption is strictly weaker than both **A1** and the no-two-equivalent-states assumption.}
1. For all $i \neq j$, $p(\cdot | a) \notin \text{co} \{Q^i(a) \cup Q^j(a)\}$.

2. For all $i \neq j$, $\text{co}(p(\cdot | a) \cup Q^i(a)) \cap \text{co}(p(\cdot | a) \cup Q^j(a)) = \{p(\cdot | a)\}$.

For $i \in I$, we set $v^i := \min_{a_i \in A} \max_{\rho^i,S^i \rightarrow A^i} E_{(s,a) \sim \mu_{\rho^i,a_i}}[r^i(s,a)]$. Proposition 4 and Theorem 4 parallel the results of Section 3.3.

**Proposition 4** Assume IPV. Then $k_1^R(\lambda) = \bar{k}_1(\lambda)$ for all $\lambda \in \Lambda$. Furthermore, under $A2$, $k_1(-e^i) \geq -v^i$ and $k_1(\lambda) = \bar{k}_1(\lambda)$ for all $\lambda \neq -e^i$.

**Theorem 4** Assume that IPV and Assumption $A1$. If $V_1^*$ has nonempty interior, then, for any $\pi_1$,

$$\lim_{\delta \rightarrow 1} T E_\delta^R(\pi_1) = V_1^*.$$  

If additionally Assumption $A2$ hold, and $V_1^{**}$ has non-empty interior, then

$$\lim \inf_{\delta \rightarrow 1} T E_\delta(\pi_1) \supseteq V_1^{**}.$$  

This theorem highlights once again that, under IPV, truthtelling is only restrictive as far as obedience goes: Assumption $A2$ ensures that deviations can be statistically detected, and the candidate payoff set must be truncated given individual rationality.

## 5 Product Monitoring

As mentioned, there are many examples in which the state-independent pure-strategy minmax payoff $v^i$ coincides with the “true” minmax payoff

$$w^i := \lim_{\delta \rightarrow 1} \min_{\sigma_i} \max_{\sigma_i} E \left[ (1 - \delta) \sum_{n \geq 1} \delta^{n-1} r^i_n \right],$$  

where the minimum is over the set of (independent) strategies by players $-i$. We denote by $\sigma_i$ the limiting strategy profile. (See Neyman 2008 for an analysis of the zero-sum undiscounted game when actions do not affect transitions.)

But the two do not coincide for all examples of economic interest. First, the state-independent pure-strategy minmax payoff rules out mixed strategies. Yet mixed strategies play a key role in some applications, e.g. the literature on tax auditing. More disturbingly, when $v^i > w^i$, it can happen that $V_1^{**} = \emptyset$. Theorem 4 becomes meaningless, as the corresponding equilibria no longer exist. On the other hand, the set

$$W := \{v \in V_1^* | v^i \geq w^i \text{ for all } i\}$$  

This definition of score can be sharpened by considering extended policies for players $-i$. 

24 This definition of score can be sharpened by considering extended policies for players $-i$. 

23
is never empty.\footnote{To see this, note that the state-independent mixed minmax payoff lies below the Pareto-frontier: clearly, the score in direction $\lambda^* = \frac{1}{\sqrt{I}}(1, \ldots, 1)$ of the payoff vector $\min_{a^{-i}} \max_{p^i : S^i \rightarrow A^i} E[p^i(s^i, a)]$ is less than $k(\lambda^*)$.}

As is also well known, even when attention is restricted to repeated games, there is no reason to expect the punishment level $w^i$ to equal the mixed-strategy minmax payoff commonly used (that lies in between $w^i$ and $\underline{w}^i$), as $w^i$ might only be obtained when players $-i$ use private strategies (depending on past action choices) that would allow for harder, coordinated punishments than those assumed in the definition of the mixed-strategy minmax payoff. Private histories may allow players $-i$ to correlate play unbeknownst to $i$. One special case in which they do coincide is when monitoring has a product structure, which rules out such correlation.\footnote{The scope for $w^i$ to coincide with the mixed minmax payoff is slightly larger, but not by much. See Gossner and Hörner (2010) for a characterization.} As this is the class of monitoring structures for which the standard folk theorem for repeated games is a characterization of (as opposed to a lower bound on) the equilibrium payoff set, we maintain this assumption throughout this section.

**Definition 3** Monitoring has product structure if there are finite sets $(Y^i)_{i=1}^I$ such that $Y = \times_i Y^i$, and $p(y \mid a) = \times_i p^i(y^i \mid a^i)$, for all $y = (y^1, \ldots, y^I) \in Y$, all $a \in A$.

As shown by FLM, product structure ensures that identifiability is implied by detectability, and that no further assumptions are required on the monitoring structure to enforce payoffs on the Pareto-frontier, hence to obtain a "Nash-threat" theorem. Our goal is to achieve a characterization of the equilibrium payoff set, so that an assumption on the monitoring structure remains necessary. We make the following assumption, which could certainly be refined.

**A3** For all $i, a$, $p(\cdot \mid a) \notin \text{co} Q^i(a)$.

Note that, given product structure, Assumption A3 is an assumption on $p^i$ only.

We maintain the nondegeneracy assumption introduced in Section 4.2. We prove that $W$ characterizes the (Bayes Nash, as well as sequential) equilibrium payoff set as $\delta \rightarrow 1$ in the IPV case. More formally:
Theorem 5 Assume that monitoring has the product structure, and that Assumption A1 and A3 hold. If $W$ has non-empty interior, the set of (Nash, sequential) equilibrium payoffs converges to $W$ as $\delta \to 1$.

Because minmaxing requires unpredictability, and as explained in Section 2.2, unpredictability might be inconsistent with truthful equilibria, this requires using strategies that are not truthful, at least during “punishments.”\textsuperscript{27} Nonetheless, we show that a slight extension of the set of strategies considered so far, to allow for silent play during punishment-like phases, suffices.

Unlike in repeated games, imposing product structure does not guarantee that the minmax strategy is stationary: players $-i$ draw inferences from the public signal $y^i$ about player $i$’s action, hence about his private state, which can be exploited to adjust the next punishment action. Our construction relies on an extension of Theorem 2, as well as an argument inspired by Gossner (1995), based on approachability theory (Blackwell, 1956). Roughly speaking, the argument is divided in two parts. First, one must extend Theorem 2 to allow for “blocks” of $T$ rounds, rather than single rounds, as the extensive form over which the score is computed. This part is delicate; in particular, the directions $-e^i$ –for which such aggregation is necessary– cannot be treated in isolation, as $\Lambda \setminus \{-e^i\}$ would no longer be compact, a property that is important in the proof of Theorem 2. Second, considering such a block in which player $i$, say, is “punished” (that is, a block corresponding to the direction $-e^i$), one must devise transfers $x$ at the end of the block, as a function of the public history, that makes players $-i$ willing to play the minmax strategy, or at least some strategy profile achieving approximately the same payoff to player $i$. The difficulty, illustrated by Example 2, is that typically there are no transfers making player $i$ indifferent over a subset of actions for different types of his simultaneously; yet minmaxing might require precisely as much.

To ensure that the distribution over action profiles during the punishment phase matches the theoretical one (computed using the realized actions taken by player $i$), we design a statistical test that a player $j \neq i$ can pass with very high probability (by conforming to the minmax strategy, for instance), independently of the other players’ strategies; and that he is very likely to fail if the distribution of his realized signals departs too much from the one that his minmax strategy would yield.\textsuperscript{28} When testing player $j$, it is critical to condition on player $i$’s realized signal, so as to incentivize player $j$ to be unpredictable.

\textsuperscript{27}We use quotation marks as there are no clearly defined punishment phases in recursive constructions (as in APS or here), unlike in the standard proof of the folk theorem under perfect monitoring.

\textsuperscript{28}This is where the IPV assumption and product monitoring are used. It ensures that player $j$’s minmax strategy can be taken to be independent of his private information, hence adapted to the public information.
6 Dropping the IPV Assumption

The IPV assumption simplifies the analysis considerably. Yet neither the independence nor the private values assumption are necessary to derive a result in the spirit of Theorem 2. Several complications arise, which reflect both new opportunities and difficulties. With correlated states, for instance, one might like to use player \(-i\)'s reports as statistical evidence in evaluating the truthfulness of player \(i\)'s report, which suggests expanding the domain of transfers of the one-shot Bayesian game, and making it easier to induce truth-telling. On the other hand, under common values, player \(i\)'s payoff is no longer independent of player \(-i\)'s state, conditional on the marginal distribution of player \(-i\)'s action. Hence, fixing this marginal distribution, player \(i\)'s incentives depend on whether player \(-i\) reports his state truthfully, which might make truth-telling harder to sustain. The next two examples illustrate.

**Example 3—A Silent Game.** This game follows Renault (2006). This is a zero-sum two-player game in which player 1 has two private states, \(s^1\) and \(\hat{s}^1\), and player 2 has a single state, omitted. Player 1 has actions \(A^1 = \{T, B\}\) and player 2 has actions \(A^2 = \{L, R\}\). Player 1’s reward is given by Figure 1. Recall that rewards are not observed. States \(s^1\)

\[
\begin{array}{c|cc}
T & L & R \\
\hline
B & 1 & 0 \\
\end{array}
\quad
\begin{array}{c|cc}
T & L & R \\
\hline
B & 0 & 0 \\
\end{array}
\]

\[
\begin{array}{c|cc}
\hat{s}^1 & L & R \\
\hline
\end{array}
\quad
\begin{array}{c|cc}
\hat{s}^1 & L & R \\
\hline
\end{array}
\]

Figure 3: Player 1’s reward in Example 3

and \(\hat{s}^1\) are equally likely in the initial round, and transitions are action-independent, with \(p \in [1/2, 1)\) denoting the probability that the state remains unchanged from one round to the next.

Pick \(M^1\) such that \(#M^1 \geq 2\), so that player 1 can disclose his state if he wishes. Will he? By revealing it, player 2 can secure a payoff of 0 by playing \(R\) or \(L\) depending on player 1’s report. Yet player 1 can secure 1/4 by choosing reports and actions at random. In fact, this is the (uniform) value for \(p = 1\) (Aumann and Maschler, 1995). When \(p < 1\), player 1 can get more than this by trading off the higher expected reward from a given action with the information that it gives away. He has no interest in giving this information away for free through informative reports. Truthful equilibria do not exist: all equilibria are babbling.
Just because we may focus on the silent game does not make it easier. Its (limit) value for arbitrary $p > 2/3$ is still unknown.\textsuperscript{29} Because the optimal strategies depend on player 2’s belief about player 1’s state, the problem of solving for them is infinite-dimensional, and all that can be done is to characterize its solution via some functional equation (see Hörner, Rosenberg, Solan and Vieille, 2010).

Non-existence of truthful equilibria in some games is no surprise. The tension between truth-telling and lack of commitment also arises in bargaining and contracting, giving rise to the ratchet effect (see Freixas, Guesnerie and Tirole, 1985). What Example 1 illustrates is that small message spaces are just as difficult to deal with as larger ones. When players hide their information, their behavior reflects their private beliefs, which calls for a state space as large as it gets.

**Example 4—Waiting for Evidence.** There are two players. Player 1 has $K + 1$ types, $S^1 = \{0, 1, \ldots, K\}$; player 2 has only two types, $S^2 = \{0, 1\}$. Transitions do not depend on actions (omitted), and are as follows. If $s^1_n = k > 0$, then $s^2_n = 0$ and $s^1_{n+1} = s^1_n - 1$. If $s^1_n = 0$, then $s^2_n = 1$ and $s^1_{n+1}$ is drawn randomly (and uniformly) from $S^1$. In words, $s^1_n$ stands for the number of rounds until the next occurrence of $s^2 = 1$. By waiting no more than $K$ rounds, all reports by player 1 can be verified.

This example makes two related points. First, in order for player $-i$ to statistically discriminate between player $i$’s states, it is not necessary that his set of signals (here, players $-i$’s states) be as rich as player $i$’s, unlike in static mechanism design with correlated types (the familiar “spanning condition” of Crémer and McLean, 1988, generically satisfied if only if $|S^{-i}| \geq |S^i|$). Two states for one player can be enough to cross-check the reports of an opponent with many more states, provided that states in later rounds are informative enough.

Second, the long-term dependence of the stochastic process implies that one player’s report should not always be evaluated on the fly. It is better to hold off until more evidence is collected. Note that this is not the same kind of delay as the one that makes review strategies effective, taking advantage of the central limit theorem to devise powerful tests even when signals are independently distributed over time (see Radner, 1986; Fang and Norman, 2006; Jackson and Sonnenschein, 2007). It is precisely because of the dependence

\textsuperscript{29}It is known for $p \in [1/2, 2/3]$ and some specific values. Peški and Toikka (2014) have recently shown that this value is non-increasing in $p$, and Bressaud and Quas (2014) have determined the optimal strategies for values of $p$ up to $\sim .7323$.\textsuperscript{27}
that waiting is useful here.

This raises an interesting statistical question: does the tail of the sequence of private states of player \(-i\) contain indispensable information in evaluating the truthfulness of player \(i\)'s report in a given round, or is the distribution of this infinite sequence, conditional on \((s_n^i, s_{n-1})\), summarized by the distribution of an initial segment of the sequence? This question appears to be open in general. In the case of transitions that do not depend on actions, it has been raised by Blackwell and Koopmans (1957) and answered by Gilbert (1959): it is enough to consider the next \(2|S^i| + 1\) values of the sequence \((s_{n'}^i)_{n' \geq n} \).30

At the very least, when types are correlated and the Markov chain exhibits time dependence, it is useful to condition player \(i\)'s continuation payoff given his report about \(s_n^i\) on \(-i\)'s next private state, \(s_{n+1}^{-i}\). Because this suffices to obtain sufficient conditions analogous to those invoked in the static case, we limit ourselves to this conditioning in this section.31

6.1 A General Theorem

In this section, \(M^i := S^i \times A^i \times S^i\) for all \(i\). This has to be interpreted as player \(i\)'s state yesterday, his action yesterday, and his state today. In the spirit of Myerson (1986), we wish to allow player \(i\) to disclose all information that is relevant to his preferences and beliefs; in this case, with correlated types, his belief about \(-i\)'s type profile depends on the action he has taken, his type yesterday and his current type. Off path, none of these are known, and a player shouldn’t find it impossible to disclose his beliefs if he happened to deviate in the previous round.

A profile \(m\) of reports is written \(m = (m_p, m_a, m_c)\), where \(m_p\) (resp. \(m_c\)) is interpreted as the report profile on previous (resp. current) states, and \(m_a\) is the reported (last round) action profile.

We set \(\Omega_{\text{pub}} := M \times Y\), and we refer to the pair \((m_n, y_n)\) as the public outcome of round \(n\). This is the additional public information available at the end of round \(n\). We also refer to \((s_n, m_n, a_n, y_n)\) as the outcome of round \(n\), and denote by \(\Omega := \Omega_{\text{pub}} \times S \times A\) the set of possible outcomes in any given round.

Let a policy \(\rho : S \to \Delta(A)\), and transfers \(x : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S \to \mathbb{R}^I\) be given. We will

---

30The reporting strategy defines a hidden Markov chain on pairs of states, reports and signals that induces a stationary process over reports and signals; Gilbert assumes that the hidden Markov chain is irreducible and aperiodic, which here need not be (with truthful reporting, the report is equal to the state), but his result continues to hold when these assumptions are dropped, see for instance Dharmadhikari (1963).

31See Obara (2008) for some of the difficulties encountered in dynamic settings when attempting to extend results from static mechanism design with correlated types.
assume that for each $i \in I$, $x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)$ is independent of $i$’s own state $t^i$.\footnote{This requirement will not be systematically stated, but it is assumed throughout.} Note that, compared to IPV, we have added the current state profile $t^{-i}$ as an argument of player $i$’s transfer, given that this profile is statistical evidence about player $i$’s state, as explained in Example 4.

Assuming states are truthfully reported and actions chosen according to $\rho$, the sequence $(\omega_n)$ of outcomes is a unichain Markov chain, and so is the sequence $(\bar{\omega}_n)$, where $\bar{\omega}_n = (\omega_{\text{pub},n-1}, m_n)$, with transition function denoted $\pi_\rho$, and with invariant measure $\mu_\rho$.

Let $\theta_{\rho,r+x} : \Omega_{\text{pub}} \times M \to \mathbb{R}^I$ denote the relative values of the players, obtained when applying Lemma 2 to the latter chain (and to all players).\footnote{There is here a slight and innocuous abuse of notation: $\theta_{\rho,r+x}$ solves the equations $v + \theta(\bar{\omega}_{\text{pub}}, m) = r(s, \rho(s)) + \mathbf{E}[x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta(\omega_{\text{pub}}, m')]$, where $v = \mathbf{E}_{\mu_\rho}[r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)]$ is the long-run payoff under $\rho$.}

Thanks to the ACOE, the condition that reporting truthfully and playing $\rho$ is a stationary equilibrium of the dynamic game with stage payoffs $r + x$ can to some extent be rephrased as saying that, for each $\bar{\omega}_{\text{pub}} \in \Omega_{\text{pub}}$, reporting truthfully and playing $\rho$ is an equilibrium in the one-shot Bayesian game in which states $s$ are drawn according to $p$ (given $\bar{\omega}_{\text{pub}}$), players submit reports $m$, then choose actions $a$, and obtain the (random) payoff

$$r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta_{\rho,r+x}(\omega_{\text{pub}}, m'),$$

where $(y, t)$ are chosen according to $p_{s,a}$ and $\omega_{\text{pub}} = (m, y)$.

