Truthful Equilibria in Dynamic Bayesian Games

Johannes Hörner[*], Satoru Takahashi[†] and Nicolas Vieille[‡]

September 24, 2013

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 may be correlated and values be interdependent. The focus is on equilibria in which players report truthfully. The characterization generalizes that for repeated games, reducing the analysis to a collection of 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).[1] These constructions are difficult to extend beyond a certain point, however. Instead, our method allows us to deal with

- moral hazard (imperfect monitoring);
- endogenous serial correlation (actions affecting transitions);
- correlated types (across players) or/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 focus is on *truthful* equilibria, in which players truthfully reveal their type at the beginning of each round, after every history.

Our main result characterizes a subset of the limit set of equilibrium payoffs as $\delta \to 1$. While the focus on truth-telling equilibria is restrictive in the absence of any commitment, it nevertheless turns out that this limit set generalizes the payoffs obtained in all known special cases so far –with the exception of the lowest equilibrium payoff in Renault, Solan and Vieille (2013), who also characterize Pareto-inferior "babbling" equilibria. When types are independent (though still possibly affected by one's own action), and payoffs are private,

[1]This is not to say that the recursive formulations of Abreu, Pearce and Stacchetti (1990, hereafter APS) cannot be adapted to such games. See, for instance, Cole and Kocherlakota (2001), Fernandes and Phelan (2000), or Doepke and Townsend (2006). These papers develop methods that are extremely useful for numerical purposes for a given discount rate, but provide little guidance regarding qualitative properties of the (asymptotic) equilibrium payoff set.

for instance, all Pareto-optimal payoffs that are individually rational (*i.e.*, dominate the stationary minmax payoff) are limit equilibrium payoffs (provided monitoring satisfies the usual identifiability conditions). In fact, with the exception of individual rationality, which could be further refined, our result is actually a characterization of the limit set of equilibrium payoffs. In this sense, this is a folk theorem. In the revelation game, there is no loss of generality in restricting attention to truthful equilibria: a revelation principle holds if players commit to actions, even if they do not commit to future reports. This implies that, when actions do not affect transitions, and leaving aside the issue of the minmax payoff, transitions do not matter for the limit set, just the invariant distribution, as is rather intuitive.

When the monitoring structure has a product structure (the condition under which the folk theorem for repeated games with public monitoring is actually a characterization of the limit equilibrium payoff set, rather than a lower bound), we show that a mild weakening of truth-telling –allowing for "punishment phases" during which communication is uninformative– yields as equilibrium payoffs a set that coincides with the theoretical upper bound on all Bayes Nash equilibrium payoffs. This provides a definitive characterization of the limit equilibrium payoff set for such games.

When types are correlated, then all feasible and individually rational payoffs can be obtained in the limit. 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.

These findings mirror standard results from static mechanism design, *e.g.* those of Arrow (1979), d'Aspremont and Gérard-Varet (1979) for the independent case, and those of d'Aspremont, Crémer and Gérard-Varet (2003) in the correlated case. This should come as no surprise, as our characterization is a reduction from the repeated game to a (collection of) one-shot Bayesian game with transfers, to which standard techniques can be applied. If there is no incomplete information about types, this one-shot game collapses to the algorithm developed by Fudenberg and Levine (1994) to characterize public perfect equilibrium payoffs, also used in Fudenberg, Levine and Maskin (1994, hereafter FLM) to establish a folk theorem under public monitoring.

Our results stand in contrast with the techniques based on review strategies (see Escobar and Toikka 2013 for instance) whose adaptation to incomplete information is inspired by the linking mechanism described in Fang and Norman (2006) and Jackson and Sonnenschein (2007). They imply that, under mild assumptions, and as is 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 requires assumptions closely related to those called for by budget-balance in static mechanism design. Indeed, our sufficient conditions are reminiscent of those in this literature, in particular the weak identifiability condition introduced by Kosenok and Severinov (2008).

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.

Consider the case of independent types for instance. Note that the long-run (or asymptotic) payoff of a player must be independent of his current state, because this state is private information and the Markov chain is irreducible. Relative to a stochastic game with observable states, there is a collapse of dimensionality as $\delta \to 1$. Yet the "transient" component of the payoff, which depends on the state, must be taken into account: a player's incentives to take a given action depend on the action's impact on later states. This component must be taken into account, but it cannot be treated as an exogenous transfer. This differs from repeated games (without persistent types): there, incentives are provided by continuation payoffs who become arbitrarily large relative to the per-round rewards as players get patient. Hence, asymptotically, the continuation play can be summarized by a transfer whose magnitude is unrestricted. But with private, irreducible types, the differences in continuation payoffs across types of a given player do not become arbitrarily large relative to the flow payoff (they fade out exponentially fast).

So: the transient component cannot be ignored, although it cannot be exploited as a standard transfer. But, for a given transfer rule and a Markov strategy, this component is easy to compute, using the *average cost optimality equation* (ACOE) from dynamic programming. This equation converts the relative future benefits of taking a particular action, given the current state, into an additional per-round reward. So it can be taken into account, and since it cannot be exploited, incentives will be provided by transfers that are independent of the type (though not of the report). After all, this independence is precisely a feature of transfers in static mechanism design, and our exclusive reliance on this channel illustrates again the lack of linkage in our analysis. What requires considerable work, however, is to show how such type-independent transfers can get implemented (after all, there are no actual transfers: all continuation payoffs must be generated by a sequence of action profiles, whose evaluation depends on the initial type); and why we can compute the transient component as if the equilibrium strategies were Markov, which they are not.

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 (Green, 1987), there is an extensive literature on recursive formulations for fixed discount factors (Fernandes and Phelan, 1999; Cole and Kocherlakota, 2001; Doepke and Townsend, 2006). In game theory, recent progress has been made in the case in which the state is observed, see Fudenberg and Yamamoto (2012) and Hörner, Sugaya, Takahashi and Vieille (2011) 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.

More related are the papers by Escobar and Toikka (2013), already mentioned, Barron (2012) and Renault, Solan and Vieille (2013). All three papers assume that types are independent across players. Barron (2012) introduces imperfect monitoring in Escobar and Toikka (2013), but restricts attention to the case of one informed player only. This is also the case in Renault, Solan and Vieille (2013). 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 The Model

We consider dynamic games with imperfectly persistent incomplete information. The stage game is as follows. The finite set of players is denoted I. Each player $i \in I$ has a finite set S^i of (private) states, 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;

2. Players simultaneously make reports $(m^i_n)_{i=1}^I \in \times_i M^i$, where M^i is a finite set. De-

pending on the context, we set M^i as either S^i or $(S^i)^2 \times A^i$, as explained below. These reports are publicly observed;

3. The outcome of a public correlation device is observed. For concreteness, it is a draw from the uniform distribution on $[0,1]$;[2]

4. Players independently choose actions $a_n^i \in A^i$. Actions taken are not observed;

5. A public signal $y_n \in Y$, a finite set, and the next state profile $s_{n+1} = (s_{n+1}^i)_{i \in I}$ are drawn according to some joint distribution $p_{s_n,a_n} \in \Delta(S \times Y)$.

Throughout, we assume that the transition function p is such that the support of $p_{\bar{s},\bar{a}}$ does not depend on $\bar{s}$ and is equal to $S \times Y(\bar{a})$ for some $Y(\bar{a}) \subseteq Y$.[3] This implies that (i) the controlled Markov chain (s_n) is irreducible under any Markov strategy, (ii) public signals, whose probability might depend on $(\bar{s}, \bar{a})$, do not allow players to rule out any type profile s. This is consistent with perfect monitoring. Note that actions might affect transitions.[4] The irreducibility of the Markov chain is a strong assumption, ruling out among others the case of perfectly persistent types (see Aumann and Maschler, 1995; Athey and Bagwell, 2008). Unfortunately, it is well known that the asymptotic analysis is very delicate without such an assumption (see Bewley and Kohlberg, 1976). On the other hand, the full-support assumption on S and the state-independence of the signal profile are for convenience: detecting deviations only becomes easier when it is dropped, but it is then necessary to specify out-of-equilibrium beliefs regarding private states.[5,6]

[2]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.

[3]Throughout the paper, we use $\bar{s}, \bar{a}, \bar{y}$, etc. when referring to the values of variables s, a, y, etc. in the "previous" round.

[4]Accommodating observable (public) states, as modeled in stochastic games, requires minor adjustments. One way to model them is to append such states as a component to each player's private state, perfectly correlated across players.

[5]We allow $Y(\bar{a}) \subsetneq Y$ to encompass the important special case of perfect monitoring, but the independence from the state $\bar{s}$ ensures that players do not need to abandon their belief that players announced states truthfully. However, note that this is not quite enough to pin down beliefs about s_{n+1} when $y_n \notin Y(a)$, when y_n is observed, yet a was supposed to be played; because transitions can depend on the action profile, beliefs about s_{n+1} depend on what players think the actual action profile played was. This specification can be chosen arbitrarily, as it plays no role in the results.

[6]In fact, our results only require that it be unichain, *i.e.* that the Markov chain defined by any Markov strategy has no two disjoint closed sets. This is the standard assumption under which the distributions specified by the rows of the limiting matrix $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} p(\cdot)^i$ are independent of the initial state; otherwise the average cost optimality equation that is used to analyze, say, the cooperative solution is no longer valid.

We also write $p_{s,a}(y)$ for the marginal distribution over signals y given (s, a), $p_{s,a}(t)$ for the marginal distribution over types t in the "next" round, etc., and extend the domain of these distributions to mixed action profiles $\alpha \in \Delta(A)$ in the customary way.

The stage-game payoff (or *reward*) of player i is a function $r^i : S \times A \to \mathbf{R}$, 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 $g^i : S \times A^i \times Y \to \mathbf{R}$, $r^i(s, a) = \mathbf{E}[g^i(s, a^i, y) \mid a]$. This interpretation is particularly natural in the case of private values (in which case we may think of $g^i(s^i, a^i, y)$ as the observed stage-game payoff), but except in that case, we do not assume that the reward satisfies this factorization property.

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 $p_1 \in \Delta(S)$, which plays no role in the analysis, given the irreducibility assumption and the focus on equilibrium payoff vectors as elements of $\mathbf{R}^I$ as $\delta \to 1$.

A special case of interest is *independent private values* (hereafter, IPV). This is the case in which (i) payoffs of a player only depend on his private state, not on the others', that is, for all (i, s, a), $r^i(s, a) = r^i(s^i, a)$, (ii) conditional on the public signal y, types are independently distributed. A more precise definition is given in Section 6.

But we do not restrict attention to private values, nor to independent types. 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.[7] 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, 2007) in static mechanism design with interdependent valuations. Hence, we assume that a player's private action, private state, the public signal and report profile is all the information available to him.[8]

[7] With this interpretation, pointed out by Athey and Segal (2013), interdependent values with observable payoffs reduce to private values *ex post*, calling for a second round of messages.

[8] However, our notion of equilibrium is sensitive to what goes into a state: by enlarging it, one weakly increases the equilibrium payoff set. For instance, one could also include in a player's state his previous realized action, which following Kandori (2003) is useful even when incomplete information is trivial and the

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 (2013), Renault, Solan and Vieille (2013), etc.[9] 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.

As we insist on sequential rationality, players are assumed to be unable to commit. Hence, the revelation principle does not apply. As is well known (see Bester and Strausz, 2000, 2001), it is not possible *a priori* to restrict attention to direct mechanisms, corresponding to the choice $M^i = S^i$ (or $M^i = A^i \times (S^i)^2$, as explained below), let alone truthful behavior.

Yet this is precisely what we restrict attention to. The next section illustrates some of the issues that this raises.

3 Some Examples

EXAMPLE 1—*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

	L	R
T	1	0
B	0	0

s^1

	L	R
T	0	0
B	0	1

$\hat{s}^1$

Figure 1: Player 1's reward in Example 1

game is simply a repeated game with public monitoring; such an enlargement is peripheral to our objective and will not be pursued here.

[9]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.

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.

Set $M^1 := \{s^1, \hat{s}^1\}$, so that player 1 can disclose his state if he wishes to. Will he? By revealing the state, player 2 can secure a payoff of 0 by playing R or L depending on player 1's report. Yet player 1 can secure a payoff of 1/4 by choosing reports and actions at random. In fact, this is the (uniform) value of this game for $p = 1$ (Aumann and Maschler, 1995). When $p < 1$, player 1 can actually 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. Silence is called for.

Just because we may focus on the silent game does not mean that it is easy to solve. Its (limit) value for arbitrary $p > 2/3$ is still unknown.[10] 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.

The surprise, then, is that the literature on Markovian games (Athey and Bagwell, 2001, 2008, Escobar and Toikka, 2013, Renault, Solan and Vieille, 2013) manages to get positive results at all: in most games, efficiency requires coordination, and thus disclosure of (some) private information. As will be clear from Section 6, existence is much easier to obtain in the IPV environment, the focus of most of these papers. Example 1 involves both interdependent values and independent types, an ominous combination in mechanism design. In our dynamic environment as well, positive results obtain whenever one or the other assumption is relaxed.

Example 2—*A Game that Leaves No Player Indifferent.* 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 $A^2 = \{L, R\}$. Rewards are given by Figure 2 (values are private).

[10] It is known for $p \in [1/2, 2/3]$ and some specific values. Pęski and Toikka (private communication) have recently shown that this value is non-increasing in p.

	L	R
T	$1,1$	$1,-1$
B	$0,-1$	$0,1$

s^1

	L	R
T	$0,1$	$0,-1$
B	$1,-1$	$1,1$

$\hat{s}^1$

Figure 2: A two-player game in which the mixed minmax payoff cannot be achieved.

The two types s^1 and $\hat{s}^1$ are i.i.d. over time and equally likely. Monitoring is perfect. To minmax player 2, player 1 must randomize uniformly, independently of his type. But clearly player 1 has a strictly dominant strategy in the repeated game, playing T in state s^1 and B in state $\hat{s}^1$. Even if player 1's continuation utility were to be chosen freely, it would not be possible to get player 1 to randomize in both states: to play B when his type is s^1, or T when his type is $\hat{s}^1$, he must be compensated by \$1 in continuation utility. But then he has an incentive to report his type incorrectly, to pocket this promised utility while playing his favorite action.

This example illustrates that fine-tuning continuation payoffs to make a player indifferent between several actions in several private states simultaneously is generally impossible to achieve with independent types. This still leaves open the possibility of a player randomizing for *one* of his types. This is especially useful when each player has only one type, like in a standard repeated game, as it then delivers the usual mixed minmax payoff. Indeed, the characterization below yields a minmax payoff somewhere in between the mixed and the pure minmax payoff, depending on the particular game considered.

This example also shows that truth-telling is restrictive even with independent private values: in the silent game, player 1's unique equilibrium strategy minmaxes player 2, as he is left guessing player 1's action. Leaving a player in the dark about one's state can serve as a substitute for mixing at the action step.

Example 3—***Waiting for Evidence.*** There are two players. Player 1 has $K+1$ types, $S^1 = \{0, 1, \ldots, K\}$, while player 2 has only two types, $S^2 = \{0, 1\}$. Transitions do not depend on actions (ignored), 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 closely 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, states) be as rich as player i's, unlike in static mechanism design with correlated types (the familiar "spanning condition" of Crémer and McLean, 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 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^i_n, s_{n-1}), summarized by the distribution of an initial segment? 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 $(s^{-i}_{n'})_{n' \geq n}$.[11]

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 s^i_n on $-i$'s next private state, s^{-i}_{n+1}. Because this suffices to obtain sufficient conditions analogous to those in the static case, we will limit ourselves to this conditioning.[12]

4 Truthful Equilibria

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 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 $(\mathfrak{m}^i, \mathfrak{a}^i) = (\mathfrak{m}^i_n, \mathfrak{a}^i_n)_{n \in \mathbb{N}}$

[11] The 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).

[12] See Obara (2008) for some of the difficulties encountered in dynamic settings when attempting to extend results from static mechanism design with correlated types.

with $\mathfrak{m}^i_n : H^i_n \times S^i \to \Delta(M^i)$, and $\mathfrak{a}^i_n : H^i_n \times S^i \times M \to \Delta(A^i)$ that specify i's report and action as a function of his private information, his current state and the report profile in the current round.[13] A strategy profile $(\mathfrak{m}, \mathfrak{a})$ defines a distribution over finite and by extension over infinite histories in the usual way, and we consider the sequential equilibria of this game.

A special class of games are "standard" repeated games with public monitoring, in which S^i is a singleton set for each player i and we can ignore the $\mathfrak{m}$-component of players' strategies. For such games, Fudenberg and Levine (1994) provide a convenient algorithm to describe and study a subset of equilibrium payoffs–public perfect equilibrium payoffs. A public perfect equilibrium (PPE) is an equilibrium in which players' strategies are public; that is, $\mathfrak{a}$ is adapted to $(H_{\text{pub},n})_n$, so that players ignore any additional private information (their own past actions). 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. Let Λ denote the unit sphere of $\mathbf{R}^I$. We refer to $\lambda \in \Lambda$ (or λ^i) as weights, although the coordinates need not be nonnegative.

Definition 1 *Fix $\lambda \in \Lambda$. Let*

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

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

(i) *α is a Nash equilibrium with payoff v of the game with payoff $r(a) + \sum_y p_a(y)x(y)$;*

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

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

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

Our purpose is to obtain a similar characterization for the broader class of games considered here. To do so while preserving the recursive nature of the equilibrium payoff set that will be described compels us to focus on a particular class of equilibria in which players report truthfully their private state in every round, on and off path, and do not condition on their earlier private information, but only on the public history and their current state.

The complete information game with transfers x that appears in the definition of the score must be replaced with a two-step Bayesian game with communication, formally defined in the next section. To describe a Bayesian game, one needs a type space and a prior distribution

[13] Recall however that a public correlation device is assumed, although it is omitted from the notations.

over this space. Clearly, in the dynamic game, player i's beliefs about his opponents' private states depend on the previous reports m_{n-1}. Hence, we are led to consider a family of Bayesian games, parameterized by such reports. In addition, as Example 3 makes clear, player i's transfer x not only depends on this parameter, but also on s_{n+1}^{-i}, the reported states of the other players in the following round.

What is a player's type, *i.e.*, what is a player's information that is relevant in round n? Certainly this includes his private state s_n^i. Because player i does not observe s_n^{-i}, his conditional belief about these states is also relevant, to predict $-i$'s behavior and because values need not be private. This is what creates the difficulty in Example 1: because player 1 does not want to disclose his state, player 2 must use all available information to predict s_n^{-i} as precisely as possible, which requires keeping track of the entire history of play.

However, when players report truthfully their information, player i knows s_{n-1}^{-i}; because player $-i$ uses a strategy that does not depend on his earlier private types $(s_{n'}^{-i})_{n'<n-1}$, player i's beliefs about those becomes irrelevant.

But knowing s_{n-1}^{-i} is not enough for i to form his beliefs about s_n^{-i}. First, s_{n-1}^i matters, because the Markov chains (s_n^i) and (s_{n-1}^i) need not be independent across players, and s_n need not be independent of s_{n-1} either. Second, a_{n-1} matters, because actions affect transitions. Regarding a_{n-1}^{-i}, player i is able to infer it from the equilibrium strategies, given the public history (including the last report).[14] And of course player i knows the action he played himself. Given that we must consider histories in which he has deviated, we must allow for this action a_n^i to be arbitrary. The natural choice for the message space is then $M^i = S^i \times A^i \times S^i$, so that player i be able to report all his private information. Along the equilibrium path, this involves repetitions. But it matters when the last report of player i was not truthful regarding his current state. Players $-i$ cannot detect such a deviation, which is "on-schedule" according to Athey and Bagwell (2008). For truthful reporting off path, the choice of M^i makes a difference: with $M^i = S^i$, player i would be asked to tell the truth regarding his "payoff-type," but possibly to lie about his "belief-type" (which would be incorrectly believed to be determined by his report of s_{n-1}^i, along with his current report).

In the IPV case, however, this enlargement is unnecessary, as past deviations do not affect i's conditional beliefs. We will then set $M^i = S^i$.

In what follows, a *type* of player i refers to the true element of M^i, to be distinguished from his state, an element of S^i.

A strategy $(\mathfrak{m}^i, \mathfrak{a}^i)$ is *public* and *truthful* if $\mathfrak{m}_n^i(h_n^i, s_n^i) = (s_{n-1}^i, a_{n-1}^i, s_n^i)$ (or s_n^i in the IPV

[14]More precisely, player i knows the mixed strategy α_{n-1}^{-i} employed.

case) for all histories h^i_n, $n \geq 1$, and $\mathfrak{a}^i(h^i_n, s^i_n, m_n)$ depends on $(h_{\text{pub},n}, s^i_n, m_n)$ only (with the obvious adjustment in the initial round). The solution concept is sequential equilibrium in public and truthful strategies.

The next section describes the family of Bayesian games formally.

5 The Main Result

In this section, $M^i := S^i \times A^i \times S^i$ for all i. 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.

5.0.1 The Average Cost Optimality Equation

Our analysis makes use of the so-called 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 μ, and a payoff function $u : S \to \mathbf{R}$. Assume that 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 $v = \mathbf{E}_\mu[u(s)]$. While this long-run payoff is independent of the initial state, discounted payoffs are not. Lemma 1 below provides a normalized measure of the differences in discounted payoffs, for different initial states. Here and it what follows, t stands for the "next" state profile, given the current state profile s.

Lemma 1 *There is $\theta : S \to \mathbf{R}$ such that*

$$v + \theta(s) = u(s) + \mathbf{E}_{t \sim p_s(\cdot)} \theta(t).$$

The map θ 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 $\lim_{\delta \to 1} \dfrac{\gamma_\delta(s) - \gamma_\delta(s')}{1 - \delta}$, where $\gamma_\delta(s)$ is the discounted payoff when starting for s. For this reason, following standard terminology, call θ the (vector of) *relative values*.

The map θ 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 μ. 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 μ and θ are usually defined as the solutions of a finite system of equations –the balance equations and the equations stated in Lemma 1. But in the ergodic case that we are concerned with, explicit formulas exist for both of them. (See, for instance, Iosifescu, 1980, p.123, for the invariant distribution; and Puterman, 1994, Appendix A for the relative values.)

5.0.2 Admissible Pairs

The characterization of FL for repeated games involves a family of optimization problems, in which one optimizes over equilibria α of the underlying stage game, with payoff functions augmented by transfers x, see Definition 1.

Because we insist on truthful equilibria, and because we need to incorporate the dynamic effects of actions on states, we must consider *policies* instead, *i.e.* maps $\rho : S \to \Delta(A)$ and transfers, such that reporting truthfully and playing ρ 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 and current public outcomes, as well as on the next reported states.

Let such a policy $\rho : S \to \Delta(A)$, and transfers $x : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ be given. We will 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.[15] Assuming states are truthfully reported and actions chosen according to ρ, the sequence (ω_n) of outcomes is a unichain Markov chain, and so is the sequence $(\tilde{\omega}_n)$, where $\tilde{\omega}_n = (\omega_{\text{pub},n-1}, s_n)$, with transition function denoted π_ρ.

Let $\theta_{\rho,r+x} : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ the relative values of the players, obtained when applying Lemma 1 to the latter chain (and to all players).

Thanks to the ACOE, the condition that reporting truthfully and playing ρ 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 ρ is an equilibrium in

[15] This requirement will not be systematically stated, but it is assumed throughout.

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

$$r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta_{\rho, r+x}(\omega_{\text{pub}}, t),$$

where (y,t) are chosen according to $p_{s,a}$ and $\omega_{\text{pub}} = (m, y)$.[16]

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 $D^i_{\rho,x}(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i)$ denote the two-step decision problem in which

Step 1 $s \in S$ is drawn according to the belief held by player i;[17] player i is informed of s^i, then submits a report $m^i \in M^i$;

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

$$r^i(s,a) + x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t^{-i}) + \theta^i_{\rho, r+x}(\omega_{\text{pub}}, t), \tag{1}$$

where $a^{-i} = \rho^{-i}(m)$ and the pair (y,t) is drawn according to $p_{s,a}$, and $\omega_{\text{pub}} := (m, y)$.

We denote by $\mathcal{D}^i_{\rho,x}$ the collection of decision problems $D^i_{\rho,x}(\bar{\omega}_{\text{pub}}, \bar{m}^i, \bar{a}^i)$.

Definition 2 *The pair (ρ, x) is* admissible *if all optimal strategies of player i in $\mathcal{D}^i_{\rho,x}$ 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, $a^i = \rho^i(s)$ is an (not necessarily unique) optimal action (Obedience).*

Some comments are in order. The condition that ρ 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 $r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}) + \theta_{\rho, r+x}(\omega_{\text{pub}}, t)$.

[16] Lemma 1 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 (M.D.P.), a policy ρ is optimal if and only if, for all states s, $\rho(s)$ maximizes the right-hand side of the equations of Lemma 1. Both results will be invoked interchangeably.

[17] Recall that player i assumes that players $-i$ report truthfully and play ρ^{-i}. Hence player i assigns probability 1 to $\bar{s}^{-i} = \bar{m}_c^{-i}$, and to previous actions being $\bar{a}^{-i} = \rho^{-i}(\bar{m}_c)$; hence this belief assigns to $s \in S$ the probability $p_{\bar{s},\bar{a})}(s \mid \bar{y})$. This is the case unless $\bar{y}$ is inconsistent with $\bar{a}^{-i} = \rho^{-i}(\bar{m}_c)$; if this is the case, use the same updating rule with some other arbitrary $\tilde{a}^{-i}$ such that $\bar{y} \in Y(\tilde{a}^{-i}, \bar{a}^i)$.

The truth-telling condition is slightly more delicate to interpret. Consider first an outcome $\bar{\omega} \in \Omega$ such that $\bar{s}^i = \bar{m}_c^i$ 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 $p_{\bar{s},\bar{a}}(\cdot \mid \bar{y})$. 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 $r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta_{\rho,r+x}(\omega_{\text{pub}}, t)$. The admissibility condition for such an outcome $\bar{\omega}$ is equivalent to requiring that truth-telling followed by ρ is an equilibrium of this Bayesian game, with "strict" incentives at the reporting step.[18]

The admissibility requirement in Definition 2 is demanding, however, in that it requires in addition truth-telling to be optimal for player i at any outcome $\bar{\omega}$ such that $(\bar{s}^{-i}, \bar{a}^{-i}) = (\bar{m}_c^{-i}, \rho^{-i}(\bar{m}_c))$, but $\bar{s}^i \neq \bar{m}_c^i$ (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 ρ^{-i}. Note that we require truth-telling to be optimal ($m^i = (\bar{s}^i, \bar{a}^i, s^i)$) even if player i has lied in the previous round ($\bar{m}_c^i \neq \bar{s}^i$) about his current state. On the other hand, Definition 2 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 2 is equivalent to saying that $\rho^i(s)$ is one best-reply to $\rho^{-i}(s)$ in the complete information game with payoff function given by (15) when $m = (\bar{s}, \bar{a}, s)$.

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 θ. Our construction involves "repeated games" strategies that are "approximately" policies, so that θ can be derived from (ρ, 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. Fortunately, this requirement is not demanding, as it will be implied by standard assumptions in the correlated case, and by some weak assumption (Assumption **3** below) on feasible policies in the IPV case.

We denote by $\mathcal{C}_0$ the set of admissible pairs (ρ, x).

[18] 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.

5.0.3 The Characterization

For given weights $\lambda \in \Lambda$, we denote by $\mathcal{P}_0(\lambda)$ the optimization program $\sup \lambda \cdot v$, where the supremum is taken over all triples (v, ρ, x) such that

- $(\rho, x) \in \mathcal{C}_0$;
- $\lambda \cdot x(\cdot) \leq 0$;
- $v = \mathbf{E}_{\mu_\rho} [r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)]$, where $\mu_\rho \in \Delta(\Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S)$ is the invariant distribution under trauth-telling and ρ, so that v is the long-run payoff induced by (ρ, 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). In what follows, budget-balance refers to this property.

We denote by $k_0(\lambda)$ the value of $\mathcal{P}_0(\lambda)$ and set $\mathcal{H}_0 := \{v \in \mathbf{R}^I, \lambda \cdot v \leq k_0(\lambda) \text{ for all } \lambda \in \Lambda\}$.

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

This result is simple enough. For instance, in the case of "standard" repeated games with public monitoring, Theorem 2 generalizes FLM, yielding the folk theorem with the mixed minmax under their assumptions.

To be clear, there is no reason to expect Theorem 2 to provide a characterization of the entire limit set of truthful equilibrium payoffs. One might hope to achieve a bigger set of payoffs by employing finer statistical tests (using the serial correlation in states), 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). There is an obvious cost in terms of the simplicity of the characterization. As it turns out, ours is sufficient to obtain all the equilibrium payoffs known in special cases, and more generally, all individually rational Bayes Nash equilibrium payoffs (including the Pareto frontier) under independent private values, as well as a folk theorem under correlated values.[19]

[19]Besides, an exact characterization would require an analysis in $\mathbf{R}^S$, mapping each type profile into a payoff for each player. When the players' types follow independent Markov chains and values are private, this makes no difference, as the players' limit equilibrium payoff *must* be independent of the initial type profile, given irreducibility and incentive-compatibility. But when types are correlated, it is possible to assign different (to be clear, long-run) equilibrium payoffs to a given player, as a function of the initial state.

Theorem 2 admits an extension 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 (ρ, 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 ρ^{LR} is the restriction of ρ to players in LR. The requirements on the long-run players are the same as in Definition 2.

5.1 Proof Overview

Here, we explain the main ideas behind the proof of Theorem 2. For simplicity, we assume perfect monitoring and action-independent transitions. For notational simplicity also, we limit ourselves to admissible pairs (ρ, x) such that transfers $x : M \times M \times A \to \mathbf{R}^I$ do not depend on previous public signals (which do not affect transitions here). This is not without loss of generality, but going to the general case is mostly a matter of notations.

Our proof is best viewed as an extension of the recursive approach of FLM to the case of persistent, private information. To serve as a benchmark, assume first that types are i.i.d. across rounds, with law $\mu \in \Delta(S)$. The game is then truly a repeated game, and the characterization of FLM applies. In that set-up, and according to Definition 2, (ρ, x) is an admissible pair if for each $\bar{m}$, reporting truthfully and then playing ρ is an equilibrium in the Bayesian game with prior distribution μ and payoff function $r(s, a) + x(\bar{m}, m, a)$ (and if the relevant incentive-compatibility inequalities are strict).

It is useful to provide a quick reminder of the FLM proof, specialized to the present set-up. Let Z be a smooth compact set in the interior of $\mathcal{H}$, and a discount factor $\delta < 1$.

Given an initial target payoff vector $v \in Z$, (and $\bar{m} \in M$), one picks an appropriately chosen direction $\lambda \in \Lambda$, and we choose an admissible contract (ρ, x) such that (ρ, x, v) is feasible in $\mathcal{P}_0(\lambda)$.[20] Players are required to report truthfully their type and to play (on path) according to ρ, and we define $w_{\bar{m},m,a} := v + \frac{1-\delta}{\delta} x(\bar{m}, m, a)$ for each $(m, a) \in M \times A$. Provided δ is large enough, the vectors $w_{\bar{m},m,a}$ belong to Z, and this construction can thus be iterated, leading to a well-defined strategy profile σ in the repeated game.[21] The expected payoff under σ is v, and the continuation payoff in step 2, conditional on public history (m, a), is equal to $w_{\bar{m},m,a}$, when computed at the *ex ante* stage, before players learn their step-2 types. The fact that (ρ, x) is admissible implies that σ yields an equilibrium in the one-shot game with payoff $(1-\delta)r(s, a) + \delta w_{\bar{m},m,a}$. A one-step deviation principle then applies, implying that σ is a sequential equilibrium of the repeated game, with payoff v.

Assume now that the type profiles (s_n) follow an irreducible Markov chain with invariant measure μ. The proof outlined above fails as soon as types as auto-correlated. Indeed, the initial type of player i now provides information over types in step 2. Hence, at the interim stage in step 1, (using the above notations) the expected continuation payoffs of player i are no longer given by $w_{\bar{m},m,a}$. This is the rationale for including the continuation private values into the definition of admissible contracts.

But this raises a difficulty. In any recursive construction such as the one outlined above, continuation private values (which help define current play) are defined by continuation play, which itself is based on current play, leading to an uninspiring circularity. On the other hand, our definition of an admissible contract (ρ, x) involves the private values $\theta_{\rho,r+x}$ induced by an indefinite play of (ρ, x). This difficulty is solved by adjusting the recursive construction in such as way that players always expect the current admissible pair (ρ, x) to be used in the foreseeable future. On the technical side, this is achieved by letting players stick to an admissible pair (ρ, x) during a random number of rounds, with a geometric distribution of parameter η. The target vector is updated only when switching to a new direction (and to a new admissible pair). The random time at which switching occurs is determined by the correlation device. The parameter η is chosen large enough compared to $1-\delta$, ensuring that target payoffs always remain within the set Z. Yet, η is chosen small enough so that the continuation private values be approximately equal to $\theta_{\rho,r+x}$: in terms of private values, it is almost as if (ρ, x) were used forever.

Equilibrium properties are derived from the observation that, by Definition 2, the incentive to report truthfully (and then to play ρ) would be strict if the continuation private

[20]If v is a boundary point, λ is an outwards pointing normal to Z at v.
[21]With $w_{\bar{m},m,a}$ serving as the target payoff vector in the next, second, step.

values were truly equal to $\theta_{\rho,r+x}$ and thus, still holds when equality holds only approximately. All the details are provided in the Appendix.

6 Independent Private Values

This section considers the special case of independent private values.

Definition 3 *The game has independent private values (IPV) if:*

- *The stage-game payoff function of i depends on his own state only: for every i and (s, a^i, y), $g^i(s, a^i, y) = g^i(s^i, a^i, y)$.*
- *The prior distribution p_1 is a product distribution: for all s,*

$$p_1(s) = \times_i p_1^i(s^i),$$

for some distributions $p_1^i \in \Delta(S^i)$.

- *The transitions of player i's state are independent of players $-i$'s private information: for every i and y, every (s^i, a^i, t^i), and pair $(s^{-i}, \alpha^{-i}, t^{-i})$, $(\tilde{s}^{-i}, \tilde{\alpha}^{-i}, \tilde{t}^{-i})$,*

$$p_{s^i,s^{-i},a^i\alpha^{-i}}(t^i \mid y, t^{-i}) = p_{s^i,\tilde{s}^{-i},a^i,\tilde{\alpha}^{-i}}(t^i \mid y, \tilde{t}^{-i}).$$

The second assumption ensures that the conditional belief of players $-i$ about player i's state only depends on the public history (independently of the play of players $-i$). Along with the third, it implies that the private states of the players are independently distributed in any round n, conditional on the public history up to that round. As is customary with IPV, this definition assumes that the factorization property holds, namely, player i's stage-game payoff only depends on a^{-i} via y, although none of the proofs uses this property.

As discussed, there is no reason to set $M^i = S^i \times A^i \times S^i$ here, and so we fix $M^i = S^i$ throughout (we nevertheless use the symbol M^i instead of S^i whenever convenient).

Our purpose is to describe explicitly the asymptotic equilibrium payoff set in the IPV case. The feasible (long-run) payoff set is defined as

$$F := \text{co}\left\{v \in \mathbf{R}^I \mid v = \mathbf{E}_{\mu_\rho}[r(s,a)],\ \text{some policy } \rho : M \to A\right\}.$$

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.

6.1 An Upper Bound on Bayes Nash Equilibrium Payoffs

Not all feasible payoffs can be Bayes Nash equilibrium payoffs, because types are private and independently distributed. As is well known, incentive compatibility restricts the set of decision rules that can be implemented in static Bayesian implementation. One can hardly expect the state of affairs to improve once transfers are further restricted to be continuation payoffs of a Markovian game. Yet to evaluate the performance of truthful equilibria, we must provide a benchmark.