However, because we insist on off-path truth-telling, we need to consider arbitrary private histories, and the formal condition is therefore more involved. Fix a player $i$. Given a triple $(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i)$, let $D_{\rho,x}(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i)$ denote the two-step decision problem in which

**Step 1** $s \in S$ is drawn according to the belief held by player $i$;\footnote{Recall that player $i$ assumes that players $-i$ report truthfully and play $\rho^{-i}$. Hence player $i$ assigns probability 1 to $\tilde{s}^{-i} = \tilde{m}_c^{-i}$, and to previous actions being drawn according to $\rho^{-i}(\tilde{m}_c)$; hence this belief assigns to $s \in S$ the probability $p_{\tilde{s},\rho}(s | \tilde{y})$. This is the case unless $\tilde{y}$ is inconsistent with $\rho^{-i}(\tilde{m}_c)$; if this is the case, use the same updating rule with some other arbitrary $\tilde{a}^{-i}$ such that $\tilde{y} \in Y(\tilde{a}^{-i}, \tilde{a}^i)$.} player $i$ is informed of $s^i$, then submits a report $m' \in M^i$;

**Step 2** player $i$ learns current states $s^{-i}$ from the opponents’ reports $m^{-i} = (\tilde{m}_c^{-i}, \tilde{a}^{-i}, s^{-i})$, and then chooses an action $a^i \in A^i$. The payoff to player $i$ is given by

$$r^i(s, a) + x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t^{-i}) + \theta_{\rho,r+x}(\omega_{\text{pub}}, m'),$$

where $a^{-i}$ is drawn according to $\rho^{-i}(s^{-i}, m_c^i)$ and the pair $(y, t)$ is drawn according to $p_{s,a}$, and $\omega_{\text{pub}} := (m, y)$.
We denote by $D_{\rho,x}^i$ the collection of decision problems $D_{\rho,x}^i(\bar{\omega}_{\text{pub}}, \bar{s}; \bar{a})$.

**Definition 4** The pair $(\rho, x)$ is admissible if all optimal strategies of player $i$ in $D_{\rho,x}^i$ report truthfully $m^i = (\bar{s}^i, \bar{a}^i, s^i)$ in Step 1 (Truth-telling); then, in Step 2, conditional on all players reporting truthfully in Step 1, $\rho'(s)$ is a (not necessarily unique) optimal mixed action (Obedience).

Some comments are in order. The condition that $\rho$ be played once states (not necessarily types) have been reported truthfully simply means that, for each $\bar{\omega}_{\text{pub}}$ and $m = (\bar{s}, \bar{a}, s)$ the action profile $\rho(s)$ is an equilibrium of the complete information one-shot game with payoff function $r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta_{\rho,r+x}(\omega_{\text{pub}}, m')$.

The truth-telling condition is slightly more delicate to interpret. Consider first an outcome $\bar{\omega} \in \Omega$ such that $\bar{s}^i = \bar{m}^i_c$ and $\bar{a}^i = \rho^i(\bar{s})$ for all $i$—no player has lied or deviated in the previous round, assuming the action to be played was pure. Given such an outcome, all players share the same belief over next types, given by $p_{s,a}(\cdot | \bar{y})$. Consider the Bayesian game in which (i) $s \in S$ is drawn according to the latter distribution, (ii) players make reports $m$, then choose actions $a$, and (iii) get the payoff $r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta_{\rho,r+x}(\omega_{\text{pub}}, m')$. The admissibility condition for such an outcome $\bar{\omega}$ is equivalent to requiring that truth-telling followed by $\rho$ is an equilibrium of this Bayesian game, with “strict” incentives at the reporting step.\(^{35}\)

The admissibility requirement in Definition 4 is demanding, however, in that it requires in addition truth-telling to be optimal for player $i$ at any outcome $\bar{\omega}$ such that $(\bar{s}^{-i}, \bar{a}^{-i}) = (\bar{m}^{-i}_c, \rho^{-i}(\bar{m}_c))$, but $\bar{s}^i \neq \bar{m}^i_c$ (or $\bar{a}^i \neq \rho^i(\bar{m}_c)$). Following such outcomes, players do not share the same belief over the next states. The same issue arises if the action profile $\rho^i(\bar{m}_c)$ is mixed. Therefore, it is inconvenient to state the admissibility requirement by means of a simple, subjective Bayesian game—hence the formulation in terms of a decision problem.

In loose terms, truth-telling is the unique best-reply at the reporting step of player $i$ to truth-telling and $\rho^{-i}$. Note that we require truth-telling to be optimal $(m^i = (\bar{s}^i, \bar{a}^i, s^i))$ even if player $i$ did misreport his previous state ($\bar{m}^i_c \neq \bar{s}^i$). On the other hand, Definition 4 puts no restriction on player $i$’s behavior if he lies in Step 1 ($m^i \neq (\bar{s}^i, \bar{a}^i, s^i)$). The second part of Definition 4 is equivalent to saying that $\rho'(s)$ is one best-reply to $\rho^{-i}(s)$ in the complete information game with payoff function given by (3) when $m = (\bar{s}, \bar{a}, s)$.

\(^{35}\)Quotation marks are needed, since we have not defined off-path behavior. What we mean is that any on-path deviation at the reporting step leads to a lower payoff, no matter what action is then taken.

30
The requirement that truth-telling be uniquely optimal reflects an important difference between our approach to Bayesian games and the traditional approach of APS in repeated games. In the case of repeated games, continuation play is summarized by the continuation payoff. Here, the future does not only affect incentives via the long-run continuation payoff, but also via the relative values. However, we do not know of a simple relationship between \( v \) and \( \theta \). Our construction involves “repeated games” strategies that are “approximately” policies, so that \( \theta \) can be derived from \((\rho, x)\). This shifts the emphasis from payoffs to policies, and requires us to implement a specific policy. Truth-telling incentives must be strict for the approximation involved not to affect them.\(^{36}\)

We denote by \( C_0 \) the set of admissible pairs \((\rho, x)\).

For given weights \( \lambda \in \Lambda \), we denote by \( P_0(\lambda) \) the optimization program \( \sup \lambda \cdot v \), where the supremum is taken over all triples \((v, \rho, x)\) such that

- \((\rho, x) \in C_0;\)
- \( \lambda \cdot x(\cdot) \leq 0; \)
- \( v = \mathbf{E}_{\mu_{\rho}}[r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)] \), where \( \mu_{\rho} \in \Delta(\Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S) \) is the invariant distribution under truth-telling and \( \rho \), so that \( v \) is the long-run payoff induced by \((\rho, x)\).

The three conditions mirror those of Definition 1 for the case of repeated games. The first condition (admissibility) and the third condition are the counterparts of the Nash condition in Definition 1(i); the second condition is the “budget-balance” requirement imposed by Definition 1(ii). We denote by \( k_0(\lambda) \) the value of \( P_0(\lambda) \) and set \( \mathcal{H}_0 := \{v \in \mathbb{R}^I, \lambda \cdot v \leq k_0(\lambda) \text{ for all } \lambda \in \Lambda\} \).

**Theorem 6** Assume that \( \mathcal{H}_0 \) has non-empty interior. Then, given \( \pi_1 \),

\[
\mathcal{H}_0 \subseteq \liminf_{\delta \to 1} \mathcal{T}E_\delta(\pi_1).
\]

This result is simple enough. For instance, in the case of “standard” repeated games with public monitoring, Theorem 6 generalizes FLM, yielding the folk theorem with the mixed minmax under their assumptions.

To be clear, there is no reason to expect Theorem 6 to provide a characterization of the entire limit set of truthful equilibrium payoffs. One might hope to achieve a larger set of

\(^{36}\)Fortunately, this requirement is not demanding, as it is implied by standard full-rank conditions in the correlated case, and by our non-degeneracy condition in the IPV case.
payoffs by employing finer statistical tests (using the serial correlation in states, for instance), just as one can achieve a bigger set of equilibrium payoffs in repeated games than the set of PPE payoffs, by considering statistical tests (and private strategies). Example 4 makes plain that using only the current report of \( -i \) as evidence for player \( i \)'s truthfulness is ad hoc. Allowing for more signals/reports comes at an obvious cost in terms of the simplicity of the characterization.

Nonetheless, as we have shown in Sections 3–5, variants of this theorem suffice to establish “folk theorems” under IPV. Similarly, with correlated types, one can use arguments based on Crémers and McLean (1988) and Kosenok and Severinov (2008) to derive a folk theorem with appropriate full rank assumptions. See the working paper for details. But Example 3 illustrates the difficulties that arise under the ominous combination of independent types and common values.

Two variations to this theorem are worth mentioning. First, Theorem 6 can be adapted to the case in which some of the players are short-run, whether or not such players have private information (in which case, assume that it is independent across rounds). As this is a standard feature of such characterizations (see FL, for instance), we will be brief. Suppose that players \( i \in LR = \{1, \ldots, L\} \), \( L \leq I \) are long-run players, whose preferences are as before, with discount factor \( \delta < 1 \). Players \( j \in SR = \{L + 1, \ldots, I\} \) are short-run players, each representative of which plays only once. We consider a “Stackelberg” structure, common in economic applications, in which long-run players make their reports first, thereupon the short-run players do as well (if they have any private information), and we set \( M^i = S^i \) for the short-run players. Actions are simultaneous. Let \( m^{LR} \in M^{LR} = \times_{i=1}^L M^i \) denote an arbitrary report by the long-run players. Given a policy \( \rho^{LR} : M \to \times_{i \in LR} \Delta(A^i) \) of the long-run players, mapping reports \( m = (m^{LR}, s^{SR}) \) (with \( s^{SR} = (s^{L+1}, \ldots, s^I) \)) into mixed actions, we let \( B(m^{LR}, \rho^{LR}) \) denote the best-reply correspondence of the short-run players, namely, the sequential equilibria of the two-step game (reports and actions) between players in \( SR \). We then modify the definition of admissible pair \( (\rho, x) \) so as to require that the reports and actions of the short-run players be in \( B(m^{LR}, \rho^{LR}) \) for all reports \( m^{LR} \) by the long-run players, where \( \rho^{LR} \) is the restriction of \( \rho \) to players in \( LR \). The requirements on the long-run players are the same as in Definition 4.

Second, signals can be private. That is, we may replace Step 2 of the decision problem \( D^i_{\rho, x} \) by: A profile \( y_n = (y^i_n) \in Y := \times_i Y^i \) of private signals and the next state profile \( s_{n+1} = (s^i_{n+1})_{i \in I} \) are drawn according to some joint distribution \( p_{s_{n+1}} \in \Delta(S \times Y) \). We then re-define a message \( m^i \) as including: player \( i \)'s state, action and signal in the last period, and player \( i \)'s current state. Transfers are then assumed to depend on the past, current and next
message profile, with the restriction, as with public monitoring, that player \( i \)'s transfer does not depend on his own future message, only on player \(-i\)'s. The definition of admissibility remains the same, given the re-defined message space, and so does the statement of the theorem.

In a sense, this more general formulation is also more natural, as the current one already reduces the program to a one-player decision-theoretic problem, in which each player must report his private information; he might as well report the signal he observed, and the payoff he received, in case of known-own payoffs. This variation mirrors Kandori and Matsushima (1998)'s extension of FLM to private monitoring; the issues that they raise regarding the possibility of a folk theorem in truthful strategies under imperfect information apply here as well.

7 Conclusion

This paper has considered a class of equilibria in games with private and imperfectly persistent information. While the structure of equilibria has been assumed to be relatively simple, to preserve tractability—in particular, we have mostly focused on truthful equilibria— it has been shown, perhaps surprisingly, that in the case of independent private values this is not restrictive as far as incentives go: all that transfers depend on are the current and the previous report. This confirms a rather natural intuition: in terms of equilibrium payoffs at least (and as far as incentive-compatibility is concerned), there is nothing to gain from aggregating information beyond transition counts. In the case of correlated values, we have shown how the standard insights from static mechanism design with correlated values generalize; in this case as well, the standard “genericity” conditions (in terms of numbers of states) suffice, provided next round’s reports by a player’s opponent are used.

Open questions remain. As explained, the payoff set identified in Theorem 6 is a subset of the set of truthful equilibria. As our characterization in the IPV case when monitoring has a product structure makes clear, this theorem can be extended to yield equilibrium payoff sets that are larger than the truthful equilibrium payoff set, but without such tweaking, it is unclear how large the gap is. If possible, an exact characterization of the truthful equilibrium payoff set (as \( \delta \to 1 \)) would be very useful. In particular, this would provide us with a better understanding of the circumstances under which existence obtains. It is striking that it does in the two important cases that are well-understood in the static case: independent private values and correlated types. Given how little is known in static mechanism design when neither assumption is satisfied, perhaps one should not hope for too much in the dynamic
case. Instead, one might hope to prove directly that such equilibria exist in large classes of games, such as games with known-own payoffs (private values, without the independence assumption).

A different but equally important question is what can be said about the dynamic Bayesian game under alternative assumptions on the communication opportunities. At one extreme, one might like to know what can be achieved without communication; at the other extreme, how to extend the analysis to the case in which a mediator is available.

References


A Perfect Monitoring Action-Independent Transitions: Proofs

A.1 Proof of Proposition 1

We here prove Proposition 1. We assume that the distribution of the initial state is µ and let a discount factor δ < 1 be given. Given a direction λ ∈ Λ, let σ be a Nash equilibrium of the game with payoff vector v ∈ R^I that maximizes λ · v among all equilibria (truthful or not) of the game. The equilibrium σ need not be truthful or in pure strategies. Consider a player i ∈ I_+(λ). Along with σ^-i, player i’s equilibrium strategy σ^i defines a distribution over histories. Let us consider an alternative strategy ñ^i where player i’s reports are replaced by realizations of the public randomization device with the same distribution (round by round, conditional on the realizations so far), and player i’s action is determined by the
randomization device as well, with the same conditional distribution (given the simulated reports) as \( \sigma^i \) would specify if this had been \( i \)'s report.\(^{37}\) The new profile \((\sigma^{-i}, \tilde{\sigma}^i)\) need no longer be an equilibrium of the game. Yet, thanks to the IPV assumption, it gives players \(-i\) the same payoff as \( \sigma \) and, thanks to the equilibrium property, it gives player \( i \) a weakly lower payoff. Most importantly, the strategy profile \((\sigma^{-i}, \tilde{\sigma}^i)\) no longer depends on the history of types of player \( i \). Clearly, this argument can be applied to all players \( i \notin I_+(\lambda) \) simultaneously, so that \( \lambda \cdot v \) is lower than the maximum inner product achieved over strategies that only depend on the history of types in \( I_+(\lambda) \). Since the distribution of the profile of states is equal to \( \mu \) in every round, the value of the latter maximization problem is equal to \( \bar{k}(\lambda) \).

### A.2 Proof of Theorem 2

We let \( Z \) be a compact set included in the interior of \( H_0 \). Given \( z \in Z \), we construct a truthful PBE \( \sigma \) with payoff \( z \). Under \( \sigma \), the play is divided into a sequence of phases of random duration. During any given phase, the players (report truthfully and) follow a policy \( \rho_\lambda : S \to A \) that depends on a direction \( \lambda \in \Lambda \). Players are incentivized to report truthfully and to follow the prescribed policy by means of “transfers,” which are implemented via adjustments in the continuation payoff, updated at the beginning of each phase.

#### A.2.1 Preliminaries

We pick \( \eta > 0 \) small enough so that the \( \eta \)-neighborhood \( Z_\eta := \{ z \in \mathbb{R}^I, d(z, Z) \leq \eta \} \) is also included in the interior of \( H_0 \). Since \( k_0 \) is continuous, there exists \( \varepsilon_0 > 0 \) such that \( \max_{z \in Z_\eta} \lambda \cdot z + 2\varepsilon_0 < k_0(\lambda) \) for all \( \lambda \in \Lambda \).

We quote without proof a classical result, which relies on the smoothness of \( Z_\eta \) (see Lemma 6 in HSTV for a related statement).

**Lemma 3** Given \( \varepsilon > 0 \), there exists \( \zeta > 0 \) such that the following holds. For every \( z \in Z_\eta \), there exists a direction \( \lambda \in \Lambda \) such that any vector \( w \in \mathbb{R}^I \) which satisfies \( \|w - z\| \leq \zeta \) and \( \lambda \cdot w \leq \lambda \cdot z - \varepsilon \zeta \) for some \( \zeta < \tilde{\zeta} \), belongs to \( Z_\eta \).

---

\(^{37}\)To be slightly more formal: in a given round, the randomization device selects a report for player \( i \) according to the conditional distribution induced by \( \sigma^i \), given the public history so far. At the same time, the device selects an action for player \( i \) according to the distribution induced by \( \sigma^i \), given the public history, including reports of players \(-i\) and the simulated report for player \( i \). The strategy \( \tilde{\sigma}^i \) plays the action recommended by the device.
Lemma 4 There is a finite set $S$ of triples $(v, \rho, x)$ such that the following holds. For every direction $\lambda \in \Lambda$, there is an element $(v, \rho, x)$ of $S$ such that (i) $(v, \rho, x)$ is feasible in $P_0(\lambda)$ with strict truth-telling incentives and (ii) $\max_{z \in \mathbb{Z}^n} \lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.

Proof. For each player $i \in I$, and any two states $s^i, \tilde{s}^i \in S^i$, there exist $a, b \in A$ such that $r^i(s^i, a) - r^i(s^i, b) > r^i(\tilde{s}^i, a) - r^i(\tilde{s}^i, b)$. This implies the existence of a family of correlated distributions $\rho_i(s^i) \in \Delta(A)$ ($s^i \in S^i$), and of a map $x^i : S^i \to \mathbb{R}$ such that

$$r^i(s^i, \rho_i(\tilde{s}^i)) + x^i(\tilde{s}^i) < r^i(s^i, \rho_i(s^i)) + x^i(s^i)$$

for every $s^i \neq \tilde{s}^i$ (see Lemma 2 in Abreu, Dutta and Smith (1994)).

We next define $\rho_s : S \to \Delta(A)$ as $\rho_s(s) := \frac{1}{|I|} \sum_{i \in I} \rho_i(s^i)$ and $x_t : S \to \mathbb{R}^I$ as $x_t(s) := \frac{1}{|I|} x^i(s^i)$. We then define $x_{ob} : \Omega_{pub} \to \mathbb{R}^I$ as $x_{ob}(s, a) = 0$ if $a^i = \rho^i_s(s)$ and set $x_{ob}(s, a)$ to be a large negative constant otherwise, and set $x_s := x_t + x_{ob}$. Recall that $\theta_{\rho_s, r+x_s} : S \to \mathbb{R}^I$ are the relative values under $\rho_s$ when transfers are $x_s$. For each player $i \in I$ and state profile $s \in S$, the sum of $r^i$ and $x^i_t$ has a strict maximum when reporting truthfully $\tilde{s}^i = s^i$ and playing $\rho^i_s(s)$. Since transitions are action-independent, this implies that for each $i$ and $s \in S$, the expected sum

$$r^i(s^i, (a^i, \rho^i_s(\tilde{s}^i, s^{-i}))) + x^i_t((\tilde{s}^i, s^{-i}), (a^i, \rho^i_s(m^i, s^{-i}))) + E_{r \sim p(\cdot | s, a^i, \rho^i_s(s))} \theta_{\rho_s, r+x_s}((\tilde{s}^i, s^{-i}), t)$$

of the current payoffs, transfers $x^i_t$ and continuation relative values $\theta_{\rho_s, r+x_s}$ has a strict maximum for $\tilde{s}^i = s^i$ and $\tilde{a}^i = \rho^i_s(s)$.

Let a direction $\lambda \in \Lambda$ be given, and subtract a constant to $x_s(\cdot)$ in order that $\lambda \cdot x_s(\cdot) < 0$. The long-run payoff associated with $(\rho_s, x_s)$ is $v_s := E_{s \sim \mu, a \sim \rho_s(s)} [r(s, a) + x_s(s, a)]$. The triple $(v_s, \rho_s, x_s)$ is then feasible in $P_0(\lambda')$ for all $\lambda'$ close enough to $\lambda$, with strict truth-telling incentives.

Let now $(v, \rho, x)$ be a feasible triple in $P_0(\lambda)$ such that $\lambda \cdot x(\cdot) < 0$ and $\lambda \cdot v > k_0(\lambda) - \varepsilon_0$. For $\varepsilon > 0$, we denote by $(\rho_\varepsilon, x_\varepsilon)$ the pair obtained when letting the p.r.d. choose between $(\rho, x)$ and $(\rho_s, x_s)$ with probabilities $1 - \varepsilon$ and $\varepsilon$ respectively. The long-run payoff associated to the pair $(\rho_\varepsilon, x_\varepsilon)$ is $v_\varepsilon := (1 - \varepsilon) v + \varepsilon v_s$.

---

38 Plainly, this is meaningful provided we view the p.r.d. as picking a pure action profile according to $\rho_\varepsilon(s)$.

39 Assuming players $-i$ report truthfully and play $\rho^i_s$. 39
Observe that, since transitions are action-independent, one has \( \theta_{\rho,v+x} (s, s') = (1 - \varepsilon) \theta_{\rho,v} (s, s') + \varepsilon \theta_{\rho,x} (s) \) for all \((s, s')\). Using once again the assumption that transitions are action-independent, this is easily seen to imply that the triple \((v_\varepsilon, \rho_\varepsilon, x_\varepsilon)\) is feasible in \(P_0(\tilde{\lambda})\) for all \(\tilde{\lambda}\) close to \(\lambda\), with strict truth-telling incentives.

In addition \(\lambda \cdot v_\varepsilon \geq k_0(\lambda) - \varepsilon_0 > \sup_{z \in Z_n} \lambda \cdot z + \varepsilon_0\) for \(\varepsilon\) small enough. The result follows, since \(\Lambda\) is compact and \(k_0\) continuous. ■

We let \(\kappa\) be a common bound on \(v, x,\) and \(\theta_{\rho,v+x}\) for \((v, \rho, x) \in \mathcal{S}\), and on \(z \in Z\) and \(r\). We pick an arbitrary \(\varepsilon_1 \in (0, \varepsilon_0)\), and set \(\varepsilon := \varepsilon_1/4\kappa\). We let then \(\tilde{\zeta}\) be obtained via Lemma 3 given \(\varepsilon\).

We assume that \(\bar{\delta} < 1\) is high enough so that the conditions (i–iv) are met for all \(\delta \geq \bar{\delta}\):

(i) \(\xi := \sqrt{1 - \delta} < \frac{1}{3}\), (ii) \(\frac{1 - \delta}{\delta \xi} < 1\), (iii) \(\zeta := 4\kappa \frac{1 - \delta}{\delta \xi} < \tilde{\zeta}\) and (iv) \(4\kappa \xi < \varepsilon_0 - \varepsilon_1\).

### A.2.2 Strategies

For simplicity, we assume that the initial state \(s_1\), together with a fictitious state \(s_0\) for round 0, is drawn according to \(\mu\). Let \(z \in Z\) be the desired equilibrium payoff. The play is divided into a sequence of phases. The durations of the successive phases form a sequence of \(i.i.d\). random variables. The initial round \(\tau_i(1)\) of phase \(k, k \geq 1\), is set as follows: \(\tau(1) = 0\); in each round \(n\), the p.r.d. decides with probability \(\xi\) whether to start a new phase.\(^{40}\)

In round \(\tau_i(1)\), a target payoff \(z_0(1) \in \mathbb{R}^I\), a direction \(\lambda_{0}(1) \in \Lambda\), and a triple \((v_0(1), \rho_0(1), x_0(1)) \in \mathcal{S}\) are updated based on past public play, together with an auxiliary target \(w_0(1) \in \mathbb{R}^I\).

We first update \(w_0(1)\) according to

\[
\xi w_0(1) + (1 - \xi) z_0 = \frac{1}{\delta} z_0(1) - \frac{1 - \delta}{\delta} v_0 + \frac{1 - \delta}{\delta} x_0 \cdot (m_{n-2}, \omega_{\text{pub}, n-1})
\]

(4)

where \(n = \tau_i(1)\).

Next, we apply Lemma 3 with \(z = w_0(1)\) to get \(\lambda_{0}(1)\) and we next apply Lemma 4 with \(\lambda_{0}(1)\) to get \((v_0(1), \rho_0(1), x_0(1)) \in \mathcal{S}\). Finally, we update \(z_0(1)\) to

\[
z_0(1) := w_0(1) + (1 - \delta) \left(1 + \frac{1 - \delta}{\delta \xi}\right) \theta_{(k)}(m_{n-1}, m_n) - \theta_{(k+1)}(m_{n-1}, m_n)
\]

(5)

where \(\theta_{(k)}\) is a short-hand notation for the relative values \(\theta_{\rho(k), v+x(k)}\) associated with the pair \((\rho, x(k))\).

\(^{40}\)Thus, the duration \(\Delta_k := \tau_{i+1(k)} - \tau_i(k)\) of phase \(k\) is such that \(\Delta_k\) follows a geometric distribution with parameter \(1 - \xi\).
Updating takes place right after the outcome of the p.r.d. is observed. The form of the left-hand side in (4) accounts for the random duration of the phases. The auxiliary variable \( w_{(k+1)} \) and the extra term in (5) (when compared to FLM) serve to adjust continuation relative values along the play, as will be apparent.

The construction is initialized with \( w_{(1)} = z \), which is used to define \( \lambda_{(1)} \), and \((v_{(1)}, \rho_{(1)}, x_{(1)})\) and \( z_{(1)} \) using (5) (with \( \theta_{(0)} := 0 \)). That this recursive construction is well-defined follows from Lemma 5 below.

**Lemma 5** For all \( k \) (and all public histories), one has \( w_{(k)} \in Z_\eta \).

**Proof.** Observe that \( \| w_{(k)} - z_{(k)} \| \leq 3\kappa(1 - \delta) \) by (5) and \( \| w_{(k+1)} - z_{(k)} \| \leq 3\kappa \frac{1 - \delta}{\xi} \) by (4) whenever \( w_{(k)} \) and \( z_{(k)} \) are defined, so that

\[
\| w_{(k+1)} - w_{(k)} \| \leq 3\kappa(1 - \delta) \left( 1 + \frac{1}{\xi} \right) \leq \zeta.
\]

Observe also that

\[
w_{(k+1)} - w_{(k)} = w_{(k+1)} - z_{(k)} + z_{(k)} - w_{(k)} = \frac{1 - \delta}{\delta\xi} \left\{ z_{(k)} - v_{(k)} + x_{(k)}(m_{n-2}, \omega_{\text{pub},n-1}) \right\} + z_{(k)} - w_{(k)} = \frac{1 - \delta}{\delta\xi} \left( w_{(k)} - v_{(k)} + x_{(k)}(m_{n-2}, \omega_{\text{pub},n-1}) \right) + \left( 1 + \frac{1 - \delta}{\delta\xi} \right) (z_{(k)} - w_{(k)})
\]

so that

\[
\lambda_{(k)} \cdot (w_{(k+1)} - w_{(k)}) \leq -\frac{1 - \delta}{\delta\xi} \varepsilon_0 + 4\kappa(1 - \delta)
\]

\[
\leq -\frac{1 - \delta}{\delta\xi} \varepsilon_1 + \frac{1 - \delta}{\delta\xi} (\varepsilon_1 - \varepsilon_0 + 4\kappa\delta\xi)
\]

\[
\leq -\varepsilon\zeta.
\]

Hence \( w_{(k+1)} \in Z_\eta \) as soon as \( w_{(k)} \in Z_\eta \). 

Given a round \( n \in [\tau(k); \tau(k+1) - 1] \) in phase \( k \), we let \( z_n := z_{(k)} \) stand for the target payoff in round \( n \), and set \((v_n, \rho_n, x_n, \theta_n) := (v_{(k)}, \rho_{(k)}, x_{(k)}, \theta_{\rho_{(k)}}x_{(k)})\). Note that \( z_n \) is measurable w.r.t. the public history available in round \( n \), including the outcome of the p.r.d.

Under \( \sigma \), each player \( i \) reports truthfully \( m_n^i = s_n^i \) at the report step. At the action step, player \( i \) plays \( \rho_n^i(m_n) \) if he reported truthfully \( m_n^i = s_n^i \). In the (off-path) event \( m_n^i \neq s_n^i \), player \( i \) plays a best-reply, denoted \( \alpha_n^i(s_n^i, m_n) \), to \( \rho^{-i}(m_n) \) in the complete information game with payoff \( r(s_n, a) + x_n(m_{n-1}, (m_n, a)) + \mathbb{E}_{s_{n+1} \sim \rho_n(\cdot|m_n)}[\theta_n(m_n, s_{n+1})] \).
A.2.3 Equilibrium properties

Given a round $n$, we denote by $\gamma_n$ the expected continuation payoff under $\sigma$, conditional on the public history at round $n$ (up to and including the outcome of the p.r.d.).\footnote{That is, denote by $\mathcal{H}_{pub,n}$ the (round $n$, public information) algebra on plays, by $P_\sigma$ the probability distribution over plays induced by $\sigma$, and by $E_\sigma$ the expectation operator under $P_\sigma$. Then $\gamma_n := (1 - \delta)E_\sigma \left[ \sum_{u=0}^{+\infty} \delta u r_{n+u} \mid \mathcal{H}_{pub,n} \right]$ in particular, $\gamma_n$ is computed under the "assumption" that $m_n = s_n$.}

**Lemma 6** One has $\gamma_n = z_n + (1 - \delta)\theta_n$.

**Proof.** Given a public history $h_{pub,n}$ (again including the outcome of the p.r.d. in round $n$), $\gamma_n$ satisfies the recursive equation

$$\gamma_n(h_{pub,n}) = (1 - \delta)r(s_n, \rho_n(s_n)) + \delta E[\gamma_{n+1} \mid h_{pub,n}]$$

where the expectation is computed over $s_{n+1} \sim p(\cdot \mid s_n)$ and over the outcome of the p.r.d. in round $n + 1$.

We prove that the sequence $(z_n + (1 - \delta)\theta_n)_n$ obeys the same recursion, that is,

$$z_n + (1 - \delta)\theta_n = (1 - \delta)r(s_n, \rho_n(s_n)) + \delta E[z_{n+1} + (1 - \delta)\theta_{n+1} \mid h_{pub,n}]$$

(9)

The claim will follow (since both sequences are bounded, a contraction argument applies).

Let $\bar{h}_{pub,n+1} = (h_{pub,n}, a_n, s_{n+1})$ be an arbitrary public extension of $h_{pub,n}$ up to round $n + 1$, ending prior to the outcome of the p.r.d. in round $n + 1$. After $\bar{h}_{pub,n+1}$, the p.r.d. chooses with probability $\xi$ whether $z_{n+1}$ is equal to $z_{k+1}$ or to $z_{(k)}$. (Abusing notations), the expectation $E[z_{n+1} + (1 - \delta)\theta_{n+1} \mid \bar{h}_{pub,n+1}]$ over the outcome of the p.r.d. is therefore

$$\begin{align*}
(1 - \xi) & \left( z_{(k)} + (1 - \delta)\theta_{(k)}(s_n, s_{n+1}) \right) + \xi \left( z_{(k+1)} + (1 - \delta)\theta_{(k+1)}(s_n, s_{n+1}) \right) \\
= & \left(1 - \xi\right) \left( z_{(k)} + (1 - \delta)\theta_{(k)}(s_n, s_{n+1}) \right) + \xi \left( w_{(k+1)} + (1 - \delta) \left(1 + \frac{1 - \delta}{\delta} \right) \theta_{(k)}(s_n, s_{n+1}) \right) \\
= & \frac{1 - \delta}{\delta} \theta_{(k)}(s_n, s_{n+1}) + \left( \frac{1}{\delta} z_{(k)} - \frac{1 - \delta}{\delta} w_{(k)} + \frac{1 - \delta}{\delta} \theta_{(k)}(s_{n-1}, \omega_{pub,n}) \right)
\end{align*}$$

(10)

while the first equality holds by virtue of (5) and the second one by (4).

Taking expectations over $\bar{h}_{pub,n+1}$ conditional on $h_{pub,n}$, the RHS in (9) is

$$\begin{align*}
& (1 - \delta)r(s_n, \rho_n(s_n)) + \delta E[z_{n+1} + (1 - \delta)\theta_{n+1} \mid h_{pub,n}] \\
= & z_{(k)} + (1 - \delta) \left\{ E_{a_n \sim \rho_n(s_n), s_{n+1} \sim p(\cdot \mid s_n)} \left[ r(s_n, a_n) + x_{(k)}(s_{n-1}, \omega_{pub,n}) + \theta_{(k)}(s_n, s_{n+1}) \right] - v_{(k)} \right\} \\
= & z_{(k)} + (1 - \delta)\theta_{(k)}(s_{n-1}, s_n) = z_n + (1 - \delta)\theta_n,
\end{align*}$$

as desired. ■
Corollary 7 $\sigma$ is a truthful PBE with expected payoff $z$.

Proof. We check that player $i$ has no profitable one-round deviation in round $n$. Let be given a private history $h_n^i$ of player $i$ up to round $n$, including the realization of $s_n^i$, and denote by $h_{\text{pub},n}$ the public part of $h_n^i$. We compute the expected continuation of player $i$ when first reporting $m_n^i$, next choosing an action contingent on reports according to some map $\beta^i : S \to A^i$, and finally switching back to $\sigma^i$.

Fix the realizations $s_n^{-i} = m_n^{-i}$ of the other players’ types, and proceed as in the proof of the previous claim. The equalities in (10) are algebraic identities, and still hold when substituting $m_n^i$ to $s_n^i$. The equality (11) also remains valid, with the appropriate changes. Specifically, the expected continuation payoff of player $i$ is given by

$$z_n^i - v_n^i + (1-\delta) \left\{ E_{a_n \sim (\beta(m_n),\rho_{n^{-i}}(m_n))} \left( r^i(s_n^i, a_n) + x_n^i(m_{n-1}, \omega_{\text{pub},n}) \right) + E_{s_{n+1} \sim p(\cdot|s_n)} \theta_n^i(m_n, s_{n+1}) \right\}.$$

Taking now the expectation over $s_n^{-i}$ it thus appears that the expected continuation payoff of player $i$ given $h_n^i$, is equal (up to the constant term $z_n^i - v_n^i$) to the interim expected payoff of $i$ in the game $\Gamma(\rho_n, x_n)$ when reporting $m_n^i$ and playing $\beta^i$ (given $(m_n^{-i}, s_n^{-i})$ and $s_n^i$). The result then follows from the optimality of truth-telling and $\rho_n$ in the game $\Gamma(\rho_n, x_n)$. ■

A.3 Proof of Proposition 2

We here prove Proposition 2. Let first the direction $\lambda$ be equal to $-\epsilon^i$, for some $i \in I$. Let $\bar{a}^{-i} \in A^{-i}$ and $\bar{\rho}^i : S^i \to A^i$ achieve the min max in the definition of $\bar{\nu}^i$, and define $\bar{\rho}^{-i} : S^i \to A^{-i}$ as $\bar{\rho}^{-i}(s^i) = \bar{a}^{-i}$. Let $x_{\bar{\rho}} : A \to \mathbb{R}^i$ be transfers such that (i) $x^j_{\bar{\rho}}(\cdot) = 0$ and, for $j \neq i$, (ii) $x^j_{\bar{\rho}}(a) = 0$ if $a^{-i} = \bar{a}^{-i}$ and $x^j_{\bar{\rho}}(a)$ is a large negative number otherwise. Then $(\bar{\nu}^i, \bar{\rho}, x_{\bar{\rho}})$ is feasible in $\mathcal{P}_0(-\epsilon^i)$. Therefore $k_0(-\epsilon^i) \geq -\underline{\nu}^i$, as desired.