To motivate this benchmark, consider first the case in which the marginal distribution over signals is independent of the states. That is, suppose for now that, for all $(s, \tilde{s}, a, y)$,

$$p_{s,a}(y) = p_{\tilde{s},a}(y),$$

so that the public signal conveys no information about the state profile, as is the case under perfect monitoring, for instance. Fix some direction $\lambda \in \Lambda$. What is the best Bayes Nash equilibrium payoff vector, if we aggregate payoffs according to the weights λ? If $\lambda^i < 0$, we would like player i to reveal his state in order to use this information against his 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. Define

$$\bar{k}(\lambda) = \max_{\rho} \mathbf{E}_{\mu_\rho} \left[\lambda \cdot r(s, a) \right],$$

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 $I(\lambda) = \emptyset$). Furthermore, let

$$V^* := \cap_{\lambda \in \Lambda} \left\{ v \in \mathbf{R}^I \mid \lambda \cdot v \leq \bar{k}(\lambda) \right\}.$$

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.

Lemma 2 *The limit set of Bayes Nash equilibrium payoffs is contained in V^*.*

Proof. Fix $\lambda \in \Lambda$. Fix also $\delta < 1$ (and recall the prior p_1 at time 1). Consider the Bayes Nash equilibrium σ of the game (with discount factor δ) 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 σ^{-i} and p_1, player i's equilibrium strategy σ^i defines a distribution over histories. Fixing σ^{-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 σ^i would specify if this had been i's report.[22] 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 σ 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 $I(\lambda)$. Maximizing this inner product over such strategies is a standard Markov decision problem, which admits a solution within the class of deterministic policies. Taking $\delta \to 1$ yields that the limit set is included in $\left\{v \in \mathbf{R}^I \mid \lambda \cdot v \leq \bar{k}(\lambda)\right\}$, and this is true for all $\lambda \in \Lambda$. ■

It is worth emphasizing that this result does not rely on the choice of any particular message space M.[23] We define

$$\rho[\lambda] \in \underset{\rho: \times_{i \in I(\lambda)} S^i \to A}{\operatorname{argmax}} \mathbf{E}_{\mu_\rho} \left[\lambda \cdot r(s,a)\right] \tag{2}$$

to be any policy that achieves this maximum, and let $\Xi := \{\rho[\lambda] : \lambda \in \Lambda\}$ denote the set of such policies.

The set V^* can be a strict subset of F, as the following example shows.

Example 4. Actions do not affect transitions. Each player $i = 1, 2$ has two states $s^i = \underline{s}^i, \bar{s}^i$, with $c(\underline{s}^i) = 2, c(\bar{s}^i) = 1$. Rewards are given by Figure 3. (The interpretation is that a pie of size 3 is obtained if at least one agent works; if both choose to work only half the

[22]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 σ^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 σ^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.

[23]Incidentally, it appears that the role of V^* is new also in the context of static mechanism design with transfers. There is no known exhaustive description of the decision rules that can be implemented under IPV, but it is clear that the payoffs in V^* (replacing μ with the prior distribution in the definition) can be achieved using the AGV mechanism on a subset of agents; conversely, no payoff vector yielding a score larger than $\bar{k}(\lambda)$ can be achieved, so that V^* provides a description of the achievable payoff set in that case as well.

	L	R
T	$3-\frac{c(s^1)}{2}, 3-\frac{c(s^2)}{2}$	$3-c(s^1), 3$
B	$3, 3-c(s^{-2})$	$0, 0$

Figure 3: Payoffs of Example 4

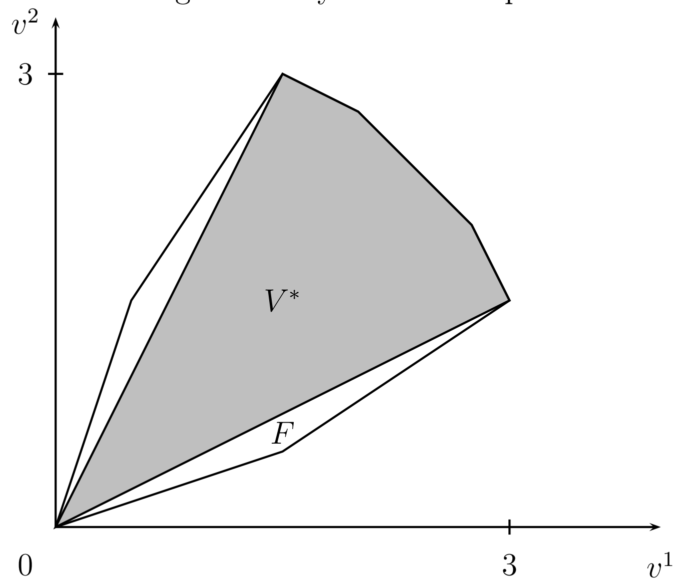

Figure 4: Incentive-compatible and feasible payoff sets in Example 4

amount of work has to be put in by each worker. Their cost of working is fluctuating.) This game satisfies the IPV assumption. From one round the next, the state changes with probability p, common but independent across players. Given that actions do not affect transitions, we can take it equal to $p = 1/2$ (i.i.d. types) for the sake of computing V^* and F, shown in Figure 4. Of course, each player can secure at least $3 - \frac{2+1}{2} = \frac{3}{2}$ by always working, so the actual equilibrium payoff set, taking into account the incentives at the action step, is smaller.[24]

So far, the distribution of public signals has been assumed to be independent of states. More information can be extracted from players when they cannot prevent public signals from revealing part of it, at least statistically. States s^i and $\tilde{s}^i$ are *indistinguishable*, denoted

[24] 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 remedies this.

$s^i \sim \tilde{s}^i$, if for all s^{-i} and all (a, y), $p_{s^i,s^{-i},a}(y) = p_{\tilde{s}^i,s^{-i},a}(y)$. Indistinguishability defines a partition of S^i and we denote by $[s^i]$ the partition cell to which s^i belongs. If signals depend on actions, this partition is non-trivial for at least one player. By definition, if $[s^i] \neq [\tilde{s}^i]$ there exists s^{-i} such that $p_{s^i,s^{-i},a} \neq p_{\tilde{s}^i,s^{-i},a}$ for some $a \in A$. Let $D^i = \{(s^{-i}, a)\} \subset S^{-i} \times A$ denote a selection of such states, along with the discriminating action profile: for all $[s^i] \neq [\tilde{s}^i]$, there exists $(s^{-i}, a) \in D^i$ such that $p_{s^i,s^{-i},a} p_{\tilde{s}^i,s^{-i},a}$.

More generally then, the best Bayes Nash equilibrium payoff in the direction $\lambda \in \Lambda$ cannot exceed

$$\bar{k}(\lambda) := \max_{\rho} \mathbf{E}_{\mu_\rho} \left[\lambda \cdot r(s, a)\right],$$

where the maximum is now over all policies $\rho : S \to A$ such that if $s^i \sim \tilde{s}^i$ and $\lambda^i \leq 0$ then $\rho(s^i, \cdot) = \rho(\tilde{s}^i, \cdot)$. Extending the definition of V^* to this more general definition of $\bar{k}$, Lemma 2 remains valid. We retain the same notation for $\rho[\lambda]$, the policies that achieve the extreme points of V^*, and Ξ, the set of such policies.

Finally, 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 $v = \mathbf{E}_{\mu_\rho}[r(s, a)]$ for some ρ that takes a constant value in A. Escobar and Toikka (2013) show that, in their environment with action-independent transitions and perfect monitoring, all individually rational (as defined below) payoffs in $\text{co}\,(Ext^{pu} \cup Ext^{po})$ are equilibrium payoffs (whenever this set has non-empty interior). Indeed, the following is easy to show.

Lemma 3 *It holds that* $\text{co}\,(Ext^{pu} \cup Ext^{po}) \subset V^*$.

In Example 4, this bound is tight, but this is not always the case.

6.2 Truth-telling

In this section, we ignore the action step and focus on the incentives of players to report their type truthfully. That is, we focus on the revelation game.

The pair (ρ, x) is *weakly truthful* if it satisfies Definition 2 with two modifications: in Step 1 of Definition 2, the requirement that truth-telling be *uniquely* optimal is dropped. That is, it is only required that truth-telling be an optimal reporting strategy, albeit not necessarily the unique one. In Step 2, the requirement that ρ^i be optimal is ignored. That is, the policy $\rho : S \to A$ is fixed.

A direction $\lambda \in \Lambda$ is *coordinate* if it is equal to e^i or $-e^i$, where e^i denotes the i-th coordinate basis vector in $\mathbf{R}^I$. The direction λ is *non-coordinate* if $\lambda \neq \pm e^i$, that is, if it has at least two nonzero coordinates. We first show that we can ignore the constraint $\lambda \cdot x \leq 0$ in all non-coordinate directions.

Proposition 1 *Let (ρ, x) be a weakly truthful pair. Fix a non-coordinate direction $\lambda \in \Lambda$. Then there exists $\hat{x}$ such that $(\rho, \hat{x})$ is weakly truthful and $\lambda \cdot \hat{x} = 0$.*

Proposition 1 implies that (exact) budget-balance comes "for free" in all non-coordinate directions. It is the undiscounted analogue of a result by Athey and Segal (2013), and its proof follows similar steps.

Proposition 1 need not hold in coordinate directions. However, we can also assume that $\lambda \cdot x(\cdot) = 0$ for $\lambda = \pm e^i$ when considering the policies $\rho[\lambda] \in \Xi$: if $\lambda = -e^i$, $\rho[\lambda]$ is an action profile that is independent of the state profile. Hence, incentives for weak truth-telling are satisfied for $x = 0$; in the case $\lambda = +e^i$, $\rho[\lambda]$ is a policy that depends on i's report only, yet it is precisely i's payoff that is maximized. Here as well, incentives for weak truth-telling are satisfied for $x = 0$.

Our next goal is to obtain a characterization of all policies ρ for which there exists x such that (ρ, x) is weakly truthful.

Along with ρ and truthful reporting by players $-i$, a reporting strategy by player i, that is, a map[25] $m^i_\rho : \Omega_{\text{pub}} \times S^i \to \Delta(M^i)$ from the previous public outcome and the current state into a report, induces a unichain Markov chain over $\Omega_{\text{pub}} \times S^i \times M^i$, with transition function q_ρ and with invariant measure $\pi^i_\rho \in \Delta(\Omega_{\text{pub}} \times S^i \times M^i)$. We define the set $\Pi^i_\rho \subset \Delta(\Omega_{\text{pub}} \times S^i \times M^i)$ as all distributions π^i_ρ that satisfy the balance equation

$$\pi^i_\rho(\omega_{\text{pub}}, t^i) = \sum_{\bar{\omega}_{\text{pub}}, s^i} q_\rho(\omega_{\text{pub}}, t^i \mid \bar{\omega}_{\text{pub}}, s^i, m^i) \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i), \text{ all } (\omega_{\text{pub}}, t^i), \tag{3}$$

and

$$\sum_{s^i \in [m^i]} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = \sum_{s^i \in [m^i]} \mu_\rho(\bar{\omega}_{\text{pub}}, m^i). \tag{4}$$

where $\mu_\rho(\bar{\omega}_{\text{pub}}, m^i)$ is the probability assigned to $(\bar{\omega}_{\text{pub}}, m^i)$ by the invariant distribution μ_ρ under truth-telling (and ρ). Equation (4) states that π^i_ρ cannot be statistically distinguished from truth-telling. As a consequence, it is not possible to prevent player i from choosing his favorite element of Π^i_ρ, as formalized by the next lemma. To state it, define

$$r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) := \mathbf{E}_{s^{-i} \mid \bar{\omega}_{\text{pub}}} \left[r^i(s^i, \rho(s^{-i}, m^i)) \right]$$

[25] Note that, under IPV, player i's private information contained in $\bar{\omega}$ is not relevant for his incentives in the current round, conditional on $\bar{\omega}_{\text{pub}}$, s^i.

as the expected reward of player i given his report, type and the previous public outcome $\bar{\omega}_{\text{pub}}$.

Lemma 4 *Given a policy ρ, there exists x such (ρ, x) is weakly truthful if and only if for all i, truth-telling maximizes*

$$\mathbf{E}_\pi \left[r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) \right] \tag{5}$$

over $\pi \in \Pi^i_\rho$.

We apply Lemma 4 to the policies that achieve the extreme points of V^*. Fix $\lambda \in \Lambda$ and $\rho = \rho[\lambda] \in \Xi$. Plainly, truth-telling is optimal for any player $i \notin I(\lambda)$, as his reports do not affect the policy. As for a player $i \in I(\lambda)$, note that if two of his reporting strategies are both in $\Pi^i_{\rho[\lambda]}$, the one that yields a higher expected payoff to him (as defined by (5)) also yields a higher score: indeed, as long as they are both in $\Pi^i_{\rho[\lambda]}$, they are equivalent from the point of view of the other players. It then follows that the maximum score over weakly truthful pairs (ρ, x) is equal to the maximum possible one, $\bar{k}(\lambda)$.

Lemma 5 *Fix a direction $\lambda \in \Lambda$. Then the maximum score over weakly truthful (ρ, x) such that $\lambda \cdot x \leq 0$ is given by $\bar{k}(\lambda)$.*

The conclusion of this section is somewhat surprising: at least in terms of payoffs, there is no possible gain (in terms of incentive-compatibility) from linking decisions (and restricting attention to truthful strategies) beyond the simple class of policies and transfer functions that we consider. In other words, *ignoring individual rationality and incentives at the action step*, the set of "equilibrium" payoffs that we obtain is equal to the set of incentive-compatible payoffs V^*. If players commit to actions, the "revelation principle" holds even if players do not commit to future reports.

If transitions are action-independent, note that this means also that the persistence of the Markov chain has no relevance for the set of payoffs that are incentive-compatible. (If actions affect transitions, even the feasible payoff set changes with persistence, as it affects the extreme policies.) Note that this does not rely on any full support assumption on the transition probabilities, although of course the unichain assumption is used (cf. Example 1 of Renault, Solan and Vieille (2013) that shows that this conclusion –that the invariant distribution is a sufficient statistic– does not hold when values are interdependent).

6.3 Obedience and Individual Rationality

Actions might be just as hard to keep track of as states. But there are well known statistical conditions under which opportunistic behavior can be kept in check when actions are imperfectly monitored. These conditions are of two kinds. First, unilateral deviations must be detectable, at least when they are profitable, so that punishments can be meted out. Second, when such deviations pertain to players that matter for budget balance, they must be identifiable, so that punishments involve surplus redistribution rather than surplus destruction. Because the signal distribution might depend on the state profile, the conditions from repeated games must be slightly amended.

In what follows, $p_{s,a}$ refers to the marginal distribution over signals $y \in Y$ only. (Because types are conditionally independent, the states of players $-i$ in round $n+1$ are uninformative about a^i, conditional on y.) Let $Q^i(s,a) := \{p_{\hat{s}^i,s^{-i},\hat{a}^i,a^{-i}} : \hat{a}^i \neq a^i, \hat{s}^i \in S^i\}$ 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. The assumption involves the extreme points (in policy space) of the relevant payoff set.

Assumption 1 *For all $\rho \in \Xi$, all s, $a = \rho(s)$; also, for all (s,a) where $(s^{-i},a) \in D^i$ for some i:*

1. *For all $i \neq j$, $p_{s,a} \notin \mathrm{co}\,(Q^i(s,a) \cup Q^j(s,a))$;*
2. *For all $i \neq j$,*
$$\mathrm{co}\left(p_{s,a} \cup Q^i(s,a)\right) \cap \mathrm{co}\left(p_{s,a} \cup Q^j(s,a)\right) = \{p_{s,a}\}\,.$$

This assumption states that deviations of players can be detected, as well as identified, even if player i has "coordinated" his deviation at the reporting and action step.

Note that Assumption **1** reduces to Assumptions **A2**–**A3** of Kandori and Matsushima (1998) in the case of repeated games.

Finally, lack of commitment curtails how low payoffs can be. Example 2 makes clear that insisting on truth-telling restricts the ability to punish players, and that the minimum equilibrium payoff in truthful strategies can be bounded above the actual minmax payoff. Nevertheless, it should be clear that this minimum is no more than the *state-independent pure-strategy* minmax payoff

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

Clearly, this is not the best punishment level one could hope for, even if it is the one used in the literature. Nevertheless, as Escobar and Toikka (2013) eloquently describe, it coincides with the actual minmax payoff (defined over all strategies available to players $-i$, see the next section) in many interesting economic examples. It does in Example 4 as well, but not in Example 2. The punishment level $-k_0(-e^i)$ delivered by the optimization program $\mathcal{P}(-e^i)$ can be strictly lower than this state-independent pure-strategy minmax payoff, but there seems to be no simple formula for it. Hence, we use $\underline{v}^i$ in what follows.

To ensure that players $-i$ are willing to punish player i, we must also impose an assumption on the signal distribution when players follow this policy. This is the counterpart of Assumption **A1** of Kandori and Matsushima (1998). Let $\underline{\rho}_i$ denote a policy that achieves $\underline{v}^i$.

Assumption 2 *For all i, for all s, $a = \underline{\rho}_i(s)$, $j \neq i$,*

$$p_{s,a} \notin \operatorname{co} Q^j(s,a).$$

Finally, we must impose a "non-degeneracy" assumption to avoid that incentives to tell the truth can never be made strict. Given $\lambda \in \Lambda$, let $\rho[\lambda]$ be a policy that maximizes $\mathbf{E}_{\mu_\rho}[\lambda \cdot r(s,a)]$ over all policies $\rho : S \to \Delta(A)$.