We now fix $\lambda \in \Lambda$, with $\lambda \neq -\epsilon^i$. We will prove that $k_0(\lambda) \geq \bar{k}(\lambda)$. Recall that $I_+ := I_+(\lambda) = \{ i \in I, \lambda^i > 0 \}$, and consider the MDP with state space $S_+ := \times_{i \in I_+} S^i$ and stage reward

$$r_\lambda(s_+, a) := \sum_{i \in I_+} \lambda^i r^i(s^i, a) + \sum_{i \notin I_+} \lambda^i r^i(\mu^i, a).$$

The (long-run) value of this MDP is equal to $\bar{k}(\lambda)$ and we let $\theta_\lambda : S_+ \to \mathbb{R}$ denote the associated relative value so that

$$\bar{k}(\lambda) + \theta_\lambda(s_+) = \max_{a \in A} \{ r_\lambda(s_+, a) \} + E_{t_+ \sim p(\cdot|s_+)}[\theta_\lambda(t_+)] \quad (12)$$

43
for all \( s_+ \in S_+ \). Pure optimal policies \( \rho : S_+ \to A \) are characterized by the property that \( \rho(s_+) \) achieves the maximum in (12) for each \( s_+ \in S_+ \).

We let \( \rho_\lambda : S_+ \to A \) be an arbitrary optimal policy. We construct transfers \( x : \delta \times \Omega_{\text{pub}} \to \mathbb{R}^I \) such that \( (\bar{\kappa}(\lambda), \rho_\lambda, x) \) is feasible in \( \mathcal{P}_0(\lambda) \), thereby showing \( k_0(\lambda) \geq \bar{\kappa}(\lambda) \).\(^{42}\)

The transfers \( x \) are obtained as the sum of transfers \( x_t : \delta \times S \to \mathbb{R}^I \), which are contingent on successive reports and provide truth-telling incentives, and of transfers inducing obedience. The transfers \( x_t \) are defined in two steps. We first define transfers \( x_1 \) of the VCG type, contingent on current reports, and rely next on AS to balance the transfers.

**Claim 8** There exists \( x_1 : \delta \to \mathbb{R}^I \) such that \((\rho_\lambda, x_1)\) is truthful.

**Proof.** For \( i \not\in I_+ \), it suffices to set \( x_1^i = 0 \) as the reports by \( i \) are ignored. Fix now \( i \in I_+(\lambda) \). For \( s_+ \in S_+ \), define \( x_1^i(s_+) \) by the equation

\[
\lambda^i x_1^i(s_+) := r_\lambda(s_+, \rho_\lambda(s_+)) - \lambda^i r_1^i(s^i, \rho_\lambda(s_+)).
\]

Observe that \( \lambda^i \theta_{r_+ x_1} : S_+ \to \mathbb{R} \) satisfies (12) as well. Hence, \( \lambda^i \theta_{r_+ x_1} = \theta_\lambda \) up to an additive constant.

Since \( \rho_\lambda(s_+) \) achieves the maximum in (12), it follows that \((\rho_\lambda, x_1)\) is truthful. \( \blacksquare \)

Note that \( x_1 = 0 \) if \( \lambda = e^i \) for some \( i \), hence the lemma holds with \( x_t = x_1 \). From now on, we assume that \( \lambda \) is not a coordinate vector and adapt the arguments from AS to the long-run setup.

For \( i \in I \), define first \( x_2^i : \delta \times S^i \to \mathbb{R} \) by \( x_2^i(\bar{s}, s^i) := E_{s^i \sim \rho(\cdot | \bar{s} = \bar{s})} [x_1^i(s)] \). Plainly, \((\rho_\lambda, x_2)\) is truthful as well. The relative values \( \theta_{x_2}^i : \delta \times S^i \to \mathbb{R} \) solve

\[
\gamma + \theta_{x_2}^i(\bar{s}, s^i) = x_2^i(\bar{s}, s^i) + E_{s^i \sim \rho(\cdot | \bar{s} = \bar{s})} [\theta_{x_2}^i(s, t^i)]
\]

where \( \gamma = E_{(\bar{s}, s^i) \sim \mu} [x_2^i(\bar{s}, s^i)] \).

Define next \( x_3^i : \delta \times S^i \to \mathbb{R} \) as

\[
x_3^i(\bar{s}, s^i) := \theta_{x_2}^i(\bar{s}, s^i) - E_{\bar{s}^i \sim \rho(\cdot | s^i)} [\theta_{x_2}^i(\bar{s}, \bar{s}^i)].
\]

**Claim 9** The pair \((\rho_\lambda, x_3)\) is truthful.

\(^{42}\)In this section, we only deal with the policy \( \rho_\lambda \), and will drop the reference to \( \rho_\lambda \) when writing relative values.
Claim 10

Let \( \bar{s} \) be equal to \( E \). Then \( \bar{t} \) is equal to \( E \).

In view of (13), the expectation of (15) (and therefore of (14)) when \( \bar{s} = \theta \), thanks to the additive term \( E_{\bar{s} \sim p(\cdot | \bar{s})} [\theta_{x_2}(\bar{s}, \bar{s})] \) which is independent of \( m^i \). Thanks to the equality \( x_3 = \theta_{x_3} \), the former expression is in turn equal to

\[
E_{\bar{s} \sim p(\cdot | \bar{s})} [\theta_{x_2}(\bar{s}, \bar{s})] + \theta_{x_3}(m, t^i)
\]

(15)

In view of (13), the expectation of (15) (and therefore of (14)) when \( s^i \sim p(\cdot | s^i) \) and \( t^i \sim p(\cdot | s^i) \), is equal to the expectation of

\[
r^i(s^i, \rho_\lambda(m)) + x_2^i(s^i, m^i) + \theta_{x_3}(m, t^i)
\]

Since \( (\rho_\lambda, x_2) \) is truthful, so is \( (\rho_\lambda, x_3) \). □

Proof. \( \mathbf{E}_{\bar{s} \sim p(\cdot | \bar{s})} [x_3^i(\bar{s}, s^i)] = 0 \) for each \( \bar{s} \), hence the equality

\[
x_3^i(\bar{s}, s^i) = x_3^i(\bar{s}, s^i) + E_{\bar{s} \sim p(\cdot | \bar{s}), t^i \sim p(\cdot | s^i)} [x_3^i(s, t^i)]
\]

holds. That is, \( x_3^i = \theta_{x_3} \).

Fix \( s \in S, s^i \in S^i \) and \( m^i \in S^i \). For given \( s^{-i} \in S^{-i}, t^i \in S^i \), and setting \( m := (s^{-i}, m^i) \), the expression

\[
r^i(s^i, \rho_\lambda(m)) + x_3^i(s^i, m^i) + \theta_{x_3}(m, t^i)
\]

is equal to

\[
r^i(s^i, \rho_\lambda(m)) + \theta_{x_2}(s^i, s^i) + \theta_{x_3}(m, t^i) + \theta_{x_3}(m, t^i)
\]

(14)

(up to the additive term \( E_{\bar{s} \sim p(\cdot | \bar{s})} [\theta_{x_2}(\bar{s}, \bar{s})] \) which is independent of \( m^i \)). Thanks to the equality \( x_3 = \theta_{x_3} \), the former expression is in turn equal to

\[
r^i(s^i, \rho_\lambda(m)) + \theta_{x_2}(s^i, s^i) + \theta_{x_3}(m, t^i) + \theta_{x_3}(m, t^i) - E_{\bar{s} \sim p(\cdot | m^i)} [\theta_{x_3}(m, \bar{t})]
\]

(15)

In view of (13), the expectation of (15) (and therefore of (14)) when \( s^{-i} \sim p(\cdot | s^{-i}) \) and \( t^i \sim p(\cdot | s^i) \), is equal to the expectation of

\[
r^i(s^i, \rho_\lambda(m)) + x_2^i(s^i, m^i) + \theta_{x_3}(m, t^i).
\]

Since \( (\rho_\lambda, x_2) \) is truthful, so is \( (\rho_\lambda, x_3) \). □

Claim 10

Let \( \mu_{ij} \in \mathbb{R} \) be arbitrary. For \( i \in I \), set

\[
x_4^i(\bar{s}, s) := x_3^i(\bar{s}, s^i) + \sum_{j \neq i} \mu_{ij} x_3^j(\bar{s}, s^i).
\]

Then \( (\rho_\lambda, x_4) \) is truthful.

Proof. Fix \( i \in I, \bar{s} \in S, s^i \in S^i \). For fixed \( s^{-i} = m^{-i} \) and \( t \), the expression

\[
r^i(s^i, \rho_\lambda(m)) + x_4^i(s^i, m) + \theta_{x_3}(m, t)
\]

(16)

is equal, thanks to \( \theta_{x_3} = x_3^i \), to

\[
r^i(s^i, \rho_\lambda(m)) + x_3^i(s^i, m^i) + \theta_{x_3}(m, t) + \sum_{j \neq i} \mu_{ij} \left( x_3^j(s, m^j) + x_3^j(m, t^j) \right).
\]

Observe that in the latter expression, and for fixed \( j \neq i \), \( x_3^j(s, m^j) \) is independent of \( m^i \) and \( E_{\bar{s} \sim p(\cdot | \bar{s})} [x_3^j(m, t^j)] = 0 \). Since \( (\rho_\lambda, x_3) \) is truthful, so is \( (\rho_\lambda, x_4) \). □
Since \( \lambda \) is not a coordinate vector, the system \( \lambda_i + \sum_{j \neq i} \lambda_j \mu_{ji} = 0 \) (\( i \in I \)) has a solution \((\mu_{ij})\). With this choice, \( \lambda \cdot x_A(\cdot) = 0 \).

We finally add transfers inducing obedience. Since \( \lambda \) is not a coordinate direction, there exist transfers \( x_{\rho_\lambda} : S \times A \to \mathbb{R}^I \) such that (i) \( \lambda \cdot x_{\rho_\lambda}(\cdot) = 0 \), (ii) \( x_{\rho_\lambda}(s, \rho_\lambda(s)) = 0 \) for each \( s \in S \) and (iii) \( x^i(s, a^i, \rho^-_\lambda(s)) \) is a large negative constant for each \( s \in S \), \( i \in I \), and \( a^i \neq \rho_\lambda^i(s) \).

The triple \((\bar{k}(\lambda), \rho_\lambda, x_A + x_{\rho_\lambda})\) is feasible in \( P_0(\lambda) \).

To conclude, we provide a short proof of the reverse inequality \( k_0(\lambda) \leq \bar{k}(\lambda) \) for all \( \lambda \in \Lambda \) (which is not needed for deriving Corollary ??). The proof uses the same idea as the proof of Proposition 1. Let \((v, \rho, x)\) be feasible in \( P_0(\lambda) \). We modify \( \rho \) and \( x \) by letting the p.r.d. pick a fictitious report \( \bar{s}^j \sim \mu^j \) for all \( j \notin I_+ \) and let actions and transfers be determined by \( \rho \) and \( x \), using these fictitious reports and the actual reports of players \( i \in I_+ \). Denote by \( \tilde{\rho} : S_+ \to x_{i \in I_+} \Delta(A^i) \) the modified policy and by \( \tilde{x} \) the modified transfers. Since \((v, \rho, x)\) is feasible and thanks to the private values assumption, all players \( j \notin I_+ \) are weakly worse off in \( \Gamma(\tilde{\rho}, \tilde{x}) \) while players \( i \in I_+ \) are unaffected. Since \( \lambda \cdot \tilde{x}(\cdot) \leq 0 \), this implies \( \lambda \cdot v \leq E_{\bar{s} \sim \mu}[\lambda \cdot r(s, \tilde{\rho}(s))] \leq \bar{k}(\lambda) \).

### B Action-dependent Transitions, Imperfect Monitoring: Proofs

#### B.1 Proof of Proposition 3

We here prove Proposition 3. Fix a direction \( \lambda \in \Lambda \) and a discount factor \( \delta < 1 \). We set \( I_+ := I_+(\lambda) \), \( I_- := I \setminus I_+ \), and \( \Delta_- := \times_{i \in I_-} \Delta(S^i) \). We introduce the MDP \( \mathcal{M}_\lambda \) in which players jointly maximize the \( \lambda \)-weighted sum of discounted payoffs, and ignore the states of players \( i \in I_- \). Formally, the state space of \( \mathcal{M}_\lambda \) is \( S_+ \times \Delta_- \) with elements denoted \((s_+, \pi_-)\), and the action set is \( A \). The transitions given \((s_+, \pi_-)\) and conditional on \( y \), are deduced from \( p \). With obvious notations, the stage reward is

\[
r_\lambda((s_+, \pi_-), a) := \sum_{i \in I_+} \lambda^i r^i(s^i, a) + \sum_{i \in I_-} \lambda^i r^i(\pi^i, a).
\]

We denote by \( v_\delta(s_+, \pi_-) \) the value of the \( \delta \)-discounted version of \( \mathcal{M}_\lambda \), starting from \((s_+, \pi_-)\).

Following the same argument as in Proposition 1, for every initial distribution \( \pi = (\pi_+, \pi_-) \in \times_{i \in I} \Delta(S^i) \) and every Nash equilibrium of the game with payoff vector \( v \in \mathbb{R}^I \),
one has $\lambda \cdot v \leq \lambda \cdot E_{s+ \sim \pi_+}[v_\delta(s_+, \pi_-)]$. Hence the result will follow from the equality

$$\lim_{\delta \to 1} v_\delta(s_+, \pi_-) = \bar{k}_1(\lambda), \text{ for all } (s_+, \pi_-). \tag{17}$$

We will prove (17) by approximating $\mathcal{M}_\lambda$ with MDPs with a finite state space and using results from the theory of such MDPs. We introduce some piece of notation, to be used later as well. Given a finite subset $K^i$ of $\Delta(S^i)$, a map $\phi^i : K^i \times Y \to K^i$ and $\eta > 0$, the pair $(K^i, \phi^i)$ is an $\eta$-approximation of $\Delta(S^i)$ if

$$\|\phi^i(k^i, y) - p^i(\cdot \mid k^i, y)\|_\infty < \eta$$

for every $k^i \in K^i$ and $y \in Y$. Intuitively, (18) entails that the exact posterior on the next state given a prior $k^i$ on the current state $s^i$ and a signal $y$, is $\eta$-close to $\phi^i(k^i, y)$. That is, the map $\phi^i$ is a good approximation of the evolution of beliefs over states.\(^{43}\)

Given a family $(K^i, \phi^i)$ of $\eta$-approximations of $\Delta(S^i)$, $i \in I$, the pair $(K, \phi)$ defined as $K = \times_{i \in I} K^i$ and $\phi(k, y) = (\phi^i(k^i, y))_i$ is said to be an $\eta$-approximation of $\Delta_\prec$.\(^{44}\)

We will without further notice assume that all $\eta$-approximations below satisfy the following communication property: for any two $k, \bar{k} \in K$, there exists a integer $N \in \mathbb{N}$, action profiles $a_1, \ldots, a_N$, and signals $y_1, \ldots, y_N$, such that (i) $p(y_n \mid a_n) > 0$ for each $n$, and (ii) the sequence $(k_n)$ defined by $k_1 = k$ and $k_{n+1} = \phi(k_n, y_n)$ is such that $k_{N+1} = \bar{k}$.\(^{44}\) Given an $\eta$-approximation $(K, \phi)$ of $\Delta_\prec$, we define $\mathcal{M}_\phi$ to be the MDP with finite state space $S_+ \times K$, action set $A$, and transitions deduced from $p(\cdot \mid a) \in \Delta(Y)$ and $\phi$. Finally, the stage reward function is (the restriction of) $r_\lambda$. Thus, $\mathcal{M}_\phi$ differs from $\mathcal{M}_\lambda$ only through the transition function, and we think of $\mathcal{M}_\phi$ as a finite state approximation of $\mathcal{M}_\lambda$. The MDP $\mathcal{M}_\phi$ is communicating.\(^{45}\)

Let $\varepsilon > 0$ be arbitrary and let $M$ be an upper bound on $\|r\|$. Since the transition function $p(\cdot \mid s, a)$ is aperiodic and irreducible, there exists a constant $c \in (0, 1)$ such that for each $(a_s)_{s \in S}$, and any two distributions $\pi$ and $\tilde{\pi}$ in $\times_{i \in I} \Delta(S^i)$, one has

$$\|\sum_{s \in S} p(\cdot \mid s, a_s)(\pi_s - \tilde{\pi}_s)\|_\infty \leq c\|\pi - \tilde{\pi}\|_\infty.$$  

Pick $\eta < \varepsilon(1 - c)/M$, and an $\eta$-approximation $(K, \phi)$ of $\Delta_\prec$.

\(^{43}\)Note though that this interpretation is valid only if $y$ is uninformative about $s^i$. Note also that we do not require that $K^i$ be a “large” subset of $\Delta(S^i)$.

\(^{44}\)The existence of communicating $\eta$-approximations is easy to establish. Not all $\eta$-approximations are communicating.

\(^{45}\)Using the full-support assumption and the communicating property of $(K, \phi)$.
In both MDPs $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$, strategies map past public signals $(y_n)$ and past (and current) states $(s_{+,n})$ of players in $I_\perp$ into an action profile. We prove in Lemma 7 below that any strategy induces approximately the same payoff in $\mathcal{M}_\lambda$ and in $\mathcal{M}_\phi$. Given a strategy $\sigma$, we denote by $\gamma_\delta(\cdot, \sigma)$ and $\gamma_{\delta,\phi}(\cdot, \sigma)$ the payoff induced in $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$ respectively, as a function of the initial state.