Assumption 3 *There exists $\lambda \in \Lambda$ such that, for all i, s^i, there are $s^{-i}, \tilde{s}^{-i}$ with $\rho[\lambda](s^i, s^{-i}) \neq \rho[\lambda](s^i, \tilde{s}^{-i})$.*

We may now state the main result of this section. Denote the set of incentive-compatible, individually rational payoffs as

$$V^{**} := \left\{ v \in V^* \mid v^i \geq \underline{v}^i, \text{ all } i \right\}.$$

Theorem 3 *Suppose that V^{**} has non-empty interior. Under Assumptions **1**–**3**, the limit set of equilibrium payoffs includes V^{**}.*

6.4 A Characterization

The previous section has provided lower bounds on the asymptotic equilibrium payoff set. This section provides an exact characterization under stronger assumptions.

As mentioned, there are many examples in which the state-independent pure-strategy minmax payoff $\underline{v}^i$ coincides with the "true" minmax payoff

$$w^i := \lim_{\delta \to 1} \min_{\sigma^{-i}} \max_{\sigma^i} \mathbf{E}\left[(1-\delta) \sum \delta^{n-1} r_n^i \right],$$

where the minimum is over the set of (independent) strategies by players $-i$. We denote by $\underline{\sigma}_i$ 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. For instance, mixed strategies play a key role in the literature on tax auditing. More disturbingly, when $\underline{v}^i > w^i$, it can happen that $V^{**} = \emptyset$. Theorem 3 becomes meaningless, as the corresponding equilibria no longer exist. On the other hand, the set

$$W := \left\{ v \in V^* \mid v^i \geq w^i \text{ for all } i \right\}$$

is never empty.[26]

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 $\underline{v}^i$), 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. One special case in which they do coincide is when monitoring has a product structure, as players $j \neq i$'s private histories do not allow them to correlate play unbeknownst to player i. As this is the class of monitoring structures for which the traditional folk theorem for repeated games is actually a characterization (rather than a lower bound) of the equilibrium payoff set, we maintain this assumption throughout this section.[27]

Definition 4 *Monitoring has product structure if there are finite sets* $(Y^i)_{i=1}^I$ *such that* $Y = \times_i Y^i$, *and*

$$p_{s,a}(y) = \times_i p^i_{s^i,a^i}(y^i),$$

for all $y = (y^1, \ldots, y^I) \in Y$.

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 that can certainly be refined.

Assumption 4 *For all* i, (s, a),

$$p_{s,a} \notin \operatorname{co} Q^i(s, a).$$

[26]To see this, note that even the state-independent mixed minmax payoff lies below the Pareto-frontier: clearly, the score in direction $\lambda^e = \frac{1}{\sqrt{I}}(1, \ldots, 1)$ of the payoff vector $\min_{\alpha^{-i}} \max_{\rho^i : S^i \to A^i} \mathbf{E}[r^i(s^i, a)]$ is less than $k(\lambda^e)$.

[27]The scope for w^i to coincide with the mixed minmax payoff is slightly larger, but not by much. See Gossner and Hörner (2010).

Note that, given product structure, Assumption **4** is an assumption on p^i only. We prove that W characterizes the (Bayes Nash, as well as sequential) equilibrium payoff set as $\delta \to 1$ in the IPV case. More formally:

Theorem 4 *Assume that monitoring has the product structure, and that Assumptions 3–4 hold. If W has non-empty interior, the set of (Nash, sequential) equilibrium payoffs converges to W as $\delta \to 1$.*

As is clear from Example 2, this requires using strategies that are not truthful, at least during "punishments."[28] Nonetheless, we show that a slight weakening of the characterization obtained so far, to allow for silent play during punishment-like phases, suffices for our purpose.

Our construction relies on an extension of Theorem 2, as well as an argument inspired by Gossner (1995), based on approachability theory. 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.

Loosely speaking, we may think of the zero-sum game that defines the minmax strategy profile $\underline{\sigma}^i$ as follows. There is no communication. Player i chooses a (without loss, Markov) strategy $\alpha^i : S^i \times \Delta(S^i) \to \Delta(A^i)$, mapping his type and the public belief about his type into a (mixed) action. This strategy must be a best-reply to the (Markov) strategy $\alpha^{-i} : \Delta(S^i) \to \times_{j \neq i} \Delta(A^j)$ used by his opponents –a function of their belief about i's type. (Here, we use α rather than $\underline{\sigma}_i$ to emphasize this Markovian structure.) This belief is correct given the optimal strategy α^i. The history of signals (y^i_n) along with the equilibrium α^i determines how this belief is updated. Player i's choice of an arbitrary Markov strategy $\tilde{\alpha}^i$ then defines

[28] 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.

a Markov chain over $S^i \times \Delta(S^i)$, whose payoff can be viewed as the expectation relative to the occupation measure $\mu[\tilde{\alpha}^i]$ over pairs $(s^i, q^i) \in S^i \times \Delta(S^i)$ that it defines. In particular, the (asymptotically) optimal strategy α^i maximizes

$$\int_{(s^i,q^i)\in S^i\times\Delta(S^i)} r^i(s^i, \alpha^{-i}(q^i))d\mu[\tilde{\alpha}^i](s^i, q^i)$$

over $\tilde{\alpha}^i$. In other words: through the actions he picks, player i controls the distribution over action profiles α^{-i} that his opponents employ. The optimal strategy of i maximizes his payoff relative to this distribution. Considering a sufficiently fine discretization of the simplex $\times_{j\neq i}\Delta(A^j)$, we may now partition the realized public histories $h^1, h^2, \ldots, h^T$, according to the partition cell which the action profile α^{-i} belongs to. This partition only depends on the public history concerning player i, as this is a sufficient statistic for the public belief about his state. For each player $j \neq i$, each partition cell, and each signal profile y^{-j} that is observed sufficiently often along the realized path, we may then compare the empirical frequency of y^j with the theoretical one if j used α^j. In this fashion, we define a test that each player j can pass with very high probability (as this is the case if he uses α^j), independently of the strategies used by players $-j$, and that, by definition, he is likely to fail if he uses a strategy inducing another distribution. Penalizing player j sufficiently for failing the test (via x), we ensure that all players $j \neq i$ must pass the test with high probability in any equilibrium of the (truncated but long enough) game with transfers, which effectively means that player i is punished.

7 Correlated Types

We now consider the case of correlated types, as defined by Assumption **6** below. As we will see, applying Theorem 2 results in an extension of the static insights from Crémer and McLean (1988) to the dynamic game.

As in the IPV case, we must distinguish truth-telling incentives from constraints imposed by individual rationality and imperfect monitoring of actions. Here, we start with the latter. Because V^* is no longer an upper bound on the Bayes Nash equilibrium payoff set, we must re-define the set of relevant policies Ξ as the set of policies that achieve extreme points of F.[29,30] These are simply the policies achieving the extreme points of the feasible (limit)

[29]If multiple policies achieve the maximum, Assumption **1**' has to be understood as asserting the existence of a selection of policies satisfying the stated requirement.

[30]To economize on new notation, in what follows we adopt the symbols used in Section 6 to denote the corresponding –although slightly different– quantities. Hopefully, no confusion will arise.

payoff set.

As before, in the statement of assumptions on the monitoring structure, $p_{s,a}$ refers to the marginal distribution over public signals only.

Assumption 1' *For all $\rho \in \Xi$, all s, $a = \rho(s)$:*

1. *For all $i \neq j$, $p_{s,a} \notin \mathrm{co}\,(Q^i(s,a) \cup Q^j(s,a))$;*

2. *For all $i \neq j$,*
$$\mathrm{co}\left(p_{s,a} \cup Q^i(s,a)\right) \cap \mathrm{co}\left(p_{s,a} \cup Q^j(s,a)\right) = \{p_{s,a}\}.$$

Because the private states of players $-i$ are no longer irrelevant when punishing player i (both because values need not be private, and because their states are informative about i's state), we must redefine the minmax payoff of player i as

$$\underline{v}^i := \min_{\rho^{-i}:S^{-i}\to A^{-i}} \max_{\rho^i:S\to A^i} \mathbf{E}_{\mu_\rho}[r^i(s,a)],$$

As before, we let $\underline{\rho}_i$ denote a policy that achieves this minmax payoff.

Assumption 2' *For all i, for all s, $a = \underline{\rho}_i(s)$, $j \neq i$,*

$$p_{s,a} \notin \mathrm{co}\, Q^j(s,a).$$

The purpose of these two assumptions is as in the IPV case: it ensures that transfers that induce truth-telling taking as given compliance with a fixed policy can always be augmented in a budget-balanced fashion so as to ensure that this compliance is optimal, whether or not a player deviates in the report he makes: with such an adjustment, even after an incorrect report (at least in non-coordinate directions), a player finds it optimal to play as if his report had been truthful. This is formally stated below.

Lemma 6 *Under Assumptions **1'–2'**, it holds that:*

- *For all non-coordinate λ, there exists $x : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ such that (i) $\lambda \cdot x(\cdot) = 0$, (ii) for all i, if players $-i$ report truthfully and play according to $\rho^{-i}[\lambda]$, then all best-replies of i at the action step specify $a^i = \rho^i[\lambda](m)$ independently of m^i.*

- *Given $\lambda = +e^i$, there exists $x : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ such that (i) $\lambda \cdot x(\cdot) = 0$, (ii) for all $j \neq i$, if players $-j$ report truthfully and play according to $\rho^{-j}[\lambda]$, then all best-replies of j at the action step specify $a^j = \rho^j[\lambda](m)$ independently of m^j,(iii) if players $-i$ report truthfully and play according to $\rho^{-i}[\lambda]$, $\rho^i[\lambda]$ is a best-reply for player i after a truthful report m^i. The same conclusions hold for $\lambda = -e^i$ and $\rho = \underline{\rho}_i$.*

We now turn to the players' incentives to report truthfully. For simplicity, we assume that the states are autocorrelated. More precisely, this is implied by Assumptions **5–6** below, which cannot hold otherwise. For the case in which states are independently distributed over time, the counterpart of Theorem 5 follows from a straightforward application of FL. To save on notation, given Lemma 6, in what follows we drop player i's previous action $\bar{a}^i$ from his report.

Throughout, fix some policy $\rho : S \to A$ and assume that actions are determined by ρ (that is, we take actions as given). Fix $\bar{m}$, $\bar{y}$, $\bar{a}$ and a player i. Having fixed actions, recall that a type of player i is a pair $\zeta^i = (\bar{s}^i, s^i)$. What evidence can be used to statistically test whether player i is reporting truthfully his type? The states s^{-i} that are announced, first;[31] the signal y (as the distribution of signals can depend on s^i) second; and last, as explained in Example 3, the next report t^{-i}.

We may use Bayes' rule to compute the distribution over (t^{-i}, s^{-i}, y), conditional on the past reports, actions and signal being $\bar{m}, \bar{a}, \bar{y}$ if player i's past and current state are $\bar{s}^i$ and s^i. This distribution is denoted

$$q_{-i}^{\bar{m},\bar{y},\bar{a}}(s^{-i}, y, t^{-i} \mid \zeta^i).$$

Detecting deviations requires that different reports induce different distributions. We must distinguish between directions $\lambda = -e^i$ and other directions. In directions $-e^i$, budget balance does not restrict the transfers that can be used to discipline players $j \neq i$, so that detection is all that is needed. We assume

Assumption 5 *For all i, $\rho = \underline{\rho}_i$, all $(\bar{m}, \bar{y}, \bar{a})$, for any $j \neq i$, $\hat{\zeta}^j \in (S^j)^2$, it holds that*

$$q_{-j}^{\bar{m},\bar{y},\bar{a}}(s^{-j}, y, t^{-j} \mid \hat{\zeta}^j) \neq \text{co}\left(q_{-j}^{\bar{m},\bar{y},\bar{a}}(s^{-j}, y, t^{-j} \mid \zeta^j) : \zeta^j \neq \hat{\zeta}^j \right).$$

If types are independent over time, and signals y do not depend on states (as is the case with perfect monitoring, for instance), this reduces to the requirement that the matrix with entries $p_{s^j}(s^{-j})$ have full row rank, a standard condition in mechanism design (see d'Aspremont, Crémer and Gérard-Varet (2003) and d'Aspremont and Gérard-Varet (1982)'s condition B). Here, beliefs can also depend on player j's previous state, $\bar{s}^j$, but fortunately, we can also use player $-j$'s future state profile, t^{-j}, to statistically distinguish player j's types.

[31] Of course, players $-i$'s reports are richer, as they are pairs $(\bar{s}^{-i}, s^{-i})$ themselves. But the information contained in $\bar{s}^{-i}$ is not useful in testing i's report, because player i already knows $\bar{s}^{-i}$, assuming that $-i$ have reported truthfully their states in the previous round.

As is well known, Assumption **5** ensures that for any minmaxing policy $\underline{\rho}_i$, truth-telling is Bayesian incentive compatible: there exists transfers $x^j(\bar{\omega}_{\text{pub}}, (m, y), t^{-j})$ for which truth-telling is optimal for $j \neq i$. This also holds for player i, as his report has no consequence on the actions played by the other players, and he is playing his (dynamic) best-reply.

In non-coordinate directions, statistical detection must be combined with budget balance, which requires statistical discrimination. As is standard, it is sufficient to consider pairwise directions (that is, weights $\lambda \in \Lambda$ for which two entries are non-zero), or, to put it differently, pairs of players i, j.

Stating the assumption requires some more notation.[32] We start with the joint distribution

$$q^{\bar{m},\bar{y},\bar{a}}(\zeta, y, t),$$

over triples (ζ, y, t), computed using Bayes rule under the assumption that $\bar{m}$ was truthful. Next, we must consider the distribution over such triples when player i uses some arbitrary reporting strategy when announcing his type $\zeta^i = (\bar{s}^i, s^i)$. Such a strategy is a map from $(S^i)^2$ into $(S^i)^2$, which can be represented by non-negative numbers $c^i = \left(c^i_{\zeta^i \hat{\zeta}^i}\right)$, with $\sum_{\hat{\zeta}^i} c^i_{\zeta^i \hat{\zeta}^i} = 1$ for all ζ^i. The interpretation is that $c^i_{\zeta^i \hat{\zeta}^i}$ is the probability with which $\hat{\zeta}^i$ is reported when player i's type is ζ^i. Truth-telling obtains under a particular reporting strategy, denoted $\hat{c}^i$: namely, for all ζ^i, $c^i_{\zeta^i \zeta^i} = 1$.

Given the prior distribution $q^{\bar{m},\bar{y},\bar{a}}$, a profile $c = (c^i)_{i \in I}$, defines a new distribution $\pi^{\bar{m},\bar{y},\bar{a}}$ over (ζ, y, t), according to

$$\pi^{\bar{m},\bar{y},\bar{a}}(\hat{\zeta}, y, t \mid c) = \sum_{\zeta} q^{\bar{m},\bar{y},\bar{a}}(\zeta, y, t) \times_j c^j_{\zeta^j \hat{\zeta}^j}.$$

Under truth-telling, this distribution $\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c})$ coincides with q. Of interest is the set of distributions that player i can induce by unilateral deviations in his report. This set is

$$\mathcal{R}^i(\bar{m}, \bar{y}, \bar{a}) := \left\{\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid c^i, \hat{c}^{-i}) : c^i \neq \hat{c}^i\right\}.$$

Again, the following is the adaptation of the assumption of Kandori and Matsushima (1998) to the current context.

Assumption 6 *For all $\rho \in \Xi$, all $(\bar{m}, \bar{y}, \bar{a})$,*

1. *For all pairs (i, j), $i \neq j$, $\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c}) \notin \text{co}\left(\mathcal{R}^i(\bar{m}, \bar{y}, \bar{a}) \cup \mathcal{R}^j(\bar{m}, \bar{y}, \bar{a})\right)$;*

[32] Some of the notation follows Kosenok and Severinov (2008).

2. For all (i,j), $i \neq j$,

$$\operatorname{co}\left(\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c}) \cup \mathcal{R}^i(\bar{m},\bar{y},\bar{a})\right) \cap \operatorname{co}\left(\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c}) \cup \mathcal{R}^j(\bar{m},\bar{y},\bar{a})\right) = \left\{\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c})\right\}.$$

Assumption **6** combines two assumptions: any deviation by a player is detectable ($\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c}) \notin \operatorname{co}\mathcal{R}^i(\bar{m},\bar{y},\bar{a})$), and unilateral deviations by two players are distinguishable (this is Assumption **6.2**). This second part is equivalent to the assumption of weak identifiability in Kosenok and Severinov (2008) for two players (whose Lemma 2 can be directly applied). The reason it is required for any pair of players (unlike in Kosenok and Severinov) is that we must obtain budget-balance also for vectors $\lambda \in \Lambda$ with only two non-zero positive coordinates (a stronger requirement than with more nonzero positive coordinates, as it restricts the set of players that can absorb a deficit or a surplus). The full strength of Assumption **6.1** is required (as in Kandori and Matsushima in their context) because we must also consider directions $\lambda \in \Lambda$ with only two non-zero coordinates whose signs are opposite.[33]

We let

$$V^{**} := \left\{v \in F \mid v^i \geq \underline{v}^i, \text{ all } i\right\}$$

denote the feasible and "individually rational" payoff set. It is then routine to show:

Theorem 5 *Assume that V^{**} has non-empty interior. Under Assumptions **1'–2'**, **5–6**, the limit set of truthful equilibrium payoffs includes V^{**}.*

As in the static case, Assumptions **5–6** are generically satisfied if $|S^{-i}| \geq |S^i|$ for all i.[34] Recall that, if these assumptions fail, it might be useful to take into account future observations. Future signals (reports by other players, in particular) are useful in statistically identifying the current state. Example 3 illustrates how powerful this channel can be.

[33] See also Hörner, Takahashi and Vieille (2012). One easy way to understand this is in terms of the cone spanned by the vectors $\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid c^i, \hat{c}^{-i})$ and pointed at $\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c})$. The first assumption is equivalent to any two such cones only intersecting at 0; and the second one states that any cone intersected with the opposite cone (of another player) also only intersect at 0. When $\lambda^i > 0 > \lambda^j$, we can rewrite the constraint $\lambda x^i + \lambda^j x^j = 0$ as $\lambda^i x^i + (-\lambda^j)(-x^j) = 0$ and the expected transfer of a player as $p(\cdot \mid c^j)x^j(\cdot) = (-p(\cdot \mid c^j))(-x^j(\cdot))$, so the condition for (λ^i, λ^j) is equivalent to the condition for $(\lambda^i, -\lambda^j)$ if one "replaces" the vectors $p(\cdot \mid c^j)$ with $-p(\cdot \mid c^j)$.

[34] Generically, for Assumption **5**, it suffices that $|S^{-i}|^2 \geq |S^i|^2$ for all i, while Assumption **6** calls for $|S^i \times S^{-i}|^2 \geq |S^i|^2 + |S^j|^2 - 1$ for all pairs (i,j), which is satisfied if $|S^i \times S^j|^2 \geq |S^i|^2 + |S^j|^2 - 1$, that is, $(|S^i|^2 - 1) \times |S^j|^2 \geq |S^i|^2 - 1$.

8 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.

For the sake of concision, there are several extensions that we have not examined. As mentioned, one can apply our results to the case of interdependent but observed values by enlarging the state and message space to include the last realized payoff. More generally, this can be applied to any type of private signals, repeated games with private monitoring being a special case.

Open questions remain. As explained, the payoff set identified in Theorem 2 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 without communication.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990). "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, **58**, 1041–1063.

Arrow, K. (1979). "The Property Rights Doctrine and Demand Revelation Under Incomplete Information," in M. Boskin, ed., *Economics and human welfare*. New York: Academic Press.

d'Aspremont, C. and L.-A. Gérard-Varet (1979). "Incentives and Incomplete Information," *Journal of Public Economics*, **11**, 25–45.

d'Aspremont, C. and L.-A. Gérard-Varet (1982). "Bayesian incentive compatible beliefs," *Journal of Mathematical Economics*, **10**, 83–103.

d'Aspremont, C., J. Crémer and L.-A. Gérard-Varet (2003). "Correlation, Independence, and Bayesian incentives," *Social Choice and Welfare*, **21**, 281–310.

Athey, S. and K. Bagwell (2001). "Optimal Collusion with Private Information," *RAND Journal of Economics*, **32**, 428–465.

Athey, S. and K. Bagwell (2008). "Collusion with Persistent Cost Shocks," *Econometrica*, **76**, 493–540.

Athey, S. and I. Segal (2013). "An Efficient Dynamic Mechanism," *Econometrica*, forthcoming.

Aumann, R.J. and M. Maschler (1995). *Repeated Games with Incomplete Information*. Cambridge, MA: MIT Press.

Barron, D. (2012). "Attaining Efficiency with Imperfect Public Monitoring and Markov Adverse Selection," working paper, M.I.T.

Bester, H. and R. Strausz (2000). "Contracting with Imperfect Commitment and the Revelation Principle: The Multi-Agent Case," *Economics Letters*, **69**, 165–171.

Bester, H. and R. Strausz (2001). "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, **69**, 1077–1088.

Bewley, T. and E. Kohlberg (1976). "The asymptotic theory of stochastic games," *Mathematics of Operations Research*, **3**, 104–125.

Blackwell, D. and L. Koopmans (1957). "On the Identifiability Problem for Functions of Finite Markov Chains," *Annals of Mathematical Statistics*, **28**, 1011–1015.

Cole, H.L. and N.R. Kocherlakota (2001). "Dynamic games with hidden actions and hidden states," *Journal of Economic Theory*, **98**, 114–126.

Crémer, J. and R. McLean (1988). "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, **56**, 1247–1257.

Dharmadhikari, S. W. (1963). "Sufficient Conditions for a Stationary Process to be a Function of a Finite Markov Chain," *The Annals of Mathematical Statistics*, **34**, 1033–1041.

Doepke, M. and R.M. Townsend (2006). "Dynamic mechanism design with hidden income and hidden actions," *Journal of Economic Theory*, **126**, 235–285.

Escobar, P. and J. Toikka (2013). "Efficiency in Games with Markovian Private Information," *Econometrica*, forthcoming.

Fang, H. and P. Norman (2006). "To Bundle or Not To Bundle," *RAND Journal of Economics*, **37**, 946–963.

Fernandes, A. and C. Phelan (1999). "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, **91**, 223–247.

Freixas, X., R. Guesnerie and J. Tirole (1985). "Planning under Incomplete Information and the Ratchet Effect," *Review of Economic Studies*, 173–191.

Fudenberg, D. and D. Levine (1994). "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, **62**, 103–135.

Fudenberg, D., D. Levine, and E. Maskin (1994). "The Folk Theorem with Imperfect Public Information," *Econometrica*, **62**, 997–1040.

Fudenberg, D. and Y. Yamamoto (2012). "The Folk Theorem for Irreducible Stochastic Games with Imperfect Public Monitoring," *Journal of Economic Theory*, **146**, 1664–1683.

Gilbert, E.J. (1959). "On the Identifiability Problem for Functions of Finite Markov Chains," *Annals of Mathematical Statistics*, **30**, 688–697.

Gossner, O. (1995). "The Folk Theorem for Finitely Repeated Games with Mixed Strategies," *International Journal of Game Theory*, **24**, 95–107.

Gossner, O. and J. Hörner (2010). “When is the lowest equilibrium payoff in a repeated game equal to the minmax payoff?” *Journal of economic theory*, **145**, 63–84.

Green, E.J. (1987). “Lending and the Smoothing of Uninsurable Income,” in Prescott E. and Wallace N., Eds., *Contractual Agreements for Intertemporal Trade*, University of Minnesota Press.

Jackson, M.O. and H.F. Sonnenschein (2007). “Overcoming Incentive Constraints by Linking Decision,” *Econometrica*, **75**, 241–258.

Hörner, J., D. Rosenberg, E. Solan and N. Vieille (2010). “On a Markov Game with One-Sided Incomplete Information,” *Operations Research*, **58**, 1107–1115.

Hörner, J., T. Sugaya, S. Takahashi and N. Vieille (2011). “Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \to 1$ and a Folk Theorem,” *Econometrica*, **79**, 1277–1318.

Hörner, J., S. Takahashi and N. Vieille (2013). “On the Limit Perfect Public Equilibrium Payoff Set in Repeated and Stochastic Games,” Cowles Discussion paper, Yale University.

Iosifescu, M. (1980). *Finite Markov Processes and Their Applications*, Wiley: Chichester, NY.

Kandori, M. (2003). “Randomization, Communication, and Efficiency in Repeated Games with Imperfect Public Information,” *Econometrica*, **74**, 213–233.

Kandori, M. and H. Matsushima (1998). “Private Observation, Communication and Collusion,” *Econometrica*, **66**, 627–652.

Kosenok, G. and S. Severinov (2008). “Individually Rational, Budget-Balanced Mechanisms and Allocation of Surplus,” *Journal of Economic Theory*, **140**, 126–161.

Mezzetti, C. (2004). “Mechanism Design with Interdependent Valuations: Efficiency,” *Econometrica*, **72**, 1617–1626.

Mezzetti, C. (2007). “Mechanism Design with Interdependent Valuations: Surplus Extraction,” *Economic Theory*, **31**, 473–488.

Myerson, R. (1986). “Multistage Games with Communication,” *Econometrica*, **54**, 323–358.

Neyman, A. (2008). "Existence of optimal strategies in Markov games with incomplete information," *International Journal of Game Theory*, **37**, 581–596.

Obara, I. (2008). "The Full Surplus Extraction Theorem with Hidden Actions," *The B.E. Journal of Theoretical Economics*, **8**, 1–8.

Pęski, M. and T. Wiseman (2013). "A Folk Theorem for Stochastic Games with Infrequent State Changes," working paper, University of Toronto.

Puterman, M.L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley: New York, NY.

Radner, R. (1986). "Repeated Partnership Games with Imperfect Monitoring and No Discounting," *Review of Economic Studies*, **53**, 43–57.

Renault, J. (2006). "The Value of Markov Chain Games with Lack of Information on One Side," *Mathematics of Operations Research*, **31**, 490–512.

Renault, J., E. Solan and N. Vieille (2013). "Dynamic Sender-Receiver Games," *Journal of Economic Theory*, **148**, 502–534.

Shapley, L.S. (1953). "Stochastic Games," *Proceedings of the National Academy of Sciences of the U.S.A.*, **39**, 1095–1100.

Wang, C. (1995). "Dynamic Insurance with Private Information and Balanced Budgets," *Review of Economic Studies*, **62**, 577–595.

Wiseman, T. (2008). "Reputation and Impermanent Types," *Games and Economic Behavior*, **62**, 190–210.

Zhang, Y. (2009). "Dynamic Contracting, Persistent Shocks and Optimal Taxation," *Journal of Economic Theory*, **144**, 635–675.

A Proof of Theorem 2

The proof is inspired by FLM but there are complications arising from incomplete information. We let Z be a compact set included in the interior of $\mathcal{H}_0$, and pick $\eta > 0$ small enough so that the η-neighborhood $Z_\eta := \{z \in \mathbf{R}^I, d(z, Z) \leq \eta\}$ is also contained in the interior of $\mathcal{H}_0$. We will prove that Z_η is included in the set of sequential equilibrium payoffs, for δ large enough.

A.1 Preliminaries

Given $\lambda \in \Lambda$, and since Z_η is contained in the interior of $\mathcal{H}_0$, one has $\max_{z \in Z_\eta} \lambda \cdot z < k(\lambda)$. Thus, one can find $v \in \mathbf{R}^I$, and $(\rho, x) \in \mathcal{C}_0$ such that $\max_{z \in Z_\eta} \lambda \cdot z < \lambda \cdot v$ and $\lambda \cdot x(\cdot) < 0$. Using the compactness of Λ, this proves Lemma 7 below.

Lemma 7 *There exists $\varepsilon_0 > 0$ and a finite set $\mathcal{S}_0$ of triples (v, ρ, x) with $v \in \mathbf{R}^I$ and $(\rho, x) \in \mathcal{C}_0$ such that the following holds. For every direction $\lambda \in \Lambda$, there is $(v, \rho, x) \in \mathcal{S}_0$ such that (v, ρ, x) is feasible in $\mathcal{P}_0(\lambda)$ and $\max_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.*

We let $\kappa_0 \in \mathbf{R}$ be a common bound on all elements of $\mathcal{S}_0$. Specifically, we assume that κ_0 is large enough so that $\|r\| \leq \kappa_0$, $\|x\| \leq \kappa_0/2$, $\|\theta_{\rho,r+x}\| \leq \kappa_0/3$ and $\|z - v\| \leq \kappa_0/2$ for each $(v, x, \rho) \in \mathcal{S}_0$ and every $z \in Z_\eta$.[35]

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

Lemma 8 *Given $\varepsilon > 0$, there exists $\bar{\zeta} > 0$ such that the following holds. For every $z \in Z_\eta$ there exists a direction $\lambda \in \Lambda$ such that if $w \in \mathbf{R}^I$ satisfies $\|w - z\| \leq \zeta$ and $\lambda \cdot w \leq \lambda \cdot z - \varepsilon\zeta$ for some $\zeta < \bar{\zeta}$, then $w \in Z_\eta$.*

We let $\varepsilon_1 < \varepsilon_0$ be arbitrary, and let $\bar{\zeta}$ be given by Lemma 8 applied with $\varepsilon := \varepsilon_1/\kappa_0$.

Next, we let $\beta < 1$ be arbitrary, and we let $\bar{\delta} < 1$ be large enough so that for $\delta \geq \bar{\delta}$ (i) $(1-\delta)^{1-\beta} \leq \delta$, (ii) $2\kappa_0\delta(1-\delta)^\beta \leq \varepsilon_0 - \varepsilon_1$ (iii) $\dfrac{(1-\delta)^\beta}{1-(1-\delta)^\beta} \leq \dfrac{\eta}{5\kappa_0}$.

Let $(\rho, x) \in \mathcal{C}_0$ be given. Fix a player $i \in I$ and a triple $(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i) \in \Omega_{\text{pub}} \times S^i \times A^i$.

Given $s^i \in S^i$, denote by $\gamma^i(\bar{\omega}_{\text{pub}}, (\bar{s}^i, \bar{a}^i, s^i) \to m^i)$ the highest expected payoff of player i in the decision problem $D^i(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i)$, when the type of player i is s^i and when reporting $m^i \in M^i$.

Since $(\rho, x) \in \mathcal{C}_0$, there is a belief of player i following $(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i, s^i)$ (deduced from ρ^{-i} wherever possible) such that truth-telling is uniquely optimal. Formally, there exists $\eta_{\rho,x} > 0$ such that

$$\eta_{\rho,x} + \gamma^i(\bar{\omega}_{\text{pub}}, (\bar{s}^i, \bar{a}^i, s^i) \to m^i) < \gamma^i(\bar{\omega}_{\text{pub}}, (\bar{s}^i, \bar{a}^i, s^i) \to (\bar{s}^i, \bar{a}^i, s^i)) \tag{6}$$

whenever $m^i \neq (\bar{s}^i, \bar{a}^i, s^i)$.

[35] The unit sphere is endowed with the L_1-norm. All other norms are sup norms.

In (6), the payoff γ^i is given by the expectation of

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

On both sides, s^{-i} is drawn according to the belief of player i. The action profile a is then drawn according to $\rho(s)$ on the right-hand side, and according to $(\tilde{a}^i, \rho^{-i}(m^i_c, s^{-i}))$ on the left-hand side, where $\tilde{a}^i$ may depend on s^i, m^i, s^{-i} and $\bar{\omega}_{\text{pub}}$. Next, y and t are drawn according to p.

We set $\eta := \min_{(v,\rho,x)\in\mathcal{S}_0} \eta_{\rho,x} > 0$.

A.2 Strategies

We let the initial state profile be commonly known and equal to $s_1 \in S$.

We let a payoff vector $z_* \in Z_\eta$, and a discount factor $\delta \geq \bar{\delta}$ be given. We here define a strategy profile σ with a payoff equal to z_* in the δ-discounted game, which we next show to be a sequential equilibrium.

Under σ, players always report truthfully, even if a deviation from σ was observed in the past. The play is partitioned into blocks of random duration. The durations of the successive blocks are iid random variables, and follow a geometric distribution with parameter $(1-\delta)^\beta$. The random decision to start a new block is made by the public randomizing device. Specifically, in each round n, the device determines whether to start a new block or not, with respective probabilities $\xi := (1-\delta)^\beta$ and $1-\xi$.

In each round n of a given block k, each player i uses a fixed policy to select his action, whenever he reported truthfully his current state ($m^i_{n,c} = s^i_n$). (Actions following lies will be specified below). This policy $\rho[k] : S \to \times_{i\in I}\Delta(A^i)$ is determined in the first round τ_k of block k, based on the available public history, including reports submitted in round τ_k.

The updating process is reminiscent of that of FLM, but somewhat more cumbersome. It is best described by introducing two "target" payoffs $w[k], z[k]$ (instead of one in FLM and HSTV), with $w[k] \in S_\eta$ for all k, as well as the triple $((\rho[k], x[k], v[k]) \in \mathcal{S}_0$ "used" in block k.

Given the values on block k, the target $w[k+1]$ is defined in the first round $n+1 = \tau_{k+1}$ of block $k+1$ by

$$\tilde{w}_{n+1} = \xi w[k+1] + (1-\xi) z[k], \tag{7}$$

where

$$\tilde{w}_{n+1} := \frac{1}{\delta} z[k] - \frac{1-\delta}{\delta} v[k] + \frac{1-\delta}{\delta} x[k](\omega_{\text{pub},n-1}, \omega_{\text{pub},n}, m_{c,n+1}). \tag{8}$$

Then, we let $\lambda[k] \in \Lambda$ be the direction associated to $w[k]$ by Lemma 8. Next, we let $(\rho[k], x[k], v[k]) \in \mathcal{S}_0$ be the triple associated to $\lambda[k]$ by Lemma 7, and we finally set

$$z[k+1] := w[k+1] + (1-\delta)\left(\left(1+\frac{1-\delta}{\delta\xi}\right)\theta[k](\omega_{\text{pub},n}, m_{n+1,c}) - \theta[k+1](\omega_{\text{pub},n}, m_{n+1,c})\right), \tag{9}$$

where $\theta[k] := \theta_{\rho[k],r+x[k]}$ and $\theta[k+1] := \theta_{\rho[k+1],r+x[k+1]}$.

We initialize these sequences by setting $w[1] = z_*$, and $\theta[0] = 0$. That is, $\lambda[1]$ is first obtained from Lemma 8, next $(\rho[1], x[1], v[1])$ from Lemma 7, and $z[1]$ is finally set to $z[1] := w[1] - (1-\delta)\theta[1](\omega_{\text{pub},0}, m_1)$, where $\omega_{\text{pub},0}$ is any element of Ω_{pub} consistent with $\rho[1]$, and m_1 is consistent with $\omega_{\text{pub},0}$ and s_1.

In FLM, the target payoff z is updated every round. In HSTV, it is updated every n rounds with $n > 1$, to account for changing states. Here instead, the target payoff is updated at random times. The fact that ξ is much larger than $1-\delta$ ensures that successive target payoffs lie in Z_η. The fact that ξ vanishes as $\delta \to 1$ ensures that the expected duration of a block increases to $+\infty$ as $\delta \to 1$.

That the recursive construction is well-defined follows from Lemma 9 below.

Lemma 9 *One has $w[k] \in Z_\eta$, for all k (and following any public history).*

Proof. Assume $w[k] \in Z_\eta$,[36] and $\|w[k] - z[k]\| \leq \kappa_0(1-\delta)$.

Note that by equation (7,

$$\xi(w[k+1] - z[k]) = \tilde{w}_{\tau_{k+1}} - z[k] = \frac{1-\delta}{\delta}(z[k] - v[k] + x[k])$$

so that

$$w[k+1] - w[k] = \frac{1-\delta}{\delta\xi}(z[k] - v[k] + x[k]) + (z[k] - w[k]).$$

Thus,

$$\|w[k+1] - w[k]\| \leq 2\frac{(1-\delta)}{\delta\xi}\kappa_0 = \zeta,$$

and

$$\begin{aligned}\lambda[k] \cdot (w[k+1] - w[k]) &\leq -\varepsilon_0 \times \frac{1-\delta}{\delta\xi} + (1-\delta)\kappa_0 \\ &\leq -\varepsilon\zeta.\end{aligned}$$

[36] Here and elsewhere, we view $w[k]$ as a random variable which is measurable wrt the public information available at the action step in round τ_k.