**Lemma 7** For every discount factor $\delta < 1$, any $s_+ \in S_+$, $\pi_- \in \Delta_-$ and $k \in K$, one has

$$|\gamma_\delta((s_+, \pi_-), \sigma) - \gamma_{\delta,\phi}((s_+, k), \sigma)| \leq \varepsilon + \frac{M(1-\delta)}{1-\delta c}.$$  

**Proof.** Fix $\sigma$ and an arbitrary play $h_\infty = (s_{+,n}, y_n, a_n)_n$. Given a player $i \in I_-$ and a round $n \in \mathbb{N}$, let $\pi_n^i \in \Delta(S^i)$ and $k_n^i$ be the $i$-th component of the state in $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$ along $h_\infty$.\(^{46}\) Along $h_\infty$, the payoff difference in $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$ is

$$\left| (1-\delta) \sum_{n=1}^{\infty} (r_\lambda(s_{+,n}, \pi_{-,n}, a_n) - r_\lambda(s_{+,n}, k_n, a_n)) \right| \leq M(1-\delta) \sum_{n=1}^{\infty} \delta^{n-1} \|\pi_{-,n} - k_n\|_\infty.$$  

The two sequences obey the recursions $\pi_n^i = p^i(\cdot \mid \pi_{n-1}^i, y_{n-1})$ and $k_n^i = \phi^i(k_{n-1}^i, y_{n-1})$ so that, by the triangle inequality, one has $\|\pi_n^i - k_n^i\|_\infty \leq c\|\pi_{n-1}^i - k_{n-1}^i\|_\infty + \eta$. Routine computations then lead to $\|\pi_n^i - k_n^i\|_\infty \leq \frac{\eta}{1-c} + c^n$, hence the payoff difference along $h_\infty$ does not exceed $\frac{M\eta}{1-c} + \frac{M(1-\delta)}{1-\delta c}$. \(\blacksquare\)

Let $v_{\delta,\phi}$ be the value of the $\delta$-discounted version of $\mathcal{M}_\phi$. Since $\mathcal{M}_\phi$ has a finite state space, by Blackwell (1962), there is a (pure) policy $\rho_* : S_+ \times K \rightarrow A$ which is optimal for all $\delta$ close enough to one. That is, $\gamma_{\delta,\phi}(\rho_*) = v_{\delta,\phi}$ for $\delta$ large enough, hence $v_\phi := \lim_{\delta \to 1} v_{\delta,\phi}$ exists. Since $\mathcal{M}_\phi$ is communicating, the limit value $v_\phi$ is independent of the initial state.

**Claim 11** For all $(s_+, \pi_-)$, one has $|\lim_{\delta \to 1} v_\delta(s_+, \pi_-) - v_\phi| \leq \varepsilon$.

**Proof.** By Lemma 7, one has both

$$|\limsup_{\delta \to 1} v_\delta - v_\phi| \leq \varepsilon \text{ and } |\liminf_{\delta \to 1} v_\delta - v_\phi| \leq \varepsilon.$$  

Since $\varepsilon$ is arbitrary, this implies the convergence of $v_\delta$ as $\delta \to 1$, with $|\lim_{\delta \to 1} v_\delta - v_\phi| \leq \varepsilon$. \(\blacksquare\)

\(^{46}\)That is, $\pi_n^i$ is the conditional distribution of $s_n^i$ given past signals, while $k_n^i$ is obtained recursively through $\phi^i$. 

48
Claim 12 \( v_\phi \leq \bar{k}_1(\lambda) + \varepsilon \).

Proof. Plainly, the policy \( \rho_* \) may also be either viewed as a strategy in \( \mathcal{M}_\lambda \), or as a policy in the initial game with state space \( S \), independent of the states of players \( i \in I_- \). Under both “interpretations”, the payoff induced by \( \rho_* \) is of course equal to \( \gamma_\delta(\cdot, \rho_*) \). According to the first interpretation, Lemma 7 applies for each \( \delta \), and \( \limsup_{\delta \to 1} \| \gamma_\delta(\cdot, \rho_*) - \gamma_{\delta, \phi}(\cdot, \rho_*) \|_\infty \leq \varepsilon \). According to the second interpretation, \( \rho_* \) induces a Markov chain over \( S \times K \). Let \( E = S \times K_E \) be an arbitrary ergodic set\(^{47} \) for this Markov chain, with invariant measure \( \mu_E \in \Delta(S \times K_E \times A) \). Given an initial state \((\bar{s}, \bar{k}) \in E\), one has \( \lim_{\delta \to 1} \gamma_{\delta, \phi}(\bar{s}, \bar{k}, \rho_*) = v_\phi \) (by the choice of \( \rho_* \)), while \( \lim_{\delta \to 1} \gamma_{\delta}(\bar{s}, \bar{k}, \rho_*) = E_{\mu_E}[\lambda \cdot r(s, a)] \). Combining these results, one gets

\[
v_\phi \leq E_{\mu_E}[\lambda \cdot r(s, a)] + \varepsilon. \tag{19}
\]

To conclude, define \( \phi_E : K_E \times Y \to K_E \) by \( \phi_E(k, y) = \phi(k, y) \) whenever \( \phi(k, y) \in K_E \), and let \( \phi_E(k, y) \in K_E \) be arbitrary otherwise. The extended policy \((\rho, K_E, \phi_E)\) is irreducible, with invariant measure \( \mu_E \). Hence \( E_{\mu_E}[\lambda \cdot r(s, a)] \leq \bar{k}_1(\lambda) \).

Combining the last two claims, \( \lim_{\delta \to 1} v_\delta(s_+, \pi_-) \leq \bar{k}_1(\lambda) + 2\varepsilon \). Since \( \varepsilon > 0 \) is arbitrary, it follows that \( \lim_{\delta \to 1} v_\delta(s_+, \pi_-) \leq \bar{k}_1(\lambda) \).

The reverse inequality \( \bar{k}_1(\lambda) \leq \lim_{\delta \to 1} v_\delta \) is straightforward. Indeed, let \( \rho_{\text{ext}} = (\rho, K, \phi) \) be an arbitrary irreducible extended policy, where \( \rho : S \times K \to \Delta(A) \) is independent of \((s^i)_{j \notin I_+}\). The policy \( \rho \) induces a strategy in \( \mathcal{M}_\lambda \), hence \( \gamma_\delta(\cdot, \rho) \leq v_\delta(\cdot) \). Letting \( \delta \to 1 \), one obtains \( E_{\mu_{\rho_{\text{ext}}}}[\lambda \cdot r(s, a)] \leq \lim_{\delta \to 1} v_\delta \).

B.2 Proof of Proposition 4 and Theorem 4

B.2.1 An overview

To unify notations, we set \( \hat{V}_{i}^{**} = -\bar{v}^i \) for \( i \in I \), and \( \hat{k}_1(\lambda) = \bar{k}_1(\lambda) \) otherwise, so that \( V_{i}^{**} = \{ z \in \mathbb{R}^I, \lambda \cdot z \leq \hat{k}_1(\lambda) \text{ for all } \lambda \} \). We observe that \( \hat{k}(\cdot) \) is lower semi-continuous, and that \( \hat{k}(-e^i) \geq -\bar{v}^i \). Thus, \( \hat{k}(\cdot) \) is lower semi-continuous as well.

We will prove the following strengthening of Proposition 4.

Lemma 8 For every \( \lambda \in \Lambda \) and \( \varepsilon > 0 \), there exists a triple \((v, \rho_{\text{ext}}, x)\), which is feasible in \( \mathcal{P}_1(\lambda) \), with strict truth-telling incentives, and such that \( \lambda \cdot v > \hat{k}_1(\lambda) - \varepsilon \).

\(^{47}\)That all ergodic sets are product sets follows from the full support assumption.
Lemma 15 readily implies \( k_1(\lambda) \geq \tilde{k}_1(\lambda) \). The following subsections are devoted to the proof of Lemma 15.

In the meantime, we deduce Theorem 4 from Lemma 15. We let \( Z \) be a compact set included in the interior of \( V_1^{**} \). Since \( Z \) is compact, there exists \( \eta > 0 \) such that the \( \eta \)-neighborhood \( Z_\eta \) of \( Z \) is also included in the interior of \( V_1^{**} \). Thus, for all \( \lambda \in \Lambda \), there is \( \varepsilon > 0 \) such that \( \max_{z \in Z_\eta} \lambda \cdot z + \varepsilon < \tilde{k}_1(\lambda) \). Hence, by compactness of \( \Lambda \) and since \( \tilde{k}_1 \) is lower semi-continuous, there is \( \varepsilon_0 > 0 \) such that

\[
\forall \lambda \in \Lambda, \max_{z \in Z_\eta} \lambda \cdot z + 2\varepsilon_0 < \tilde{k}_1(\lambda).
\]  

(20)

**Lemma 9** There exists a finite set \( S \) of triples \((v, \rho_{ext}, x)\) such that the following holds. For every direction \( \lambda \in \Lambda \), there is an element \((v, \rho_{ext}, x)\) of \( S \) such that

1. \((v, \rho_{ext}, x)\) is feasible in \( P_1(\lambda) \) with strict truth-telling incentives.
2. \( \max_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0 < \lambda \cdot v \).

**Proof.** Given \( \lambda \in \Lambda \), apply Lemma 15 with \( \varepsilon = \varepsilon_0 \). Plainly, by adding a small constant to \( x \), we may assume that in addition \( \lambda \cdot x(\cdot) < 0 \) for fixed \( \lambda \), hence \((v, \rho_{ext}, x)\) is feasible in \( P_1(\lambda') \) for all \( \lambda' \) in a neighborhood of \( \lambda \). Using (20), the inequality \( \lambda \cdot v > \tilde{k}_1(\lambda) - \varepsilon_0 \) implies

\[
\lambda \cdot v > \sup_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0.
\]

Since both sides of the inequality are continuous in \( \lambda \) and since \( \Lambda \) is compact, Lemma 9, and therefore Theorem 4, follows from Lemma 15. \( \blacksquare \)

Lemma 9 is the exact analog of Lemma 4 (in the proof of Theorem 2). Inspection of the proof of Theorem 2 then shows that Lemma 3 is still valid here and that all subsequent arguments based on Lemmas 3 and 4 remain valid (that is, both the construction of strategies in Section 3 and the results from Section A.2.3 readily extend to the present, more general setup).\(^{48}\) Thus, Theorem 4 follows.

\(^{48}\)The only, quite minor modification is as follows. Elements of \( S \) are now triples \((\rho_{ext}, x, v)\), where \( \rho_{ext} = (\rho, K, \phi) \) is an extended policy, and the auxiliary set \( K \) changes with \( \rho \). At the beginning of the \( k \)-th block, once the extended policy \((\rho(k), K(k), \phi(k))\) has been selected as a function of past public play, an initial state in \( K(k) \) still needs to be specified. This choice is irrelevant for the proofs.
B.2.2 Step 1: There is a strictly truthful pair \((\rho_{ext,0}, x_0)\)

We start to prove Lemma 9. In this first step, we construct a specific \textit{ex post}, strictly truthful (pure) pair \((\rho_{ext,0}, x_0)\). It will later be used as a perturbation, and will thus play a role analog to that of the pair \((\rho_*, x_*)\) in Section 3.

By A1 and as in Section 3, there exists for each \(i \in I\) a family \(\mu^i(s^i) \in \Delta(A)\) of distributions, and transfers \(\tau^i : S^i \to \mathbb{R}\) such that, for each \(s^i\), the map \(s^i \mapsto \tau^i_\mathcal{E}(s^i) + E_{a \sim \mu(s^i)}[\phi_{a,r}(s^i)]\) has a strict maximum at \(\bar{s}^i = s^i\). We assume w.l.o.g. that \(\mu^i(s^i)\) has full support.

For \(s \in S\), define then \(\mu(s) := \frac{1}{|I|} \sum_{i \in I} \mu^i(s^i)\) and \(T^i(s) := \frac{1}{|I|} \tau^i(s^i)\) so that, for each \(i \in I\) and \(s \in S\), the map \(\bar{s}^i \mapsto T^i(\bar{s}^i, s^{-i}) + E_{a \sim \mu(s^i, s^{-i})}[\phi_{a,r}(s^i)]\) has a strict maximum at \(\bar{s}^i = s^i\).

Let \(\eta_0 > 0\) to be fixed later, and set \(K_0 = A\). Under the extended policy \(\rho_{ext,0} = (\rho_0, K_0, \phi_0)\), players repeat the same action profile \(a \in K_0\) until the p.r.d. picks at a random time a (possibly) different new action profile according to a distribution which is contingent on the states reported in that round.

Formally, given the recommendation \(a_0\) in the previous round, and reports \(m\) in the current round, the p.r.d. picks a recommended action profile \(a'_0 \in A\), which is equal to \(a_0\) with probability \(1 - \eta_0\), and drawn according to \(\mu(m)\) otherwise. We set \(\rho_0(m, a_0) = a'_0\), \(\phi_0(m, a_0, y) = a'_0\), and \(x_0(m, a_0) = -\gamma_{a'_0} + \eta_0 T(m)\).

Thus, \((\rho_0, K_0, \phi_0)\) is irreducible. Denote by \(\mu_{\eta_0}\) the invariant measure and by \(\gamma_{\eta_0} \in \mathbb{R}^I\) and \(\theta_{\eta_0} : S \times K_0 \to \mathbb{R}^I\) the long-run payoff and relative values respectively, including transfers.

**Lemma 10** Both \(\lim_{\eta_0 \to 0} \mu_{\eta_0}\) and \(\lim_{\eta_0 \to 0} \theta_{\eta_0}\) exist. In addition, the difference \(\lim_{\eta_0 \to 0} \theta_{\eta_0}(s, k_0) - \theta_{\eta_0,r}(s)\) only depends on \(k_0\).

**Proof.** The distribution \(\mu_{\eta_0}\) is the unique solution to a linear system with coefficients affine in \(\eta_0\). Therefore, \(\eta_0 \mapsto \mu_{\eta_0}\) is a rational function and, being bounded, has a limit as \(\eta_0 \to 0\). We refer to the online appendix for the proof relative to \(\theta_{\eta_0}\). The proof uses similar arguments, but the proof that \(\eta_0 \mapsto \theta_{\eta_0}\) is bounded is more delicate. \(\square\)

\(^{49}\)Consistent with our usage, the dependence of \(\rho_0, x_0\) and \(\phi_0\) on the outcome of the p.r.d. does not appear explicitly.

\(^{50}\)Recall that \(\gamma_{\bar{a}}\) is the long-run payoff induced by the constant policy \(\bar{a}\). That is, \(\gamma_{\bar{a}} = E_{a \sim \mu_\bar{a}}[r(s, a)] = r(\mu_{\bar{a}}, a)\).
Lemma 11 For $\eta_0$ small enough, the pair $(\rho_{ext,0}, x_0)$ is ex post strictly truthful.

Proof. We need to prove that in the one-shot Bayesian game $\Gamma(\rho_{ext,0}, x_0)$ and given a state profile $(s, a) \in S \times K_0$, each player $i$ finds it strictly optimal to report $s^i$ (assuming obedience to $\rho_{ext,0}$). Fix $(s, a) \in S \times K_0$. The actual payoff of player $i$ when reporting $\tilde{s}^i \in S^i$ is

$$r^i(s^i, a') + x^i_0((\tilde{s}^i, s^{-i}), a') + \theta^i_{\eta_0}(t, a'),$$

where $a' \in A$ is the p.r.d recommendation and $t \sim p(\cdot \mid s, a')$.

Taking expectations over $a'$ and $t$, and since $x^i_0(\tilde{s}^i, s^{-i}) = \eta_0 T(\tilde{s}^i, s^{-i}) - \gamma_{\tilde{a}^-}$, the expected payoff when reporting $\tilde{s}^i$ is

$$(1 - \eta_0) \left\{ r^i(s^i, a) - \gamma_{\tilde{a}} + E_{t \sim p(\cdot \mid s, a)} \left[ \theta^i_{\eta_0}(t, a) \right] \right\} + \eta_0 \left\{ E_{a' \sim \mu(\tilde{s}^i, s^{-i})} \left[ r^i(s^i, a') + T^i(\tilde{s}^i, s^{-i}) - \gamma_{\tilde{a}^-} + E_{t \sim p(\cdot \mid s, a') } \theta^i_{\eta_0}(t, a') \right] \right\}$$

The first term is independent of $\tilde{s}^i$. As for the second, observe that, for fixed $a'$, the term between brackets converges as $\eta_0 \to 0$ to

$$r^i(s^i, a') + T^i(\tilde{s}^i, s^{-i}) - \gamma_{\tilde{a}^-} + E_{t \sim p(\cdot \mid s, a') } \theta^i_{\tilde{a}^-}(t)$$

(up to an additive constant), which is equal to $T^i(\tilde{s}^i, s^{-i}) + \theta^i_{\tilde{a}^-}(s^i)$. By the choice of $\mu$ and $T$, the expectation of the latter term under $a' \sim \mu(\tilde{s}^i, s^{-i})$ has a strict maximum for $\tilde{s}^i = s^i$. Therefore, it is ex post strictly optimal for player $i$ to report truthfully.

B.2.3 Step 2: $\lambda$ is not a coordinate direction

We here deal with the more difficult case where $\lambda$ is not a coordinate direction. We rely on Proposition 5 below which deals with the following question. Let be given an irreducible MDP with state space $\Omega$, action set $B$, transition function $q(\cdot \mid \omega, b)$, reward $r : \Omega \times B \to \mathbb{R}$ (all sets being finite). Assume that successive states are observed by a first agent, who makes a report to a second one, who in turn chooses an action, the reward of both agents being $u$. Plainly, if the second agent follows a stationary optimal policy, it is weakly optimal for the first one to be truthful. According to Proposition 5, there are arbitrarily small report-contingent transfers and an optimal policy in the perturbed MDP, see P1, such that truth-telling is strictly optimal whenever the report affects the action (distribution) being played, see P2.
Proposition 5 For each $\varepsilon > 0$, there exists $x : \Omega \times B \to \mathbb{R}$ and $\rho : \Omega \to \text{int}\Delta(B)$ such that the following holds, with $\theta := \theta_{\rho, u+x}$:

**P1** $\|x(\cdot)\| < \varepsilon$ and $\rho$ is an optimal policy in the MPD with reward $r + x$.

**P2** For every $\omega, \tilde{\omega} \in \Omega$,

$$r(\omega, \rho(\omega)) + x(\omega, \rho(\omega)) + \mathbb{E}_{q(\cdot|\omega, \rho(\omega))} \theta(\omega') \geq r(\omega, \rho(\tilde{\omega})) + x(\tilde{\omega}, \rho(\tilde{\omega})) + \mathbb{E}_{q(\cdot|\omega, \rho(\tilde{\omega}))} \theta(\omega'),$$

and a strict inequality holds whenever $\rho(\tilde{\omega}) \neq \rho(\omega)$.

Proposition 5 is immediate when transitions are action-independent. In that case indeed, and for $\omega \in \Omega$, set $B(\omega) := \arg\max_B r(\omega, \cdot)$ and let $\rho(\omega)$ be the uniform distribution over $B(\omega)$ and set $x(\omega) := \eta|B(\omega)|$. For $\eta > 0$ small enough, the pair $(\rho, x)$ satisfies **P1** and **P2**.

The proof is significantly more involved under action-dependent transitions. It is provided in the online appendix.

We now proceed in three (sub-)steps. We first rely on Proposition 5 to prove in Lemma 12 the existence of an extended policy $\rho_{\text{ext},1}$ and of transfers $x_1$ such that the long-run payoff under $\rho_{\text{ext},1}$ is close to $\bar{k}_1(\lambda)$ and such that truth-telling incentives are ex post strict unless reports do not affect the action being played. By perturbing the latter extended policy with the policy $\rho_{\text{ext},0}$ defined in Step 1, we next prove in Lemma 13 the existence of an ex post strictly truthful pair $(\rho_{\text{ext},1}, x)$ such that the long-run payoff under $\rho_{\text{ext}}$, excluding transfers – is close to $\bar{k}_1(\lambda)$. We conclude using AS and the action-identifiability assumption A2.

Lemma 12 For all $\varepsilon > 0$, there exists an irreducible extended policy $\rho_{\text{ext},1} = (\rho_1, K_1, \phi_1)$ where $\rho_1 : S_+ \times K_1 \to \Delta(A)$ and transfers $x_1 : S_+ \times K_1 \to \mathbb{R}^I$, s.t.

**C1** $\mathbb{E}_{\mu_{\rho_{\text{ext},1}}} [\lambda \cdot r(s, a)] > \bar{k}_1(\lambda) - \varepsilon$;

**C2** For all $(s, k) \in S \times K$, all $i \in I_+$ and $\bar{s}^i \neq s^i$ such that $\rho(s_+, k) \neq \rho(\bar{s}^i, s_+^{-i}, k)$, player $i$ ex post strictly prefers reporting $s^i$ over $\bar{s}^i$ in $\Gamma(\rho_{\text{ext},1}, x_1)$ (at the state profile $(s, k)$).

**Proof.** Proposition 5 holds for finite MDPs, hence we will have to rely on finite state approximations of the MPD $\mathcal{M}_\lambda$. We use the notations from Section B.1. Let $\varepsilon > 0$ be given. We let $(K_1, \phi_1)$ be an $\eta$-approximation of $\Delta_+$ such that the limit value $v_\phi$ of the MDP $\mathcal{M}_\phi$ induced by $(K_1, \phi_1)$ is close to $\bar{k}_1(\lambda)$: $|v_\phi - \bar{k}_1(\lambda)| < \frac{\varepsilon}{3}$. In addition, we assume that $\eta > 0$ is small enough\(^51\) so that, for each irreducible $\rho : S_+ \times K_1 \to \Delta(A)$, one has

$$\left| \mathbb{E}_{(s_+, k, a) \sim \rho} [r_\lambda(s_+, k, a)] - \mathbb{E}_{(s, a) \sim \mu_\rho} [\lambda \cdot r(s, a)] \right| < \frac{\varepsilon}{3},$$

(21)

\(^51\)It suffices to take $\eta < (1 - c)\varepsilon/3M$, see Section B.1.
where $\mu_\rho \in \Delta(S \times K_1 \times A)$ is the invariant distribution induced by $\rho$. Inequality (21) reads as follows: the two expectations are the long-run payoffs induced by $\rho$ in the MDPs $\mathcal{M}_\phi$ and $\mathcal{M}_\lambda$ respectively. Consequently, the long-run $\lambda$-weighted payoff induced by any such policy $\rho$ is close to the payoff induced in $\mathcal{M}_\phi$.

With this choice of $(K_1, \phi_1)$, we apply Proposition 5 to the MDP $\mathcal{M}_\phi$ with $\varepsilon/3$, and get $\rho_1 : S_+ \times K_1 \to \text{int} \Delta(A)$ and $\bar{x} : S_+ \times K \times A \to \mathbf{R}$. Abusing notations, we will also view $\rho_1$ and $\bar{x}$ as maps defined on $S \times K_1$ and $S \times K_1 \times A$, independent of $s^i$ for $i \in I_-$.

To repeat, the pair $(\rho_{\text{ext},1}, x)$ is such that for all $\omega, \tilde{\omega} \in S_+ \times A$, and denoting by $q$ the transition function in $\mathcal{M}_\phi$, one has

$$r_\lambda(\omega, \rho_1(\omega)) + \bar{x}(\omega, \rho_1(\omega)) + E_q(\omega, \rho_1(\omega)) \theta_{\rho_1, r_\lambda + \bar{x}}(\omega') > r_\lambda(\omega, \tilde{\rho}_1(\omega)) + \bar{x}(\omega, \tilde{\rho}_1(\omega)) + E_q(\omega, \tilde{\rho}_1(\omega)) \theta_{\tilde{\rho}_1, r_\lambda + \bar{x}}(\omega')$$

whenever $\rho_1(\omega) \neq \tilde{\rho}_1(\omega)$ and

$$E_{\mu_{\rho_1}}[r_\lambda(\omega, \rho_1(\omega))] \geq v_\phi - \frac{\varepsilon}{3} \geq \bar{k}_1(\lambda) - \frac{2\varepsilon}{3}.$$  

Together with (21), this proves C1.

Next, we follow Claim 9 and introduce transfers of the VCG type. For $i \in I_+$, we define $x^i_1 : S_+ \times K_1 \to \mathbf{R}$ by

$$\lambda^i x^i_1(\omega) := r_\lambda(\omega, \rho_1(\omega)) - \lambda^i r^i(\omega, \rho_1(\omega)) + \lambda^i \bar{x}^i(\omega, \rho_1(\omega))$$

so that, as in Claim 9, one has $\lambda^i \theta_{\rho_1, r + x_1} = \theta_{\rho_1, \lambda + \bar{x}}$. Therefore, $\rho_1$ inherits the following truth-telling property in the one-shot game $\Gamma(\rho_{\text{ext},1}, x_1)$: at each state $(s^+, k) \in S_+ \times K_1$ and for each $i \in I_+$, player $i$ strictly prefers reporting $s^i$ over $\tilde{s}^i$ whenever $\rho_1(s^+, k) \neq \rho_1(\tilde{s}^i, s_{-i}^i, k)$.

For $i \in I_-$, set $x^i_1 = 0$. Since $\rho_1$ is independent of $s^i \in S^i$, the latter property also holds for all $i \in I_-$. Thus, C2 holds.

**Lemma 13** For all $\varepsilon > 0$, there exists an irreducible extended policy $\rho_{\text{ext}} = (\rho, K, \phi)$ such that

C'1 $\lambda \cdot E_{\rho_{\text{ext}}}[\lambda \cdot r(s, a)] \geq \bar{k}_1(\lambda) - \varepsilon$.

C'2 The pair $(\rho_{\text{ext}}, x)$ is ex post strictly truthful.

**Proof.** We construct $(\rho, K, \phi)$ as a perturbation of $(\rho_1, K_1, \phi_1)$ using $(\rho_0, K_0, \phi_0)$ so that the play alternates between long phases in $S \times K_1$ following and long but much shorter, phases in $S \times K_0$.  

54
Let $\eta_1 > 0$ be small, to be fixed later. We define the extended policy $\rho_{ext} = (\rho, K, \phi)$ as follows. We set $K = K_0 \cup K_1$. In each round, given the current public auxiliary state $k \in K$ and reports $s \in S$, the p.r.d. updates the public state to $k' \in K$ as follows. If $k \in K_1$, $k'$ is set to $k$ with probability $1 - \eta_1^2$, and $k' \sim \mu(s)$ otherwise. If $k \in K_0$, $k'$ is set to a fixed $\bar{k}_1 \in K_1$ with probability $\eta_1^2$, and otherwise determined as under $\rho_{ext,0}$ (i.e., set to $k$ with probability $1 - \eta_0$, and otherwise drawn according to $\mu(s)$).

We then define $\rho : S \times K \to \Delta(A)$ as $\rho(s,k) = \rho_0(s,k')$ if $k' \in K_0$ and $\rho(s,k) = \rho_1(s,k')$ if $k' \in K_1$. We also set $\phi(s,k,y) = \phi_1(k',y)$ if $k' \in K_1$ and $\phi(k,y) = k'$ if $k' \in K_0$.

Transfers $x : S \times K \to \mathbf{R}^I$ are defined as $x(s,k) = x_1(s,k')$ if $k' \in K_1$, $x(s,k) = x_0(s,k') = \eta_0 T(s) - \gamma \tilde{k}_1$ if both $k, k' \in K_0$ and $x(s,k) = T(s) - \gamma \tilde{k}_1$ if $k \in K_1$ and $k' \in K_0$.

The irreducibility of $(\rho, K, \phi)$ follows from that of both $(\rho_0, K_0, \phi_0)$ and $(\rho_1, K_1, \phi_1)$. We denote by $\mu_{\eta_1} := \mu_{\rho_{ext}}$ the invariant measure as a function of $\eta_1$, and by $\theta_{\eta_1} := \theta_{\rho_{ext,r+x}} : S \times K \to \mathbf{R}^I$ the relative values. As in Section B.2.2, $\mu := \lim_{\eta_1 \to 0} \mu_{\eta_1}$ is well-defined. In addition, note that the limit transition function is the one induced by $\rho_{ext,0}$ and $\rho_{ext,1}$ on $S \times K_0$ and $S \times K_1$ respectively. Since transitions from $K_0$ to $K_1$ (resp. from $K_1$ to $K_0$) occur with probability $\eta_1$ (resp., $\eta_1^2$), one has $\mu_{\eta_1}(S \times K_0) = \frac{1}{1 + \eta_1}$. As a consequence,

$$\lim_{\eta_1 \to 0} E_{\mu_{\eta_1}}[\lambda \cdot r(s,a)] = E_{\mu_{\rho_{ext,1}}} [\lambda \cdot r(s,a)] > \bar{k}_1(\lambda) - \varepsilon,$$

hence $C'$ hold for $\eta_1$ small enough.

We turn to $C''$. As in Section B.2.2 (see proof on line), $\theta := \lim_{\eta_1 \to 0} \theta_{\eta_1}$ is also well-defined, and for $s \in S$, the differences $\theta(s,k_0) - \theta_0(s,k_0)$ and $\theta(s,k_1) - \theta_1(s,k_1)$ are independent of $k_0 \in K_0$ and $k_1 \in K_1$ respectively (where $\theta_n = \theta_{\rho_{n,r+x_n}}$ for $n = 0,1$).

Fix $(s,k) \in S \times K$, $i \in I$ and $\bar{s}^i \in S^i$. If $k \in K_0$ the strict incentive to report $s^i$ follows from the strict truthfulness of $(\rho_{ext,0}, x_0)$, for $\eta_1$ small enough.

Assume now that $k \in K_1$. If $\rho(s^i, s^{-i}, k) \neq \rho(\bar{s}^i, s^{-i}, k)$ for some $s^{-i} \in S^{-i}$, player $i$ strictly prefers reporting $s^i$ over $\bar{s}^i$ in $\Gamma(\rho_{ext,1}, x_1)$. And therefore in $\Gamma(\rho_{ext}, x)$ as well, for $\eta_1$ small enough. Assume finally that $\rho(s^i, s^{-i}, k) = \rho(\bar{s}^i, s^{-i}, k)$ for all $s^{-i} \in S^{-i}$. Then the expected payoff of player $i$ in $\Gamma(\rho_{ext}, x)$, conditional on the p.r.d. picking $k' = k$ is the same under both reports $s^i$ and $\bar{s}^i$. On the other hand, conditional on the p.r.d. picking some $a' \in K_0$, the expected payoff of player $i$ converges as $\eta_1 \to 0$ to

$$E_{a' \sim \mu(s^{-i})} [r^i(s^i, a') + T^i(*, s^{-i}) - \gamma \alpha_{a'} + E_{t \sim p(\cdot | s, a')} \hat{\theta}(t, a')]$$

---

52 Whatever be the choice of $\omega_{pub}$ by nature in $\Gamma(\rho_{ext,1}, x_1)$. Indeed, player $i$ is ex post indifferent between $s^i$ and $\bar{s}^i$ when $p(s) = \rho(\bar{s}^i, s^{-i})$ and strictly prefers $s^i$ over $\bar{s}^i$ if $p(s) = \rho(s^i, s^{-i})$. The claim thus follows since the belief of player $i$ over $S^{-i}$ has full support.
which, by the choice of $\eta_0$, has a strict maximum for $s^i$. The strict truthfulness of $(\rho_{\text{ext}}, x)$ follows, provided $\eta_1$ is small enough. ■

We now conclude the proof of Lemma 15. Inspection of the proofs of Claims 8, 9 and 10 shows that the successive modifications of the transfers in AS preserve strict inequalities and do not rely on transitions being action-independent. That is, the same sequence of claims leads here to the existence of $x_4 : \Omega_{\text{pub}} \times S \times K \to \mathbb{R}^I$ (with $\Omega_{\text{pub}} = S \times K \times Y$) such that $\lambda \cdot x_4(\cdot) = 0$ and all truth-telling incentives in $\Gamma(\rho_{\text{ext}}, x_4)$ are strict.

We finally add a component to transfers, so as to ensure obedience. This is standard. By A2, and since $\lambda$ is not a coordinate direction, there exists for each $a \in A$ transfers $x_a : Y \to \mathbb{R}^I$ such that $\lambda \cdot x_a(\cdot) = 0$, $\mathbf{E}_{y \sim p(\cdot | a)} x_a(y) = 0$ and $\mathbb{E}_{y \sim p(\cdot | \tilde{a}_i, a_i)} x_i^a(y)$ is a large negative constant for each $i \in I$ and $\tilde{a}_i \neq a_i$. We then view the policy $\rho : S \times K \to \Delta(A)$ as being implemented by means of the p.r.d. that picks a recommended action profile $a \in A$ based on the reports, leading to transfers $x_a(\cdot)$. We abbreviate this to $x_\rho : S \times K \times Y \to \mathbb{R}^I$ and finally set $x := x_4 + x_\rho$. Since $\lambda \cdot x_\rho(\cdot) = 0$, the expected weighted payoff induced by $\rho_{\text{ext}}$ in $\Gamma(\rho_{\text{ext}}, x)$ is

$$\mathbf{E}_{\mu_{\rho_{\text{ext}}}} [\lambda \cdot r(s, a)] \geq \bar{k}_1(\lambda) - \varepsilon,$$

and the triple $(\mathbf{E}_{\mu_{\rho_{\text{ext}}}} [r(s, a)], \rho_{\text{ext}}, x)$ is feasible in $\mathcal{P}_1(\lambda)$.

**B.2.4 Step 3: $\lambda$ is a coordinate direction**

We continue with the case $\lambda = +e^i$. The proof involves a variation upon the ideas of Section B.2.3 but is much simpler. We denote by $\mathcal{M}^i$ the MDP faced by the players when jointly maximizing the payoff of player $i$. The MDP $\mathcal{M}^i$ has $S^i$ as state space, $A$ as action set, and the reward and transitions are $r^i$ and $p$. Plainly, the limit value of $\mathcal{M}^i$ is $\bar{k}_1(e^i)$.

We let an arbitrary $\varepsilon > 0$ be given, and let $\bar{x} : S^i \times A \to \mathbb{R}$ and $\rho_1 : S^i \to \Delta(A)$ be obtained by applying Proposition 5 to $\mathcal{M}^i$. We will obtain strict truth-telling incentives by means of a perturbation argument. Before doing so, we first modify $\bar{x}$ to get strict obedience incentives.

We view $\rho_1 : S \to \Delta(A)$ as a map defined over $S$ (independent of $s^{-i}$) and recall that $\theta_{\rho_1, r + \bar{x}} : S \to \mathbb{R}^I$ are the relative values associated with $\rho_1$ and $\bar{x}$. By A2, for $j \neq i$ and for each $a \in A$, there exists $x_a : Y \to \mathbb{R}^I$ that induce strict obedience to $a$:

$$r^j(s^j, \tilde{a}^j, a_{-j}) + \mathbf{E}_{p(\cdot | \tilde{a}^j, a_{-j})} x^j_a(y) + \mathbf{E}_{p(\cdot | s^j, \tilde{a}^j, a_{-j})} \theta_{\rho_1, r + \bar{x}}^j(s^j, a) + \mathbf{E}_{p(\cdot | s^j, a_{-j})} \theta_{\rho_1, r + \bar{x}}^j(t) < r^j(s^j, a) + \mathbf{E}_{p(\cdot | a)} x^j_a(y) + \mathbf{E}_{p(\cdot | s^j, a)} \theta_{\rho_1, r + \bar{x}}^j(t)$$

for each $s \in S$ and $\tilde{a}^j \neq a^j$.
For \( j = i \), we ask for more. For any \( s^i \in S^i \) and since \( \rho_1 \) is optimal in the MDP with payoff \( r^i + \bar{x} \), any action \( a \in A \) in the support of \( \rho_1(s^i) \) maximizes
\[
r^i(s^i, a) + \bar{x}(s^i, a) + E_{t \sim p(\cdot | s^i, a)} \theta_{\rho_1, r+\bar{x}}(t^i).
\]
Since \( \|\bar{x}\| < \varepsilon \), the components \( x^i_a \) can be chosen so that the following holds:

**B1**: \((22)\) is modified and strengthened to
\[
r^i(s^i, a) + \bar{x}(s^i, a) + E_{y \sim p(\cdot | \tilde{a}^i, a^{-i})} x^i_a(y) + E_{t \sim p(\cdot | s, \tilde{a}^i, a^{-i})} \theta_{\rho_1, r+\bar{x}}(t) < r^i(s^i, a) + \bar{x}(s^i, a) + E_{y \sim p(\cdot | a)} x^i_a(y) + E_{t \sim p(\cdot | s^i, a)} \theta_{\rho_1, r+\bar{x}}(t^i)
\]
for every \( \tilde{s}^i \in S^i \) and \( \tilde{a}^i \neq a^i \).