Hence the result, by the choice of $\lambda[k]$, of $\bar{\delta}$ and since $\zeta < \bar{\zeta}$. Finally, the inequality $\|z[k+1] - w[k+1]\| \leq (1-\delta)\kappa_0$ follows from (9) and the choice of κ_0 and $\bar{\delta}$. ■

It is convenient to denote by $\mathcal{T}_n = (z_n, w_n, \rho_n, x_n, v_n, \theta_n)$ the relevant values at round n. That is, we have $w_n = w[k]$ for every $\tau_k \leq n < \tau_{k+1}$, and the other variables are defined accordingly. For instance, ρ_n thus stands for the action policy in round n. We stress that $\mathcal{T}_n$ is known only when the output of the randomizing device has been observed in round n.

Fix a player i, and let h^i_n be a private history of player i up to the action choice in round n, such the current report of player i on his own type is correct in round n: $m^i_{n,c} = s^i_n$. Recall that at h^i_n, the belief of player i assigns probability one to $s^{-i}_n = m^{-i}_{n,c}$, so that the expected continuation payoff of player i under σ is well-defined, and only depends on $\omega_{\text{pub},n-1}$, m_n and on $\mathcal{T}_n$.[37] We denote it by $\gamma^i_\sigma(\omega_{\text{pub}}, m; \mathcal{T})$.

To complete the description of σ, we finally let h^i_n be a (private) history at which m^i_n is *not* truthful: $m^i_{n,c} \neq s^i_n$. At such an history, we let σ^i prescribe an action which maximizes the discounted sum of the current payoff and of expected continuation payoffs, that is, the expectation of

$$(1-\delta)r^i(s_n, (a^i, a^{-i}_n)) + \delta\gamma^i_\sigma(\omega_{\text{pub},n}, m_{n+1}; \mathcal{T}_{n+1}),$$

where $a^{-i}_n = \rho^{-i}(s_n)$, $(y_n, s_{n+1}) \sim p_{s_n, a^i, a^{-i}_n}$, $\omega_{\text{pub},n} = (m_n, y_n)$, $m_{n+1} = (s_n, a_n, s_{n+1})$ and the expectation is taken over y_n, m_{n+1} and $\mathcal{T}_{n+1}$.

Theorem 2 follows from Proposition 2, which is proven in the next section.

Proposition 2 *The following holds.*

1. *For $\mathcal{T} = (w, z, \rho, x, v, \theta)$, one has $\gamma_\sigma(\omega_{\text{pub}}, m; \mathcal{T}) = z + (1-\delta)\theta(\omega_{pub}, m)$ for every $(\omega_{\text{pub}}, m) \in \Omega_{\text{pub}} \times M$.*[38]

2. *The profile σ is a sequential equilibrium.*

A.3 Proof of Proposition 2

The rationale behind the twisted recursive formula (9) is the simple observation below.

[37] This is true even if $\omega_{\text{pub},n-1}$ and m_n are inconsistent.

[38] There is a slight abuse of notation here, as $\gamma_\sigma(\omega_{\text{pub}}, m; \mathcal{T})$ is only defined for sets $\mathcal{T}$ which can possibly arise along the play, and the equality in the Proposition thus only holds for those. We hope no confusion should arise.

Lemma 10 *For any public history $h_{pub,n+1}$ including reports m_{n+1} in round $n+1$, one has*

$$\tilde{w}_{n+1}(h_{pub,n+1}) + \frac{1-\delta}{\delta}\theta_n(\omega_{pub,n}, m_{n+1}) = \mathbf{E}[z_{n+1}(h_{pub,n+1}) + (1-\delta)\theta_{n+1}(\omega_{pub,n}, m_{n+1})]. \quad (10)$$

Recall that $\tilde{w}_{n+1}$ is given by (7) and is measurable wrt the public information available before the random device is observed in round $n+1$. Here, the expectation is taken with respect to the random output of the randomizing device in round $n+1$. The equality (10) is an algebraic identity which uses the updating formulas. It relates targets and relative values before and after the randomizing device is observed.

Proof. Let $h_{\text{pub},n+1}$ be given, and let k denote the current block: $\tau_k < n+1 \leq \tau_{k+1}$. For clarity, we drop the arguments $h_{\text{pub},n+1}$, $\omega_{\text{pub},n}$ and m_{n+1} from all equalities below. With probability $1-\xi$, $\mathcal{T}_{n+1} = \mathcal{T}_n$ and with probability ξ, z_{n+1} is given by (9). Thus,

$$\begin{aligned}
\mathbf{E}[z_{n+1} + (1-\delta)\theta_{n+1}] &= (1-\xi)\left(z_n + (1-\delta)\theta_n\right) \\
&\quad +\xi\left(w[k+1] + (1-\delta)(1+\frac{1-\delta}{\delta\xi})\theta_n - (1-\delta)\theta_{n+1} + (1-\delta)\theta_{n+1}\right) \\
&= \tilde{w}_{n+1} + (1-\delta)\left((1-\xi) + \xi\left(1+\frac{1-\delta}{\delta\xi}\right)\right)\theta_n \\
&= \tilde{w}_{n+1} + (1-\delta)\left(1+\frac{1-\delta}{\delta}\right)\theta_n \\
&= \tilde{w}_{n+1} + \frac{1-\delta}{\delta}\theta_n.
\end{aligned}$$

■

The previous lemma has the following important consequence, which ultimately ensures that the expected continuation payoff at round n is equal to the target z_n, adjusted for the relative value θ_n.

Lemma 11 *Let h_n^i be a private history of player i up to the action choice in round n such that $m_{n,c}^i = s_n^i$. Denote by $h_{pub,n}$ the public part of h_n^i. One has*

$$z_n^i(h_{pub,n}) + (1-\delta)\theta_n^i(\omega_{pub,n-1}, m_n) = \mathbf{E}_\sigma[(1-\delta)r^i(s_n, a_n) + \delta\left(z_{n+1}^i(h_{pub,n+1}) + (1-\delta)\theta_{n+1}^i(\omega_{pub,n}, m_{n+1})\right)].$$

Here, the expectation $\mathbf{E}_\sigma$ is computed under the belief held by player i at h_n^i, assuming all players play σ from h_n^i on. In concise form, we will write

$$z_n^i + (1-\delta)\theta_n^i = \mathbf{E}_\sigma\left[(1-\delta)r_n^i + \delta\left(z_{n+1}^i + (1-\delta)\theta_{n+1}^i\right)\right].$$

Proof. Since $(v_n, \rho_n, x_n) \in \mathcal{S}_0$, one has

$$v_n^i(h_{\text{pub},n}) + \theta_n^i(\omega_{\text{pub},n-1}, m_n) = \mathbf{E}_\sigma \left[r^i(s_n, a_n) + x_n^i(\omega_{\text{pub},n-1}, \omega_{\text{pub},n}, s_{n+1}^{-i}) + \theta_n^i(\omega_{\text{pub},n}, m_{n+1}) \right]. \tag{11}$$

Again leaving all arguments implicit, we first rewrite (11) as

$$v_n + (1-\delta)\theta_n = \mathbf{E}\left[(1-\delta) r_n + \delta \left(v_n + \frac{1-\delta}{\delta} x_n + \frac{1-\delta}{\delta}\theta_n \right) \right].$$

Adding $z_n - v_n$ on both sides yields

$$\begin{aligned} z_n + (1-\delta)\theta_n &= \mathbf{E}\left[(1-\delta) r_n + \delta \left(v_n + \frac{1}{\delta}(z_n - v_n) + \frac{1-\delta}{\delta} x_n + \frac{1-\delta}{\delta}\theta_n \right) \right] \\ &= \mathbf{E}\left[(1-\delta) r_n + \delta(\tilde{w}_{n+1} + \frac{1-\delta}{\delta}\theta_n) \right] \\ &= \mathbf{E}\left[(1-\delta) r_n + \delta(z_{n+1} + (1-\delta)\theta_{n+1}) \right]. \end{aligned}$$

where the last equality follows from Lemma 10. ■

For later use, we note that, by the best-reply property of ρ_n^i, the equality (11) still holds (resp. a weak inequality $\leq$ holds) if the mixed action $\rho_n^i(m_n^i)$ is replaced by any action a^i in its support (resp. not in its support). This implies that, when an arbitrary action a^i is substituted to $\rho_n^i(m_n)$, the conclusion of Lemma 11 still holds with equality or a weak inequality $\leq$, depending on whether a^i belongs to the support of $\rho_n^i(m_n)$ or not.

Lemma 11 immediately implies, in more probabilistic terms, that

$$z_n + (1-\delta)\theta_n = \mathbf{E}_\sigma \left[(1-\delta) r_n + \delta \left(z_{n+1} + (1-\delta)\theta_{n+1} \right) \mid \mathcal{H}_{\text{pub},n} \right], \ \mathbf{P}_\sigma - a.s.,$$

where $\mathcal{H}_{\text{pub},n}$ is the σ-algebra corresponding to the public information available *after* the randomizing device is observed in round n.[39] Of course, the continuation payoffs $\gamma_\sigma(\omega_{\text{pub},n-1}, m_n; \mathcal{T}_n)$ also satisfy the recursive equation

$$\gamma_\sigma(\omega_{\text{pub},n-1}, m_n; \mathcal{T}_n) = \mathbf{E}_\sigma \left[(1-\delta) r_n + \delta \gamma_\sigma(\omega_{\text{pub},n}, m_{n+1}; \mathcal{T}_{n+1}) \mid \mathcal{H}_{\text{pub},n} \right].$$

Since all quantities are bounded, this implies that

$$\gamma_\sigma(\omega_{\text{pub},n-1}, m_n; \mathcal{T}_n) = z_n(h_{\text{pub},n}) + (1-\delta)\theta_n(\omega_{\text{pub},n-1}, m_n), \tag{12}$$

for every n and every public history $h_{\text{pub},n}$ of positive probability given σ.

[39]Since $\mathcal{H}_{\text{pub},n}$ is finite, the statement actually means that the equality holds for every $h_{\text{pub},n}$ of positive probability under σ.

We now check that σ is a sequential equilibrium. Fix a player i. Note first that, by construction, player i has no profitable one-step deviation at the action phase following a lie. Using the remark following the proof of Lemma 11, this remains true following a truthful report.

Let now h_n^i be an arbitrary private history of player i – excluding reports at round n. We need to prove that truth-telling is optimal for player i at round n following h_n^i. By (12), the objective of player i is to maximize

$$\mathbf{E}\left[(1-\delta)r_n^i + \delta\gamma_\sigma^i(\omega_{\text{pub},n}, m_{n+1}; \mathcal{T}_{n+1})\right] = \mathbf{E}\left[(1-\delta)r_n^i + \delta\left(z_{n+1}^i + (1-\delta)\theta_{n+1}^i(\omega_{\text{pub},n}, m_{n+1})\right)\right]$$

where the expectation is computed given the belief held by player i at h_n^i, when facing σ^{-i}, as a function of the report of player i.

For each report m_n^i and by Lemma 10, this expectation is also given by

$$\mathbf{E}\left[(1-\delta)r_n^i + \delta\tilde{w}_{n+1}^i + (1-\delta)\theta_n^i\right] = (1-\delta)\mathbf{E}\left[r_n^i + x_n^i + \theta_n^i\right] + \mathbf{E}[z_n^i] \tag{13}$$

(beware that we are taking expectations prior to reports in round n, hence neither z_n^i nor the maps x_n^i and θ_n^i are known at h_n^i).

For any untruthful report m_n^i, we compare the expectation $\mathbf{E}_{\text{lie}}$ when reporting m_n^i to that when reporting truthfully, $\mathbf{E}_{\text{truth}}$. We condition on the outcome of the public randomizing device. With probability $1-\xi$, the current block continues at least to round $n+1$, so that $\mathcal{T}_n = \mathcal{T}_{n-1}$. Since $(\rho_{n-1}, x_{n-1}, v_{n-1}) \in \mathcal{S}_0$, and since the belief of player i at h_n^i is deduced from ρ_{n-1}, the conditional $\mathbf{E}_{\text{truth}}$ exceeds the conditional $\mathbf{E}_{\text{lie}}$ by at least $\eta(1-\delta)$.

With probability ξ, the play switches a new block in round n. Conditional on switching, lying may possibly improve both expectations on the right-hand side of (13). Yet, the improvement of $(1-\delta)\mathbf{E}[r_n^i + x_n^i + \theta_n^i]$ is of at most $2\kappa_0(1-\delta)$, and the improvement of $\mathbf{E}[z_n^i]$ is, given (9) and the choice of $\bar{\delta}$, of at most $3\kappa_0(1-\delta)$.

Hence,

$$\mathbf{E}_{\text{truth}} - \mathbf{E}_{\text{lie}} \geq (1-\xi) \times \eta(1-\delta) - \xi \times 5\kappa_0(1-\delta),$$

which is nonnegative since $\dfrac{\xi}{1-\xi} \leq \dfrac{\eta}{5\kappa_0}$.

B Proofs for independent private values

Proof of Lemma 3. Fix a non-coordinate direction λ. Suppose first that $\lambda^i \leq 0$, all i. Consider the vector $v \in Ext^{po}$ that maximizes $\lambda \cdot v$, and the corresponding policy ρ. This policy implements a distribution over $\Delta(A)$. Consider the constant (and hence weakly admissible, for $x = 0$) policy that uses the public randomization device to replicate this distribution (independently of the reports). The IPV assumption ensures that all players are weakly worse off. Hence $\bar{k}(\lambda) \geq \lambda \cdot v$, and so $\text{co}(Ext^{pu} \cup Ext^{po}) \subset \{v \in \mathbf{R}^i : \lambda \cdot v \leq \bar{k}(\lambda)\}$ (if another constant policy improves the score further, consider it instead). Suppose next that $\lambda^i < 0$ for all $i \in J \subsetneq I$, and there exists i such that $\lambda^i > 0$. Again, consider the vector $v \in Ext^{po}$ that maximizes $\lambda \cdot v$. Because $v \in Ext^{po}$, v also maximizes $\hat{\lambda} \cdot v$ over $v \in Ext^{po}$, for some $\hat{\lambda} \geq 0$, $\hat{\lambda} \in \Lambda$. Furthermore, we can take $\hat{\lambda}^i = 0$ for all $\lambda^i < 0$. Such a vector is achieved by a policy that only depends on $(s^i)_{i \notin J}$, because of private values. Truth-telling is trivial for these types, and hence this policy is also weakly admissible in the direction λ. Hence again $\bar{k}(\lambda) \geq \lambda \cdot v$, and so $\text{co}(Ext^{pu} \cup Ext^{po}) \subset \{v \in \mathbf{R}^i : \lambda \cdot v \leq \bar{k}(\lambda)\}$. Directions $\lambda \geq 0$ are unproblematic, as both efficient and constant policies are weakly admissible for some choice of x.

Coordinate directions $\lambda = e^i$ are also immediate: the vector that maximizes v^i over $\text{co}(Ext^{pu} \cup Ext^{po})$ is part of the Pareto-frontier, and the corresponding policy is weakly admissible using AGV. Scores in the directions $\lambda = -e^i$ are (at least) $-\underline{v}^i$, hence $V^* \cap \text{co}(Ext^{pu} \cup Ext^{po}) \subset \{v \in \mathbf{R}^i : -e^i \cdot v \leq \bar{k}(-e^i)\}$. ■

B.1 Proof of Proposition 1

Given a policy $\rho : S \to A$, we denote by q_ρ the transition probability over $\Omega_{\text{pub}} \times S^i$, induced by ρ and truth-telling. More generally, we use the notation p_ρ whenever expectations/laws should be computed under the assumption that states are truthfully reported, actions chosen according to ρ, and transitions determined using p. For instance, $p_\rho(s^{-i} \mid \bar{\omega}_{\text{pub}})$ is the (conditional) law of s^{-i} under $p_{\bar{s},\rho(\bar{s})}$, given $\bar{y}$. Given the IPV assumption, it is thus $\times_{j \neq i} p^j(s^j \mid \bar{s}^i, \rho^j(\bar{s}), \bar{y})$.

Fix a weakly truthful pair (ρ, x), with $\rho : S \to A$ and $x : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$. For $i \in I$, $(\bar{\omega}_{\text{pub}}, s^i) \in \Omega_{\text{pub}} \times S^i$, set

$$\xi^i(\bar{\omega}_{\text{pub}}, s^i) := \mathbf{E}_{s^{-i} \sim p_\rho(\cdot|\bar{\omega}_{\text{pub}})}[x^i(\bar{\omega}_{\text{pub}}, s^i, s^{-i})].$$

Plainly, the pair (ρ, ξ) is weakly truthful as well.

The next lemma is the long-run analog of Claim 1 (outdated reference) in Athey and Segal (2013, AS). The logic of the proof is identical.

Lemma 12 *Define* $\tilde{x} : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ *by*

$$\tilde{x}^i(\bar{\omega}_{\text{pub}}, s)(= \tilde{x}^i(\bar{\omega}_{\text{pub}}, s^i)) = \theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, s^i) - \mathbf{E}_{\tilde{s}^i \sim p_\rho(\cdot|\bar{\omega}_{\text{pub}})}[\theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, \tilde{s}^i)].$$

Then $(\rho, \tilde{x})$ is weakly truthful.

Proof.

We first argue that $\theta^i_{\tilde{x}}(\cdot) = \tilde{x}^i(\cdot)$ (up to an additive constant, as usual). It suffices to prove that $\tilde{x}^i$ is a solution to the linear system (with unknowns θ)

$$\theta^i(\bar{\omega}_{\text{pub},1}, s_1) - \theta^i(\bar{\omega}_{\text{pub},2}, s_2) = \tilde{x}^i(\bar{\omega}_{\text{pub},1}, s_1) + \mathbf{E}_{p_\rho(\cdot|\bar{\omega}_{\text{pub},1},s_1)}[\theta^i(\omega, t)] - \left(\tilde{x}^i(\bar{\omega}_{\text{pub},2}, s_2) + \mathbf{E}_{p_\rho(\cdot|\bar{\omega}_{\text{pub},2},s_2)}[\theta^i(\omega, t)]\right),$$

(for $(\bar{\omega}_{\text{pub}_1}, s_1), (\bar{\omega}_{\text{pub},2}, s_2) \in \Omega_{\text{pub}} \times S^i$). But this follows from the fact that for each $(\bar{\omega}_{\text{pub}}, s) \in \Omega_{\text{pub}} \times S$, one has

$$\mathbf{E}_{\pi_\rho(\cdot|s)}[\tilde{x}^i(\omega_{\text{pub}}, t^i)] = \mathbf{E}_{y \sim p_\rho(\cdot|s)}\left[\mathbf{E}_{t \sim p_\rho(\cdot|\omega_{\text{pub}})} \tilde{x}^i(\omega_{\text{pub}}, t^i)\right] = 0.$$

Fix next a player i, and an outcome $\bar{\omega} = (\bar{s}, \bar{m}, \bar{s}, \bar{y})$ such that $\bar{s}^{-i} = \bar{m}^{-i}$ and $\bar{a}^{-i} = \rho^{-i}(\bar{m})$. Since (ρ, ξ) is weakly truthful, for each such $\bar{\omega}$ and $s^i \in S^i$, the expectation of

$$r^i(s^i, \rho(s^{-i}, m^i)) + \xi^i(\bar{\omega}_{\text{pub}}, m^i) + \theta^i_{\rho,r}(\omega_{\text{pub}}, t) + \theta^i_{\rho,\xi}(\omega_{\text{pub}}, t) \tag{14}$$

is maximized for $m^i = s^i$. Here, the expectation is to be computed as follows. First, s^{-i} is drawn according to the belief of i which, given the IPV assumption, is equal to $p_\rho(\cdot \mid \bar{\omega}_{\text{pub}})$; next, (y, t) is drawn $\sim p_{s,\rho(m^i, s^{-i})}$, and $\omega_{\text{pub}} = (s^{-i}, m^i, y)$.

To prove that $(\rho, \tilde{x})$ is admissible, we need to prove that the similar expectation of

$$r^i(s^i, \rho(s^{-i}, m^i) + \tilde{x}^i(\bar{\omega}_{\text{pub}}, m^i) + \theta^i_{\rho,r}(\omega_{\text{pub}}, t) + \theta^i_{\rho,\tilde{x}}(\omega_{\text{pub}}, t) \tag{15}$$

is maximized for $m^i = s^i$ as well. Fix $m^i \in M^i$. Using $\theta^i_{\tilde{x}} = \tilde{x}^i$, and the definition of $\tilde{x}^i$, the expectation of the expression in (15) is equal to the expectation of

$$r^i(s^i, \rho(s^{-i}, m^i)) + \theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, m^i) + \theta^i_{\rho,r}(\omega_{\text{pub}}, t) + \theta^i_{\rho,\xi}(\omega_{\text{pub}}, t^i) - \mathbf{E}_{\tilde{s}^i \sim p_\rho(\cdot|\omega_{\text{pub}})} \theta^i_{\rho,\xi}(\omega_{\text{pub}}, \tilde{s}^i), \tag{16}$$

up to the constant $\mathbf{E}_{\tilde{s}^i \sim p_\rho(\cdot|\bar{\omega}_{\text{pub}})} \theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, \tilde{s}^i)$, which does not depend on m^i.

Next, observe that by definition of $\theta_{\rho,\xi}$, one has

$$\theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, m^i) = \xi^i(\bar{\omega}_{\text{pub}}, m^i) + \mathbf{E}_{p_\rho} \theta^i_{\rho,\xi}(\omega_{\text{pub}}, \tilde{s}^i),$$

again up to a constant that does not depend on m^i.

Thus, (and up to a constant), the expression in (15) has the same expectation as the expression in (14), so that the weak truthfulness of $(\rho, \tilde{x})$ follows from that of (ρ, ξ).

■

Corollary 6 *Let $\mu_{ij} \in \mathbf{R}$ be arbitrary. For $i \in I$, set*

$$\hat{x}^i(\bar{\omega}_{\text{pub}}, m) = \tilde{x}^i(\bar{\omega}_{\text{pub}}, m^i) + \sum_{j \neq i} \mu_{ij} \tilde{x}^j(\bar{\omega}_{\text{pub}}, m^j).$$

Then $(\rho, \bar{x})$ is weakly truthful.

Proof. It is enough to check that, at any $\bar{\omega}_{\text{pub}}$, the expectation of $\theta_{\rho,\tilde{x}^j}(\omega_{\text{pub}}, \tilde{s}^j) = \tilde{x}^j(\omega_{\text{pub}}, \tilde{s}^j)$ does not depend on m^i. But this expectation is zero (as in Claim 2 in AS). ■

Proof of Proposition 1. Set $\hat{x}^i = \tilde{x}^i$ for $i \notin I(\lambda)$. For $i \in I(\lambda)$ set

$$\hat{x}^i(\bar{\omega}_{\text{pub}}, s) := \tilde{x}^i(\bar{\omega}_{\text{pub}}, s^i) - \frac{1}{|I(\lambda)| - 1} \sum_{j \neq i} \frac{\lambda^j}{\lambda^i} \tilde{x}^j(\bar{\omega}_{\text{pub}}, s^j)$$

and apply the previous corollary. ■

Proof of Lemma 4. We focus on a fixed player i. Fix a pair (ρ, x). The optimality of truth-telling given (ρ, x) is equivalent to truth telling solving the following Markov decision process:

- The state space is $\Omega_{\text{pub}} \times S^i$, with elements $(\bar{\omega}_{\text{pub}}, s^i)$;
- The action set is $M^i = S^i$: today's announced state;
- The reward is $r^i(s^i, \rho(s^{-i}, m^i)) + x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})$;
- Transitions are given by $p_{(\bar{\omega}_{\text{pub}}, s^i), m^i}(\omega_{\text{pub}}, t^i) = 0$ if ω_{pub} does not specify m^i as i's report, or does not specify a signal y in the support of the distribution determined by $\rho(s^{-i}, m^i)$. Otherwise, it is derived from p and ρ in the obvious way.

Let us define, as in Section 6,

$$r^i_\rho((\bar{\omega}_{\text{pub}}, s^i), m^i) = \mathbf{E}_{s^{-i}|\bar{\omega}_{\text{pub}}} r^i(s^i, \rho(s^{-i}, m^i)), \qquad x^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = \mathbf{E}_{s^{-i}, y|\bar{\omega}_{\text{pub}}, \rho} x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}).$$

Under our unichain assumption, there is an equivalent LP formulation (see Puterman, Ch. 8.8). Namely, agent i choose $\pi_\rho \in \hat{\Pi}^i_\rho$ to maximize

$$\mathbf{E}_{\pi^i_\rho}\left[r^i_\rho x^i_\rho,\right]$$

where $\hat{\Pi}^i(\rho) \subset \Delta(\Omega_{\text{pub}} \times S^i \times M^i)$ is the set of distributions π such that, for all states $(\omega_{\text{pub}}, t^i)$,

$$\sum_{\tilde{m}^i} \pi^i_\rho(\omega_{\text{pub}}, t^i, \tilde{m}^i) = \sum_{\bar{\omega}_{\text{pub}}, s^i, m^i} p_{(\bar{\omega}_{\text{pub}}, s^i), m^i}(\omega_{\text{pub}}, t^i) \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i).$$

This is equation 3 from Section 6. Our goal is to examine which ρ are weakly truthful for some choice of x^i_ρ.

Suppose first that $p_{s,a}(y)$ is independent of s. Then $x^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i)$ is a function of $\bar{\omega}_{\text{pub}}, m^i$ only, so we write $x^i_\rho(\bar{\omega}_{\text{pub}}, m^i)$. We can consider the zero-sum game in which we pick x^i_ρ in some large but bounded set $[-M, M]$ to minimize

$$\sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) + \sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} (\pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) - \mu(\bar{\omega}_{\text{pub}}, m^i)) x^i_\rho(\bar{\omega}_{\text{pub}}, m^i).$$

This is a game between player i who chooses $\pi^i_\rho \in \hat{\Pi}^i(\rho)$ and the designer who chooses $x^i_\rho \in [-M, M]^{\Omega_{\text{pub}} \times S^i}$. By the minmax theorem, we can think of i moving first. It is then clear that any optimal strategy for i must specify, for all $(\bar{\omega}_{\text{pub}}, m^i)$,

$$\sum_{s^i} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = \mu(\bar{\omega}_{\text{pub}}, m^i),$$

and we can pick x^i_ρ to be the optimal strategy of the designer.[40] Thus, we may restrict further player i to choose from $\Pi^i(\rho)$ where $\pi^i_\rho \in \Pi^i(\rho)$ if and only if $\pi^i_\rho \in \hat{\Pi}^i(\rho)$ and $\sum_{s^i} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = \mu(\bar{\omega}_{\text{pub}}, m^i)$ for all $(\bar{\omega}_{\text{pub}}, m^i)$. Note that for $\pi^i_\rho \in \Pi^i(\rho)$, the original objective of the LP becomes

$$\sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) + \sum_{\bar{\omega}_{\text{pub}}, m^i} \mu(\bar{\omega}_{\text{pub}}, m^i) \hat{x}^i(\bar{\omega}_{\text{pub}}, s^i, m^i),$$

and because the second term does not involve π^i_ρ, it is irrelevant for the maximization. Hence, transfers cannot achieve more than the restriction to $\Pi^i(\rho)$, and the policy ρ is weakly truthful if and only if the solution to the program $\hat{\mathcal{P}}^i(\rho)$: maximize

$$\sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i)$$

over $\pi^i_\rho \in \Pi^i(\rho)$ is given by $\pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = 0$ if $m^i \neq s^i$, all $(\bar{\omega}_{\text{pub}}, s^i)$ (and so $\pi^i_\rho(\bar{\omega}_{\text{pub}}, m^i, m^i) = \mu(\bar{\omega}_{\text{pub}}, m^i)$). Invoking Proposition 1 to balance the budget, this concludes the proof of

[40] Pick as M any common upper bound on the Lagrangian coefficients on the optimization program:

$$\max \sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i)$$

over $\pi^i_\rho \in \hat{\Pi}^i(\rho)$ subject to for all $(\bar{\omega}_{\text{pub}}, m^i)$,

$$\sum_{s^i} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = \mu(\bar{\omega}_{\text{pub}}, m^i).$$

Lemma 4 for non-coordinate directions. In coordinate directions $\lambda = \pm e^i$, note that setting $x = 0$ provides the appropriate truth telling incentives (the only player who makes a report that affects the outcome is player i when $\lambda = e^i$, but in this case the policy $\rho[e^i]$ trivially is a solution to the LP for $x^i_\rho = 0$.

Consider next the case of state-dependent signaling. Then x^i is a function of $(\bar{\omega}_{\text{pub}}, [s^i], m^i)$ only. Then we consider the zero-sum game with payoff

$$\sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) r^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i)$$
$$+ \sum_{(\bar{\omega}_{\text{pub}}, s^i, m^i)} (\pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) - \mu(\bar{\omega}_{\text{pub}}, m^i)) x^i_\rho(\bar{\omega}_{\text{pub}}, [s^i], m^i),$$

and it follows from the minmax theorem again that player i's optimal strategy specifies, for all $(\bar{\omega}_{\text{pub}}, m^i)$,

$$\sum_{s^i \in [m^i]} \pi^i_\rho(\bar{\omega}_{\text{pub}}, s^i, m^i) = \sum_{s^i \in [m^i]} \mu(\bar{\omega}_{\text{pub}}, m^i).$$

The remainder follows as in the first case. ■

Proof of Lemma 5. Follows immediately from Lemma 4 and the discussion following its statement in Section 6. Note that the proof of Lemma 4 also covers the case of state-dependent signaling, so Lemma 5 holds for this case as well, given the definition of $\bar{k}(\lambda)$ for that case. ■

C Proof of Theorem 3

First, note that Assumption **1** ensures that for all $\rho \in \Xi$, all s, $a = \rho(s)$ (also, for all (s, a) where $(s^{-i}, a) \in D^i$ for some i) and for all $d > 0$,

1. For each i, there exists $\hat{x}^i : S \times Y \to \mathbf{R}$ such that, for all $\hat{a}^i \neq a^i$, all $\hat{s}^i$,

$$\mathbf{E}[\hat{x}^i(s, y) \mid a, s] - \mathbf{E}[\hat{x}^i(s, y) \mid a^{-i}, \hat{a}^i, s^{-i}, \hat{s}^i] > d;$$

   (The expectation is with respect to the signal y.)

2. For every pair i, j, $i \neq j$, $\lambda^i \neq 0, \lambda^j \neq 0$, there exists $\hat{x}^h : S \times Y \to \mathbf{R}$, $h = i, j$,

$$\lambda^i \hat{x}^i(s, y) + \lambda^j \hat{x}^j(s, y) = 0, \tag{17}$$

   and for all $\hat{a}^h \neq a^h$, all $\hat{s}^h$,

$$\mathbf{E}[\hat{x}^h(s, y) \mid a, s] - \mathbf{E}[\hat{x}^h(s, y) \mid a^{-h}, \hat{a}^h, \hat{s}^h, s^{-h}] > d.$$

See Lemma 1 of Kandori and Matsushima (1998). By subtracting the constant $\mathbf{E}[\hat{x}^i(s,y) \mid a, s]$ from all values $\hat{x}^i(s,y)$ (which does not affect (17), since (17) must also hold in expectations), we may assume that, for our fixed choice of a, it holds that, for all s, $\hat{x}^i$ is such that $\mathbf{E}[\hat{x}^i(s,y) \mid a, s] = 0$, all i.

Given this normalization, we have that

$$\mathbf{E}[\hat{x}^i(s,y) \mid a^{-i}, \hat{a}^i, s^{-i}, \hat{s}^i] < -d,$$

for any choice $(\hat{s}^i, \hat{a}^i)$ that does not coincide with (s^i, a^i) (in which case the expected transfer is zero). Intuitively, the transfer $\hat{x}^i$ ensures that, when chosen for high enough d, it never pays to deviate in action, even in combination with a lie, rather than reporting the true state and playing the action profile a that is agreed upon, holding the action profile to be played constant across reports $\hat{s}^i$, given s^{-i}. Deviations in reports might also change the action profile played, but the difference in the payoff from such a change is bounded, while d is arbitrary.

More formally, fix some pure policy $\rho : S \to A$ with long-run payoff v. There exists $\theta : S \to \mathbf{R}^I$ such that, for all s,

$$v + \theta(s) = r(s, \rho(s)) + \mathbf{E}_{p_{s,\rho(s)}}[\theta(t)].$$

Consider the M.D.P. in which player i chooses reports $m^i \in M^i = S^i$ and action $\hat{\rho}^i : M^i \times S^{-i} \to A^i$, and his realized reward is $r^i(s, a^i, \rho^{-i}(m^i, s^{-i})) + \hat{x}^i(m^i, s^{-i}, y)$. Then we may pick $d > 0$ such that, given $\hat{x}^i$, every optimal policy specifies $\hat{\rho}^i(m^i, s^{-i}) = \rho^i(m^i, s^{-i})$. Note also that because of our normalization of $\hat{x}^i$, the private values in this M.D.P. are equal to θ^i if player i sets $m^i = s^i$.

The argument is similar for the case of coordinate directions. In case $\lambda = -e^i$ (resp. $+e^i$) use Assumption **2** (resp. again **1**) and follow Kandori and Matsushima (1998, Case 1 and 2, Theorem 1).

This transfer addresses deviations at the action step. Adding to $\hat{x}$ the transfer constructed in the proof of Lemma 4 yields a transfer that ensures that deviations are suboptimal at either step.

D Proof of Theorem 4

We here prove Theorem 4, recalled below.

Theorem 4 *Assume that monitoring has the product structure, and that Assumptions 3–4 hold. If W has non-empty interior, the set of (Nash, sequential) equilibrium payoffs converges to W as $\delta \to 1$.*

We organize the proof as follows. We first define modified scores $\tilde{k}(\lambda)$ and the corresponding set $\tilde{\mathcal{H}} := \{v \in \mathbf{R}^I : \lambda \cdot v \leq \tilde{k}(\lambda) \text{ all } \lambda \in \Lambda\}$. We next observe that the IPV assumption, together with Assumption X, ensures that $\tilde{\mathcal{H}} = W$.

To a large extent, the rest of the proof is an adaptation of the proof of Theorem 2, to which we will extensively refer to avoid duplications.

D.1 Alternative scores

Fix an arbitrary $s_* \in S$. We define finite-horizon games, parameterized by final payoffs. Given a horizon $T \in \mathbf{N}$, final transfers $x : \Omega_{\text{pub}}^T \to \mathbf{R}^I$, and a "relative values" map $\theta : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$, we define $G(T, x, \theta)$ as the T-stage repetition of the underlying stage game with (not necessarily truthful) communication and initial state profile s_*. The game $G(T, x, \theta)$ ends with the draw of s_{T+1} in round $T + 1$.

Payoffs in $G(T, x, \theta)$ are given by

$$\frac{1}{T}\left(\sum_{n=1}^{T} r(s_n, a_n) + x(h_{\text{pub},T}) + \theta(\omega_{\text{pub},T}, s_{T+1})\right),$$

where $h_{\text{pub},T}$ is the public history up to round T.

We denote by C a uniform bound on $\theta_{\rho,r}$, when ρ ranges through the set of all policies.

For $\lambda \in \Lambda$ and $T \in \mathbf{N}$, we define the maximization problem $\tilde{\mathcal{P}}_T(\lambda) : \tilde{k}_T(\lambda) := \sup \lambda \cdot v$, where the supremum is taken over all (σ, x, θ), such that

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

Set $\tilde{k}(\lambda) = \limsup_T \tilde{k}_T(\lambda)$, and $\tilde{\mathcal{H}} := \{v \in \mathbf{R}^I : \lambda \cdot v \leq \tilde{k}(\lambda) \text{ all } \lambda \in \Lambda\}$.