**B2** \( \|x^i_a\| < k\varepsilon \) for some constant \( k \) that only depends on the primitives of the model and not on \( \varepsilon \).

**B3** \( x^i_a(\cdot) \leq 0 \) and \( E_{y \sim p(\cdot)} x^i_a(y) \) is independent of \( a \in A \).

The substantive properties are **B1** and **B2**. Once they hold, **B3** follows by subtracting a small constant from \( x^i_a \).

We view \( \rho_1 : S \to \Delta(A) \) as being implemented by means of the p.r.d picking a recommendation \( a \sim \rho_1(s) \), and transfers being then given by \( x^i_1(s, y) := \bar{x}(s^i, a) + x^i_a(y) \) and \( x^j_1(s, y) = x^j_a(y) \) for \( j \neq i \). The properties of \( x^i_a \) and of \( \bar{x} \) ensure that the pair \((\rho_1, \bar{x})\) is strictly obedient and satisfy the same truth-telling incentives as the pair \((\rho_1, \bar{x})\). Observe that \( x^i_1(s, y) \leq (k+1)\varepsilon \). Since \( \rho_1 \) is optimal in the MDP with reward \( r + \bar{x} \), one has
\[
E_{p_{\text{ext}, 1}} \left[ r^i(s^i, a) + x^i_1(s, y) \right] \geq \bar{k}_1(e^i) − (k+1)\varepsilon.
\]

We now recall the strictly truthful pair \((\rho_{\text{ext}, 0}, x_0)\) with \( \rho_{\text{ext}, 0} = (\rho_0, K_0, \phi_0) \), from Step 1 in Section B.2.2. Once again, we supplement \( x_0 \) with transfers inducing obedience. For \( a \in A \), we let \( x_a : Y \to \mathbb{R}^J \) be such that (i) \( E_{p(\cdot | a)} x_a(y) = 0 \), and (ii) \( E_{p(\cdot | \tilde{a}^i, a^{-i})} x_a^i(y) \) is a large negative constant for each \( j \in J \) and \( \tilde{a}^i \neq a^i \). We next subtract the same constant to all maps \( x_a^i(\cdot) \) \( (a \in A) \) to get \( x_a^i(\cdot) \leq 0 \). (With an abuse of notation), transfers \( x_0 = S \times K_0 \times Y \to \mathbb{R}^J \) are now defined by \( x_0(s, k_0, y) = x_0(s, k_0) + x_a(y) \) where \( a \in A \) is selected by the p.r.d. as specified in \((K_0, \phi_0)\).\(^{53}\) With this updated definition of \( x_0 \), the pair \((\rho_{\text{ext}, 0}, x_0)\) is both strictly truthful and strictly obedient.

\(^{53}\)Recall from Section B.2.2 that the p.r.d. sets \( a \in A \) to be equal to \( k_0 \) with probability \( 1 - \eta_0 \) and otherwise draws \( a \sim \mu(s) \).
We now perturb. For $\eta > 0$, we define the irreducible extended policy $\rho_{ext} = (\rho, K, \phi)$ from $\rho_1$ and $\rho_{ext, 0}$ and transfers $x : S \times K \times Y \to \mathbf{R}^I$ from $x_0$ and $x_1$, exactly as $\rho_{ext}$ and $x$ were obtained in Step 2 from $\rho_{ext, 1}$ and $\rho_{ext, 0}$. As in Step 2, it follows that for $\eta > 0$ small, the pair $(\rho_{ext}, x)$ is both strictly truthful and obedient - hence the triple $(E_{\mu_{\rho_{ext}}}[r(s, a) + x(s, k, y)], \rho_{ext}, x)$ is feasible in $\mathcal{P}(e^i)$. Finally, since transitions from $\rho_1$ to $\rho_{ext, 0}$ (resp. from $\rho_{ext, 0}$ to $\rho_1$) occur with probability $\eta^2$ (resp. $\eta$) in each round, the expectation $E_{\mu_{\rho_{ext}}}[r^i(s, a) + x^i(s, k, y)]$ is arbitrarily close to $E_{\mu_{\rho_{ext}, 1}}[r^i(s, a) + x^i_1(s, a)]$ for $\eta > 0$ small enough. The result follows.

The case $\lambda = -e^i$ is analogous. Let $\bar{a}^{-i} \in A^{-i}$ achieve the min in the definition of $\underline{v}^i$. Let next $\tilde{M}^i$ be the MDP faced by player $i$ when maximizing his own payoffs against the constant policy $\bar{a}^{-i}$. Hence $\tilde{M}^i$ has $S^i$ as state space, $A^i$ as action set, and the rewards and transitions in $\tilde{M}^i$ are deduced from $r^i$ and $p$ given $\bar{a}^{-i}$. We then repeat the proof of the case $\lambda = -e^i$.

C Proof of Theorem 5

The proof of Theorem 5 consists in adding a layer of complexity to the proof of Theorem ?? to deal with negative, unit directions. We will extensively refer to the latter to avoid duplications. We will work under the assumption that the distribution of signals is independent of the current states. The proof for the general case is more cumbersome, but does not involve additional insights.

C.1 Alternative Scores

We first define modified scores $k_2(\lambda)$ and the corresponding set $\mathcal{H}_2$. We next observe that the IPV assumption, together with Assumption 1', ensures $\mathcal{H}_2 = W$.

Fix an arbitrary $s_* \in S$. We define a class of finite-horizon games, parameterized by final payoffs. Given a horizon $T \in \mathbf{N}$, final transfers $x : \Omega_{pub}^T \to \mathbf{R}^I$, and $\theta : \Omega_{pub} \times S \to \mathbf{R}^I$, we define $G(T, x, \theta)$ as the $T$-round repetition of the underlying stage game with communication, starting from the commonly known state profile $s_*$. The game $G(T, x, \theta)$ ends with the draw of $s_{T+1}$ in round $T + 1$.

\footnote{With obvious changes. Transfers $x^i$ to player $i$ are now required to be non-negative.}
Payoffs in $G(T, x, \theta)$ are given by
\[
\frac{1}{T} \left( \sum_{n=1}^{T} r(s_n, a_n) + x(h_{\text{pub}, T+1}) + \theta(\omega_{\text{pub}, T}, s_{T+1}) \right),
\]
where $h_{\text{pub}, T+1}$ is the public history in the $T$ rounds. Information and play is as in the infinite horizon game.

Denote by $C$ a uniform bound on $\theta_{\rho,r}$, when $\rho$ ranges through the set of all policies. For $\lambda \in \Lambda$ and $T \in \mathbb{N}$, we define the maximization problem $\tilde{P}_T(\lambda) : \tilde{k}_T(\lambda) := \sup \lambda \cdot v$, where the supremum is taken over all $(\sigma, x, \theta)$, such that

- $\sigma$ is a sequential equilibrium of $G(T, x, \theta)$ with payoff $v$.
- $\lambda \cdot x(\cdot) \leq 0$ and $\lambda \cdot \theta(\cdot) \leq C$.

Set $k_2(\lambda) = \limsup T \tilde{k}_T(\lambda)$, and $\mathcal{H}_2 := \{ v \in \mathbb{R}^I : \lambda \cdot v \leq k_2(\lambda) \text{ for all } \lambda \in \Lambda \}$.

**Proposition 6** One has $\mathcal{H}_2 = W$.

Proposition 6 follows from Lemmas 14 and 15 below.55

**Lemma 14** For $\lambda \neq -e^i$, one has $k_2(\lambda) = \bar{k}(\lambda)$.

**Proof.** Fix $\lambda \in \Lambda$ with $\lambda \neq -e^i$ for all $i \in I$. Let a weakly truthful pair $(\rho, x)$ be given with $\lambda \cdot x(\cdot) = 0$, set $v := E_{\mu}[r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})]$, and $\theta := \theta_{\rho,r+x}$. Given an integer $T \in \mathbb{N}$, define $x_T : \Omega^T_{\text{pub}} \to \mathbb{R}^I$ as
\[
x_T(h_{\text{pub}, T}) = \sum_{n=1}^{T} x(\omega_{\text{pub}, n-1}, \omega_{\text{pub}, n}),
\]
where $\omega_{\text{pub}, 0} \in \Omega_{\text{pub}}$ is arbitrary and $\omega_{\text{pub}, 1} = (s_*, y_1)$. Let $\sigma_T$ be the strategy profile in $G(T, x_T, \theta)$ defined as: (i) each player $i$ reports truthfully $m^i_n = s^i_n$ in all rounds, irrespective of past play, (ii) in each round $n$, player $i$ plays $\rho^i(m_n)$ if $m^i_n = s^i_n$, and any action $a^i$ which maximizes the expectation of
\[
r^i(s^i_n, \rho^{-i}(m_n), a^i) + x^i(\omega_{\text{pub}, n-1}, \omega_{\text{pub}, n}) + \theta^i(\omega_{\text{pub}, n}, s_{n+1})
\]
otherwise. Denote by $\tilde{\gamma}_T(\sigma_T)$ the expected payoff of $\sigma_T$ in $G(T, x, \theta)$.

55Only the inclusion $\mathcal{H}_2 \supseteq W$ is relevant for the proof.
Since \((\rho, x)\) is weakly truthful, it is easily checked that \(\sigma_T\) is a sequential equilibrium in \(G(T, x, \theta)\), hence \(\lambda \cdot \tilde{\gamma}_T(\sigma_T) \leq \tilde{k}_T(\lambda)\). On the other hand, by the irreducibility assumption, one has \(\lim_{T \to +\infty} \tilde{\gamma}_T(\sigma_T) = E_{\mu_\theta} [r(s, a) + x(\tilde{\omega}_{\text{pub}}, \omega_{\text{pub}})] = v\), so that \(\lambda \cdot v \leq k_2(\lambda)\). Using Lemma 2, this shows that \(\tilde{k}(\lambda) \leq k_2(\lambda)\), as desired.

We next prove that \(k_2(\lambda) \leq \tilde{k}(\lambda)\). Fix \(\varepsilon > 0\). Given \(T \in \mathbb{N}\), pick a feasible triple \((\sigma, x, \theta)\) in \(\tilde{\mathcal{P}}_T(\lambda)\) which achieves \(k_2(\lambda)\) up to \(\varepsilon\). Mimicking the argument in Lemma 2, there is a profile \(\tilde{\sigma}_T\) which only depends on the states of players in \(I(\lambda)\) and such that \(\lambda \cdot \tilde{\gamma}_T(\sigma_T) \leq \lambda \cdot \tilde{\gamma}_T(\tilde{\sigma}_T)\). Since \(\lambda \in \Lambda\), \(\lambda \cdot \theta(\cdot) \leq C\) and \(\lambda \cdot x(\cdot) \leq 0\), one has

\[
\lambda \cdot \tilde{\gamma}_T(\tilde{\sigma}_T) \leq \lambda \cdot \gamma_T(\sigma_T) + \frac{C}{T},
\]

where \(\gamma_T(\sigma_T)\) is the payoff induced in the \(T\)-round game \(G(T, 0, 0)\) with no final payoffs. Denote by \(v_T(\lambda) := \sup_\sigma \lambda \cdot \gamma_T(\sigma)\) the value of the \(\lambda\)-weighted \(T\)-round game, where the supremum is taken over \(\sigma : \times_{i \in I(\lambda)} S_i^I \to A\). By the irreducibility assumption, \(\lim_{T \to +\infty} v_T(\lambda) = \tilde{k}(\lambda)\). Let now \(T \to +\infty\) in (23) to get \(k_2(\lambda) - \varepsilon \leq \tilde{k}(\lambda)\). The result follows.

**Lemma 15** For \(\lambda = -e^i, k_2(\lambda) = -w_i^i\).

**Proof.** In all games \(G(T, x, \theta)\) considered for this lemma, the reports will be “babbling,” that is, a player sends the same report independently of his type in any given period. We set \(\theta = 0\). Given \(k \in \mathbb{N}\), let \(A^j_k := \{\alpha^j \in \Delta(A^j) : k\alpha^j(a_i^j) \in \mathbb{N} \text{ for all } a_i^j \in A^j\}\). The set \(A^j_k\) consists of those mixed action profiles that assign rational probabilities with denominator \(k\). For any \(\alpha^j \in \Delta(A^j)\), there exists \(\alpha^j_k \in A^j_k\) such that \(d(\alpha^j_k, \alpha^j) \leq |A^j|/k\); similarly, for all \(\alpha^{-i} \in \times_{j \neq i}\Delta(A^j)\), \(d(\alpha^{-i}_k, \alpha^{-i}) \leq |A^{-i}|/k\), for some \(\alpha^{-i}_k \in A^{-i}_k := \times_{j \neq i} A^j_k\). 57 We write \(\Sigma^j_k\) for the strategies of \(j\) with values in \(A^j_k\), and we let

\[
\sigma_k = \arg \min_{\sigma^{-i} \in \Sigma^{-i}_k} \max_{\sigma^i} \limsup_T \frac{1}{T} E_\sigma \left[ \sum_{n=1}^T g^i(s_n^i, a_n^i, y_n) \right]
\]

be minmax strategies when players \(-i\) are constrained to strategies in \(\Sigma^{-i}_k\). Given the product structure, these strategies may be taken measurable with respect to the history of signals of player \(i\), and we write \(h^i_{\text{pub}, n} \in H_{\text{pub}, n} = (Y^i)^{n-1}\) for such public histories. We write \(w_i^i(k)\) for the limiting expected payoff of player \(i\) under \(\sigma_k\). Using the irreducibility assumption, it follows from standard arguments that \(\lim_{k \to +\infty} w_i^i(k) = w_i^i\).

---

56When supplemented with appropriate beliefs.
57Throughout the lemma, we use the Euclidean distance.
58And that min and max are indeed achieved, so that \(\sigma_k\) is well-defined.
Given $T \in \mathbb{N}$, we also write $w^i(k, T)$ for the highest expected payoff of player $i$ over the first $T$ rounds, when facing $\sigma_{-i}^k$. Given a realized public history $h_t^{i, \text{pub}, T+1} \in H_t^{i, \text{pub}, T+1}$ and $\alpha_{k}^{-i} \in A_{k}^{-i}$, we let

$$T(\alpha_{k}^{-i}) = \{n = 1, \ldots, T: \sigma_{k}^{-i}(h_t^{i, \text{pub}, n}) = \alpha_{k}^{-i}\}$$

denote the rounds at which $\sigma_{k}^{-i}$ prescribes $\alpha_{k}^{-i}$.\(^{59}\) Let $\alpha_{k}^{-i} \in A_{k}^{-i}$ be the empirical distribution of signals observed in those stages. For $j \neq i$ and $y^j \in Y^j$, we also denote by $f[\alpha_{k}^{-i}](y^j)$ the empirical frequency of $y^j$ over the stages in $T(\alpha_{k}^{-i})$. We now let

$$D^j(h_{\text{pub}, T+1}) = \sum_{\alpha_{k}^{-i} \in A_{k}^{-i}} \frac{|T(\alpha_{k}^{-i})|}{T} \sum_y \left| f[\alpha_{k}^{-i}](y) - f[\alpha_{k}^{-i}](y^j) \right| \mathbb{P}[y^j | \alpha_{k}^j],$$

and, given $\phi > 0$, we define the test:

$$\tau_{\phi}^j(h_{\text{pub}, T+1}) = \begin{cases} 1 & \text{if } D^j(h_{\text{pub}, T+1}) < \phi, \\ 0 & \text{otherwise.} \end{cases}$$

We can finally state one claim that directly parallels one of Gossner (1995).

**Claim 13** Given $\varepsilon > 0$ and $\phi > 0$, there exists $T_0$ such that, if $T \geq T_0$,

$$\mathbb{P}_{\sigma_{j}^i, \sigma_{-j}^i} \left[ \tau_{\phi}^j(h_T) = 0 \right] < \varepsilon,$$

for all $j \neq i$ and all strategy profiles $\sigma^{-j}$.

In words, if player $j$ uses $\sigma_{j}^i$, he is very likely to pass the test $\tau_{\phi}^j$ no matter players $-j$’s strategy profile. The proof of Claim 13 relies on approachability theory, see Gossner (1995) for details.

Given $\varepsilon > 0$, we let $\phi < \frac{2\varepsilon}{\bar{r}(I - 1)}$, and let $T_0$ be given by Claim 13 applied with $\varepsilon/2\bar{r}$ and $\phi$.\(^{60}\) Given $T \geq T_0$, we pick $M > 0$ such that

$$-T\bar{r} - \varepsilon M > T\bar{r} - 2\varepsilon M,$$

or equivalently, $M > T\frac{\bar{r}}{\varepsilon}$. That is, $M$ is a punishment sufficiently large (for failing the test) that getting the worst reward for $T$ rounds followed by a probability of failing the test of up

---

\(^{59}\)Here, $h^{i, \text{pub}, n}$ refers to an initial segment of $h^{i, \text{pub}, T+1}$.

\(^{60}\)Here, $\bar{r}$ is a uniform bound on all payoffs in the game.
to $\varepsilon$ exceeds the payoff from the highest reward for $T$ rounds followed by a probability of failing the test of at least $2\varepsilon$.

We next set $x^i(\cdot) = 0$ and, for $j \neq i$,

$$x^j(h_{pub,T+1}) = \begin{cases} -M & \text{if } \tau^j_\phi(h_{pub,T+1}) = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (24)$$

The second claim states that, if all players $j \neq i$ pass the test with high probability, player $i$ is effectively punished.

**Claim 14** For every sequential equilibrium $\sigma$ of $G(T, x, \theta)$, one has

$$\frac{1}{T}E_{x,\sigma} \left[ \sum_{n=1}^{T} g^i(s^i_n, a^i_n, y_n) \right] \leq w^i(k, T) + 2\varepsilon.$$

**Proof.** By the condition on $M$, one has $P_\sigma [\tau^j_\phi(h_T) = 0] < 2\varepsilon_1$ in all equilibria of $G(T, x, \theta)$. Take any strategy profile in $G(T, x, \theta)$ such that $P_\sigma [\tau^j_\phi(h_T) = 0] < \frac{\varepsilon}{2F}$ for all $j \neq i$. On the event $\cap_{j \neq i} \{\tau^j_\phi(h_T) = 1\}$, one has for all $j \neq i$,

$$\sum_{y} \sum_{\alpha_k^{-i}} \frac{|T(\alpha_k^{-i})|}{T} \left| f[\alpha_k^{-i}](y) - f[\alpha_k^{-i}](y^j) P[y^j | \alpha_k^j] \right| < \phi,$$

which implies, by repeated substitution,

$$\sum_{y} \sum_{\alpha_k^{-i}} \frac{|T(\alpha_k^{-i})|}{T} \left| f[\alpha_k^{-i}](y) - f[\alpha_k^{-i}](y^i) \times_{j \neq i} P[y^j | \alpha_k^j] \right| < (I - 1)\phi. \quad (25)$$

We have that

$$\frac{1}{T} \sum_{n=1}^{T} g^i(s^i_n, a^i_n, y_n) \leq \frac{1}{T} \sum_{\alpha_k^{-i}} \left| \sum_{n \in T(\alpha_k^{-i})} g^i(s^i_n, a^i_n, y_n) - \sum_{\tilde{y}^{-i}} g^i(s^i_n, a^i_n, (\tilde{y}^{-i}, y_n)) P[\tilde{y}^{-i} | \alpha_k^{-i}] \right|$$

$$+ \frac{1}{T} \sum_{\alpha_k^{-i}} \sum_{n \in T(\alpha_k^{-i})} \sum_{\tilde{y}^{-i}} g^i(s^i_n, a^i_n, (\tilde{y}^{-i}, y_n)) P[\tilde{y}^{-i} | \alpha_k^{-i}].$$

By (25), the first sum is bounded by $\varepsilon/2$ on the event $\cap_{j \neq i} \{\tau^j_\phi(h_T) = 1\}$, and by $\bar{r}$ on its complement, which is of probability at most $\frac{\varepsilon}{2F}$. The expectation of the second sum under an arbitrary profile $\sigma$ does not depend on $\sigma^{-i}$ and is equal to the payoff induced by $(\sigma^i, \sigma_k^{-i})$. This implies the result. ■
Claim 14 implies $\tilde{k}_T(-e^i) \geq -w^i(k, T) - \varepsilon$ for all large $T$. Letting first $T \to +\infty$, then $k \to +\infty$, we get $k_2(-e^i) \geq -w^i_i - \varepsilon$, hence $k_2(\lambda) \geq -w^i_i$ since $\varepsilon$ is arbitrary. The reverse inequality is obvious. ■

**C.2 The Strategies**

Given $z \in Z$, and an initial distribution of states $p \in \Delta(S)$, we will construct a sequential equilibrium $\sigma$ with payoff $z$ for $\delta$ close enough to one.

As in Theorem ??, the play is divided into an infinite sequence of blocks, with odd blocks serving as transition blocks. Even blocks are now either “regular,” or devoted to the punishment of a single player. The behavior in odd and in regular even blocks is identical to that in Theorem ??, In contrast, the duration of a punishment even block is fixed and set equal to $(1 - \delta)^{-\beta}$ rounds.

The nature of an even block $k$ is dictated by the direction $\lambda[k] \in \Lambda$. If $\lambda[k]$ is close to $-e^i$ for some $i$, block $k$ is devoted to the punishment of player $i$. It is otherwise regular.

**C.2.1 Punishment Blocks**

The equilibrium behavior in punishment blocks relies on an elaborate version of Lemma 15, which we now introduce. Given $T \in \mathbb{N}$, $x : M \times Y^T \to \mathbb{R}^I$, $\delta < 1$ and $m \in M$, we denote by $G(m, \delta, x, T)$ a discounted $T$-round version of $G(T, x, \theta_*)$ without communication and initial state $m \in M$. That is, in each round $n = 1, \ldots, T$, players observe their private states $(s^i_n)$ choose actions $(a^i_n)$, and $(y_n, s_{n+1}) \in Y \times S$ is drawn according to $p_{s_n, a_n}$.

The payoff vector is

$$
\frac{1 - \delta}{1 - \delta^{T+1}} \left\{ \sum_{n=1}^T \delta^{n-1} r(s_n, a_n) + \delta^T x(m, \vec{y}) + \delta^T \theta_*(s_{T+1}) \right\},
$$

where $\vec{y} = (y_1, \ldots, y_T)$ is the sequence of public signals received along the play.

**Lemma 16** For every $\varepsilon_2 > 0$, there is a constant $\kappa \in \mathbb{R}$ and $\bar{\delta} < 1$ such that, for every player $i \in I$ and every discount factor $\delta \geq \bar{\delta}$, the following holds.

With $T = (1 - \delta)^{-1/2}$, there exists $x[i] : M \times \Omega^T_{\text{pub}} \to \mathbb{R}^I$ and $\gamma[i] \in \mathbb{R}^I$ such that:

(a) $\|x[i]\| \leq \kappa_T$ and $x^i[i](\cdot) \geq 0$.

(b) $|\gamma^i[i] - w^i_i| < \frac{\varepsilon_2}{2}$.

(c) $\gamma[i]$ is a sequential equilibrium payoff of $G(m, \delta, x[i], T)$ for every $m \in S$. 

63
Plainly, \( x \) and \( \gamma \) can then be chosen such that \( |\gamma^i[i] - w^i| < \varepsilon_2 \) and \( x^i[i](\cdot) > \frac{\varepsilon}{T} \).

**Proof.** Fix \( \varepsilon_2 > 0, i \in I \) and \( m \in M \). We set \( \varepsilon := \varepsilon_2/18, \kappa = \bar{r}/\varepsilon \) and prove that the conclusion holds for \( 2\kappa \).

The choice of \( \kappa \) guarantees that, for each \( T \), \( -T\bar{r} + \kappa T(1 - \varepsilon) > T\bar{r} + \kappa T(1 - 2\varepsilon) \).

Pick \( \delta_1 < 1 \) such that the same holds for each \( \delta \in (\delta_1, 1) \), when payoffs are discounted with \( \delta \) and \( T = (1 - \delta)^{-1/2} \):

\[
- \sum_{n=1}^{T} \delta^{n-1}\bar{r} + \delta^{2}\kappa T(1 - \varepsilon) > \sum_{n=1}^{T} \delta^{n-1}\bar{r} + \delta^{2}\kappa T(1 - 2\varepsilon).
\]

Pick now \( k \in \mathbb{N} \) such that \( |w^i - w^i(k)| < \varepsilon \), then \( \bar{T} \) such that \( |w^i(k) - w^i(k, T)| < \varepsilon \) for all \( T \geq T \).

We follow closely Lemma 15. We take \( \phi \) and \( T_0 \) as specified after Claim 13. We let \( \delta_2 < 1 \) be such that \( T = (1 - \delta)^{-1/2} \max(\bar{T}, T_0) \) for all \( \delta \geq \delta_2 \) and \( \delta_3 < 1 \) such that the normalized \( \delta \)-discounted sum of payoffs in the first \( T = (1 - \delta)^{-1/2} \) stages differs from the arithmetic mean by at most \( \varepsilon \), for all \( \delta \geq \delta_3 \).

Let \( \delta \geq \max(\delta_1, \delta_2, \delta_3) \) be arbitrary. Define as before \( x^i[i](\cdot) = 0 \), and, for \( j \neq i \), \( x^j[i] \) as in (24) with \( M = \kappa T \). Pick an arbitrary equilibrium \( \sigma[i, m] \) of \( G(m, \delta, x^i[i], T) \). It follows, as in Claim 14, that the (discounted) payoff \( \tilde{w}^i_m \) of player \( i \) under \( \sigma[i, m] \) does not exceed \( w^i[k, T] + 2\varepsilon \leq w^i + 4\varepsilon \).

Observe also that \( \tilde{w}^i_m \geq w^i - \varepsilon \) provided \( \delta \) is close enough to one. Hence \( \|\tilde{w}^i_m - \tilde{w}^i_m'\| < 5\varepsilon \).

For all \( j \in I \), define \( x^j_m[i] \) by adding to \( x^j[i] \) the quantity \( \max_{m'} \left( \tilde{w}^j_{m'}[i] - \tilde{w}^j_{m'}[i] \right) \) properly normalized. The added constant does not affect incentives, but ensures that the new equilibrium payoff vector, \( \gamma^j[i] \), is independent of \( m \in S \).

Given this redefinition of \( x \), we have that \( |\gamma^j[i] - w^j| < 9\varepsilon = \varepsilon_2/2 \) and \( x^j[i] \geq 0 \) as desired.

\[ \Box \]

### C.2.2 The Parameters

As in Theorem 5, given \( Z \), pick first \( \eta > 0 \) such that \( Z_\eta \) is contained in the interior of \( \mathcal{H}_2 \), and \( \varepsilon_0 > 0 \) such that \( \max_{Z_\eta} \lambda \cdot z < k_2(\lambda) - 2\varepsilon_0 \) for all directions \( \lambda \in \Lambda \). Let \( \kappa_R \) be obtained when applying Lemma 16 with \( \varepsilon := \varepsilon_0 \).

Pick \( \varepsilon_R < \varepsilon_0/\kappa_R \), and set \( \bar{\Lambda} := \Lambda \setminus \bigcup_i B(-e^i, \varepsilon_R) \). Replicating with the compact set \( \bar{\Lambda} \) the same compactness argument as in Section ??, we may assume wlog that the transfers \( x \) are

---

61Recall the definition of \( w^i(k, T) \) from Lemma 15. Note however that the definition has to be amended, to reflect the fact that the initial state profile is \( m \) and no longer \( s_* \). We still use the same notation.
picked from a finite set of maps $\mathcal{X}$ as $\lambda$ varies through $\tilde{\Lambda}.\theta_{\delta}$ and $x \in \mathcal{X}$, valid for all $\delta < 1$. Pick next $\beta_{\ast} \in (0, 1/2)$.

As in the proof of Theorems 6 and ??, we fix $\kappa_{2}$ large enough, let $\varepsilon_{1} \in (0, \varepsilon_{0})$, set $\varepsilon := \varepsilon_{1}/2\kappa_{1}$, and then let $\tilde{\zeta}$ be given by Lemma ?? applied with $\varepsilon$. Given these values, we finally let $\tilde{\delta}$ be close enough to one, so that a finite number of inequalities hold for all $\delta \geq \tilde{\delta}$. Again, we omit the exact conditions on $\kappa_{2}$ and $\tilde{\delta}$ under which the computations below are valid.

C.2.3 The Updating Process

We follow Section ???. Consider a block $k + 1$, starting in round $n + 1 := \tau_{k+1}$. If $k + 1$ is even, we define first $w[k + 1]$ and $\tilde{w}_{n+1}$ by (??) and (??), and we pick $\lambda[k + 1] \in \Lambda$ so that the conclusion of Lemma ?? holds.

If $\lambda[k + 1] \in \tilde{\Lambda}$, so that block $k + 1$ is regular, we define $(\rho[k + 1], x[k + 1], v[k + 1], \theta[k + 1])$ as in Section ???, and let $z[k + 1]$ be defined by (??).

If instead $\lambda[k + 1] \in B(-e^{i}, \varepsilon_{3})$ for some $i \in I$, we set $w[k + 1] = z[k + 1] := 1 - \delta T z[k] - \frac{1 - \delta^{T}}{\delta T} \gamma[i] + (1 - \delta)x[i](m_{\tau_{k}}, y_{\tau_{k}}, \ldots, y_{\tau_{k}+1-1}).$

The process is initialized as in Theorem 5.

C.2.4 The Strategies

Fix a player $i$. Let block $k$ be an $i_{\ast}$-punishment block. If the report of $i$ in round $\tau_{k}$ was truthful ($m_{\tau_{k}} = s_{\tau_{k}}^{i}$) player $i$ plays $\sigma^{i}[m_{\tau_{k}}, i_{\ast}]$ up to round $\tau_{k+1} = \tau_{k} + T$. If instead player $i$ lied about his state in the initial round, $\tau_{k}$, of the punishment phase, player $i$ plays a sequential rational strategy against $\sigma^{-i}[m_{\tau_{k}}, i_{\ast}]$ in the game $G(s_{\tau_{k}}^{i}, m_{\tau_{k}}^{-i}, \delta, x[i_{\ast}], T)$.

In any block which is not a punishment block, the strategy of player $i$ is defined as in the proof of Theorem 5.

That $\sigma$ is well-defined follows from the next lemma.
Lemma 17 One has $w[k] \in Z_\eta$ for $k$ even.

Proof. We proceed as in Lemma ??. Assume that $w[k] \in Z_\eta$ for some even $k$. It suffices to deal with the case where block $k$ is a $i_*$-punishment block, for some $i_* \in I$. From the updating formula, it follows that

$$
\|w[k + 1] - z[k]\| \leq \frac{1 - \delta^T}{\delta^T} \kappa_2 + (1 - \delta)\kappa_2 T,
$$

so that

$$
\|w[k + 2] - w[k]\| \leq \|w[k + 2] - w[k + 1]\| + \|z[k] - w[k]\| + \|w[k + 1] - z[k]\| \\
\leq \frac{1 - \delta}{\delta \xi_*} \kappa_2 + \frac{1 - \delta^T}{\delta^T} \kappa_2 + (1 - \delta)\kappa_2 T + (1 - \delta)\kappa_2.
$$

Denote by $\zeta$ the right-hand side.

On the other hand,

$$
\lambda[k] \cdot (w[k + 2] - w[k]) \leq \frac{1 - \delta}{\delta \xi_*} \kappa_2 + (1 - \delta)\kappa_2 + \lambda[k] \cdot (w[k + 1] - z[k]).
$$

Since $\lambda[k] \cdot (z[k] - \gamma[i]) \leq -2\varepsilon_0 \times \frac{1 - \delta^T}{\delta}$ and $\lambda[k] \cdot x[i] \leq 0$, it follows from elementary computations and the choice of $\delta$ that

$$
\lambda[k] \cdot (w[k + 2] - w[k]) \leq -\varepsilon_1 \xi,
$$

hence $w[k + 2] \in Z_\eta$. □

C.2.5 The Equilibrium Property

Fix a player $i \in I$. As in Theorem ??, the construction of the strategy profile $\sigma$ ensures that the continuation payoff of player $i$ at the action step of a given round $n$ is given by

$$
\gamma^i(\omega_{\text{pub},n-1, s_n; T_n}) = z_n + (1 - \delta)\theta_*(s_n) \text{ or } z_n + (1 - \delta)\theta_*(\omega_{\text{pub},n-1, s_n})
$$

whenever $m^i_n = s^i_n$ and round $n$ is not part of a punishment block. In addition, the continuation payoff in the first stage of a $i_*$-punishment block is $\gamma^i[i_*]$, again if $m_n^i = s_n^i$.

For use below, we make the following observation. Fix $m \in M$, $i \in I$, $s^i \in S^i$, and consider the variant $\tilde{G}^i_i(m, \delta, x[i_*], T)$ of $G(m, \delta, x[i_*], T)$ in which the initial state of $i$ is $s^i$ instead of $m^i$.\footnote{But final transfers are still given by $x[i_*](m, \tilde{y}).$} Thanks to the irreducibility property, the highest payoff of $i$ in $\tilde{G}^i_i(m, \delta, x[i_*], T)$ when facing $\sigma^{-i}[i_*, m]$ differs from the payoff $\gamma^i[i_*]$ induced by $\sigma[i_*, m]$ in $G(m, \delta, x[i_*], T)$ (and

62But final transfers are still given by $x[i_*](m, \tilde{y}).$}
therefore from the payoff induced by $\sigma[i^*, (m^{-i}, s^i)]$ in $G((m^{-i}, s^i), \delta, x[i^*], T))$ by at most $(1 - \delta)\bar{\kappa}$, where $\bar{\kappa}$ is a constant that only depends on $\kappa_2$ and on the primitives of the game. Thus, misreporting at the beginning of a punishment block does not benefit much.

That player $i$ cannot profitably deviate at the action step of a given round $n$ follows as in Theorem ??, unless $n$ is part of a punishment block, in which case it follows from the sequential rationality of $\sigma$ in that block.

That player $i$ cannot profitably deviate by lying in a regular block also follows as in Theorem ??, On the other hand, players babble in punishment blocks.

We now place ourselves at the reporting step of a round $n$ in a transition block. There are two cases: either $n$ is the first stage of the transition block, following a $i^*$-punishment block; or it is not. In the former case, the belief of $i$ is derived from the public history and the strategies $\sigma[i^*]$; in the latter, it is derived using $\rho^*$. In both cases, the belief of $i$ over $s^{-i}_n$ has full support, and the optimality of truth-telling follows along the lines of Theorem ??, using (i) the ex post optimality of truth-telling under $\rho^*$ and (ii) the fact that misreporting in the first round of a punishment block has only a minor impact (of the order of $(1 - \delta)$, see above) on the continuation payoff of player $i$. 

67