D.2 The proof of $\tilde{\mathcal{H}} = W$

We first note that the discussion in Section 6 implies Lemma 13 below.

Lemma 13 *For all $\lambda \neq -e^i$, one has $\bar{k}(\lambda) = k_*(\lambda)$.*

The equality $\tilde{\mathcal{H}} = W$ follows from Lemmas 14 and 15 below.

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

Proof.

Fix $\lambda \in \Lambda$ with $\lambda \neq -e^i$ for all $i \in I$. We first prove that $k(\lambda) \leq \tilde{k}(\lambda)$. Let an arbitrary feasible triple $(\rho, x, v) \in \mathcal{P}(\lambda)$ and an integer $T \in \mathbf{N}$ be given. Set $\theta := \theta_{\rho, r+x}$, and define $x_t : \Omega^T_{\text{pub}} \to \mathbf{R}^I$ 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 σ_T be the strategy profile in $G(T, x_T, \theta)$ defined as follows: (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

$$r^i(s^i_n, a^i, \rho^{-i}(m_n)) + x^i(\omega_{\text{pub},n-1}, \omega_{\text{pub},n}) + \theta^i(\omega_{\text{pub},n}, s_{n+1})$$

otherwise.

Since $(\rho, x) \in \mathcal{S}_0$, it is easily checked that σ_T (when supplemented with appropriate beliefs) is a sequential equilibrium in $G(T, x, \theta)$. Hence, $\lambda \cdot \tilde{\gamma}_T(\sigma_T) \leq \tilde{k}(\lambda)$ where $\tilde{\gamma}_T(\sigma_T)$ is the expected payoff of σ_T in $G(T, x, \theta)$. On the other hand, $\tilde{\gamma}_T(\sigma_T)$ converges to $\mathbf{E}_{\mu_\rho}[r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})]$ as $T \to +\infty$, hence $\lim_{T \to +\infty} \gamma_T(\sigma_T) = \lambda \cdot v$. This shows that $k(\lambda) \leq \tilde{k}(\lambda)$, as desired.

We next prove that $\tilde{k}(\lambda) \leq \bar{k}(\lambda)$. Fix $\varepsilon > 0$. Given $T \in \mathbf{N}$, pick a feasible triple (σ, x, θ) in $\tilde{\mathcal{P}}_T(\lambda)$ which achieves $\tilde{k}(\lambda)$ up to ε. Mimicking the argument in Lemma **??**, there is a profile $\tilde{\sigma}_T$ which only depends on the types 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 $\lambda \in \Lambda$, $\|\theta\| \leq C$ and $\lambda \cdot x(\cdot) \leq 0$, one has

$$\lambda \cdot \tilde{\gamma}(\tilde{\sigma}_T) \leq \lambda \cdot \gamma_T(\sigma_T) + \frac{C}{T}, \tag{18}$$

where $\gamma_T(\sigma_T)$ is the payoff induced in the T-stage 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 λ-weighted T-stage game. By the unichain assumption on transitions, $\lim_{T\to+\infty} v_T(\lambda) = \bar{k}(\lambda)$. Let now $T \to +\infty$ in (18) to get $\tilde{k}(\lambda) - \varepsilon \leq \bar{k}(\lambda)$, and the result follows. ■

Lemma 15 *For* $\lambda = -e^i$, $\tilde{k}(\lambda) = -w_i^i$.

Proof. Throughout the finite-horizon 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 $A_k^j := \{\alpha^j \in \Delta(A^j) : \alpha^j(a_l^j) = \frac{m}{k}$, for some $m = 0, \ldots, k$ and all $a_l^j \in A^j\}$, where $\alpha^j(a_l^j)$ is the probability assigned to action a_l^j by mixed action profile α^j. The set A_k^j consists of those mixed action profiles that to each action assign a rational probability with k as denominator. For any $\alpha^j \in \Delta(A^j)$, there exists $a_k^j \in A_k^j$ such that $d(a_k^j, \alpha^j) \leq |A^j|/k$; similarly, for all $\alpha^{-i} \in \times_{j\neq i}\Delta(A^j)$, $d(a_k^{-i}, \alpha^{-i}) \leq |A^{-i}|/k$, for some $a_k^{-i} \in A_k^{-i} := \times_{j\neq i} A_k^j$. (Throughout, we use the Euclidean distance). We write Σ_k^j for the strategies of j that take values in A_k^j and only depend on the history of i's signals. Finally, we let

$$\underline{\sigma}_k = \arg \min_{\sigma^{-i}\in\Sigma_k^{-i}} \max_{\sigma^i} \lim_T \frac{1}{T} \mathbf{E}_\sigma \left[\sum_{t=1}^{\infty} g^i(s^i, a^i, y) \right]$$

for the minmax payoff of player i when players $-i$ are constrained to strategies with values in A_k^{-i}. We write w_k^i for the expectation of i's payoff under $\underline{\sigma}_k$. Given $T \in \mathbf{N}$, we also write $w_{k,T}^i$, $\underline{\sigma}_{k,T}^{-i}$ for the average payoff (and strategy) over the first T rounds (as well as $\underline{\sigma}_T^{-i}$ for the average payoff over the first T rounds under $\underline{\sigma}^i$). Given the product structure, these strategies are measurable with respect to the history of signals of player i, and we write $h_{\text{pub},t}^i \in H_{\text{pub},t}^i$ for such public histories (*i.e.*, $H_{\text{pub},t}^i = (Y^i)^{t-1}$). Given $j \neq i$, k, $h_{\text{pub},T}^i \in H_{\text{pub},T}^i$, let

$$T(a_k^{-i}) = \{t = 0, \ldots, T : \underline{\sigma}_k^{-i}(h_{\text{pub},T}^i) = a_k^{-i}\}$$

denote the rounds at which $\underline{\sigma}_k^{-i}$ prescribes a_k^{-i}. (Here, $h_{\text{pub},t}^i$ refers to an initial segment of $h_{\text{pub},T}^i$.) Given any history $h_{\text{pub},T}$ and any a_k^{-i}, $y \in Y$, let

$$n[a_k^{-i}](y) = |\{t \in T(a_k^{-i}) : y_t = y\}|$$

denote the number of times the signal profile y is observed when players $-i$ are supposed to play a_k^{-i}, and similarly, for all $j \neq i$,

$$n[a_k^{-i}](y^j) = \sum_{y^{-j}} n[a_k^{-i}](y)$$

be the number of times signal y^j was observed during those rounds. Let

$$f[a_k^{-i}](y) = \begin{cases} \frac{n[a_k^{-i}](y)}{|T(a_k^{-i})|} & \text{if } |T(a_k^{-i})| \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and similarly $f[a_k^{-i}](y^j)$, $f[a_k^{-i}](y^{-i})$. For $j \neq i$, $h_{\text{pub},T}$, we let

$$D^j(h_{\text{pub},T}) = \sum_{a_k^{-i} \in A_k^{-i}} \frac{|T(a_k^{-i})|}{T} \sum_y \left| f[a_k^{-i}](y) - f[a_k^{-i}](y^{-j}) \mathbf{P}[y^j \mid a_k^j] \right|,$$

and also, given $\eta > 0$, define the *test*:

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

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

Claim 7 *For every $\varepsilon > 0$ and $\eta > 0$, $j \neq i$, there exists T_0 such that, if $T \geq T_0$,*

$$\mathbf{P}_{\underline{\sigma}_k^j(\delta), \sigma^{-j}} \left[\tau_\eta^j(h_T) = 0 \right] < \varepsilon.$$

for all strategy profiles σ^{-j}.

In words, if player j uses $\underline{\sigma}_k^j$, he is almost sure to pass the test τ_η^j no matter players $-j$'s strategy profile. (Again, without loss, strategies are public). This is where approachability is used, see Gossner (1995) for details.

Given $\varepsilon > 0$, we first pick $M > 0$ such that

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

(where $\bar{r}$ is an upper bound on the magnitude of payoffs) or

$$M > T\frac{\bar{r}}{\varepsilon}. \tag{19}$$

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 to ε exceeds the payoff from the highest reward for T rounds followed by a probability of failing the test by at least a probability 2ε. Note that this reward is a function of the public history only.

We set $x^i(h_{\text{pub},T}) = 0$ for all $h_{\text{pub},T}$, and, given η, for $j \neq i$,

$$x^i(h_{\text{pub},T}) = \begin{cases} -M & \text{if } \tau^j_\eta(h_{\text{pub},T}) = 0, \\ 0 & \text{otherwise.} \end{cases}$$

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

Claim 8 *Fix $k \in \mathbf{N}$. For every $\eta > 0$, there exists $T_1 \in \mathbf{N}$, for all $T \geq T_1$, if σ is a (sequential) equilibrium of $G(T, x, \theta)$, where M satisfies* (19)*, then*

$$\frac{1}{T}\mathbf{E}_{s_0,\sigma}\left[\sum_{t=1}^{T} g^i(s^i, a^i, y)\right] \leq w^i_{k,T} + \eta.$$

Proof. Let $T \geq T_0$, and $\varepsilon = \frac{\eta}{4\bar{r}(I-1)}$. Given ε, by the choice of M, it must hold that

$$\mathbf{P}_\sigma\left[\tau^j_\eta(h_T) = 0\right] < 2\varepsilon$$

in all equilibria of $G(T, x, \theta)$.

Consider the constrained optimization program: maximize

$$v^i_T = \frac{1}{T}\mathbf{E}_{s_0,\sigma}\left[\sum_{t=1}^{T} g^i(s^i_t, a^i_t, y_t)\right]$$

over strategy profiles σ (σ^j measurable with respect to histories (h^j_t)), such that

$$\mathbf{P}_\sigma\left[\tau^j_\eta(h_T) = 0\right] < 2\varepsilon,$$

for all $j \neq i$. Note that

$$\sum_y \sum_{a^{-i}_k} \frac{|T(a^{-i}_k)|}{T}\left|f[a^{-i}_k](y) - f[a^{-i}_k](y^{-j})\mathbf{P}[y^{-j} \mid \underline{\sigma}^j_k]\right| < 2\varepsilon$$

implies that, by repeated substitution,

$$\sum_y \sum_{a^{-i}_k} \frac{|T(a^{-i}_k)|}{T}\left|f[a^{-i}_k](y) - f[a^{-i}_k](y^i) \times_{j\neq i} \mathbf{P}[y^j \mid \underline{\sigma}^j_k]\right| < 2(I-1)\varepsilon.$$

We have that

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1} g^i(s^i_t, a^i_t, y_t) &\leq \sum_{y, a_k^{-i}} \sum_{\{t: y_t = y, \underline{\sigma}^{-i}_{k,t} = a_k^{-i}\}} \left| g^i(s^i_t, a^i_t, y) - \sum_{y'} g^i(s^i_t, a^i_t, y')\mathbf{P}[y' \mid a_k^{-i}] \right| \\ &+ \sum_{a_k^{-i}} \sum_{\{t: \underline{\sigma}^{-i}_{k,t} = a_k^{-i}\}} \sum_{y'} g^i(s^i_t, a^i_t, y')\mathbf{P}[y' \mid a_k^{-i}] \\ &\leq 2\bar{r}(I-1)\varepsilon + \sum_{a_k^{-i}} \sum_{\{t: \underline{\sigma}^{-i}_{k,t} = a_k^{-i}\}} \sum_{y'} g^i(s^i_t, a^i_t, y')\mathbf{P}[y' \mid a_k^{-i}],\end{aligned}$$

and the expectation of the last term is maximized by the strategy $\underline{\sigma}^i_T$, for a payoff $w^i_{k,T}$.

Hence

$$v^i_T \leq w^i_{k,T} + \eta.$$

■

Finally, note that $w^i_{k,T} \to w^i_T$ as $k \to \infty$, so that, given any $\eta > 0$, we can pick $T \in \mathbf{N}$ such that $w^i_T < w^i_i + \frac{\eta}{2}$, and $k \in \mathbf{N}$ such that $w^i_{k,T} < w^i_T + \frac{\eta}{2}$ (hence, $w^i_{k,T} < w^i_i + \eta$) and so that, given Claim 8, all equilibria of $G(T, x, \theta)$ yield a payoff no more than $w^i_i + 2\eta$. It follows then that $\tilde{k}(\lambda) > -w^i_i - 2\eta$, and since $\eta > 0$ is arbitrary, the result follows. ■

D.3 Adaptation of the proof of Theorem 2

For simplicity, we will write the proof under the following Assumption **3**', similar to Assumption **3**, and will comment on the adjustments to be made under the latter one.

Assumption 3' There exists a policy $\rho_* : S \to \Delta(A)$, transfers $x_* : S \to \mathbf{R}^I$ such that the following holds with $\theta_* := \theta_{\rho_*, r + x_*}$. For each player $i \in I$, any two $\bar{s}^i, \tilde{s}^i \in S^i$, and any $s^{-i} \in S^{-i}$, one has

$$\begin{aligned}& r^i(\bar{s}^i, \rho_*(\bar{s}^i, s^{-i}) + x_*(\bar{s}^i, s^{-i}) + \mathbf{E}_{p_{\bar{s}^i, s^{-i}, \rho_*(\bar{s}^i, s^{-i})}}[\theta_*(t)] \\ \geq\ & r^i(\bar{s}^i, \rho_*(\tilde{s}^i, s^{-i}) + x_*(\tilde{s}^i, s^{-i}) + \mathbf{E}_{p_{\bar{s}^i, s^{-i}, \rho_*(\tilde{s}^i, s^{-i}}}[\theta_*(t)],\end{aligned}$$

with strict inequality for at least one $s^{-i} \in S^{-i}$.

D.3.1 The strategies

Given $z \in Z_\eta$, and an initial distribution of states $p \in \Delta(S)$, we will construct a sequential equilibrium σ with payoff z for large enough δ.

The play is divided into an infinite sequence of blocks. Odd blocks serve as transition blocks, and are typically much shorter than even blocks. Even blocks may either be "regular," or devoted to the punishment of a single player. A regular block has a random duration, which follows a geometric duration of parameter $(1-\delta)^{\beta}$. A punishment block has a duration of $(1-\delta)^{-\beta}$ rounds. Transition blocks have a random duration, with a geometric duration of parameter $(1-\delta)^{\beta_*}$.

As in the proof of Theorem 2, the end of a block of random duration is "decided" by the public randomization device, and this thus gets known only *after* reports have been submitted.

Again, the durations of the successive blocks are independent r.v.'s. For $k \geq 1$, we denote by $\tau[k]$ the first round of block k. Note that an even block k starts in round $\tau[k]$, after reports have been submitted, while an odd block k starts before or after reports are submitted in round $\tau[k]$, depending on whether block $k-1$ was a punishment block, or a regular block.

Except during punishment blocks, reports under σ^i are always truthful, even if a deviation from σ was observed in the past. In punishment blocks, reports are fully uninformative.

In any odd block, actions are specified by ρ^i_*, whenever player i reported truthfully in the current round n: $m^i_n = s^i_n$.

The nature of an even block k is dictated by a direction $\lambda[k] \in \Lambda$. If $\lambda[k] \in B(-e^i, \varepsilon_3)$ for some i, block k is a i-punishment block, and players switch to $\sigma[i, m_{\tau[k]}]$. (recall – reports $m_{\tau[k]}$ in round $\tau[k]$ have been submitted before the block starts). If instead $\lambda[k] \in \tilde{\Lambda}$, actions in block k are given by $\rho[k]$ (following a truthful report, as for transition blocks). The direction $\lambda[k]$ is updated from one even block to the next even block, together with targets $z[k]$ and $w[k]$ in $\mathbf{R}^I$, much as in the proof of Theorem 2. This updating process is discussed next.

D.4 Preliminaries: IR

The equilibrium construction also relies on an elaborate version of Lemma 15, which we now introduce. Given $T \in \mathbf{N}$, $x : M \times Y^T \to \mathbf{R}^I$, $\delta < 1$ and $m \in M$, we denote by $G(m, \delta, x, T)$ a discounted T-stage version of $G(T, x, \theta_*)$ without communication and initial state $m \in M$. That is, in each round $n = 1, \ldots, T$, players first observe their private states (s^i_n) and 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

$$\frac{1}{1-\delta^T} \left\{ \sum_{n=1}^{T} (1-\delta)\delta^{n-1} r(s_n, a_n) + (1-\delta)\delta^T x(m_1, \vec{y}) + (1-\delta)\delta^T \theta_*(s_{T+1}) \right\},$$

with $\vec{y} = (y_1, \ldots, y_T)$.

Lemma 16 *For all $\varepsilon > 0$, there is $\beta < 1$, a constant $\kappa \in \mathbf{R}$, and $\bar{\delta} < 1$ such that, for every player $i \in I$ and discount factor $\delta \geq \bar{\delta}$, the following holds. For $T = (1-\delta)^{-\beta}$, there exists $x[i] : M \times \Omega_{pub}^T \to \mathbf{R}^I$ with $\|x[i]\| \leq \kappa T$ and $x^i[i](\cdot) \geq 0$, and $\gamma[i] \in \mathbf{R}^I$ such that (i) $\|\gamma[i] - w\| < \frac{\varepsilon}{2}$ and (ii) γ is a sequential equilibrium payoff of $G(m, \delta, x, T)$ for every $m \in S$.*

Plainly, x and γ can then be chosen such that $\|\gamma[i] - w\| < \varepsilon$ and $x^i[i] > \frac{\varepsilon}{2}$. Given δ, i and m, we denote by $\sigma[i, m]$ the equilibrium profile in $G(m, \delta, x[i], T)$ with payoff $\gamma[i]$.

Proof. ■

D.4.1 Preliminaries: IC

We here apply Athey Segal (2013) to construct transfers to be used in regular blocks. Because a player may be indifferent between different reports, these transfers cannot be defined independently of the discount factor, and the definition should take into account that transitions between blocks perturb incentives.

For given $x : S \times S \to \mathbf{R}^I$, $\beta < 1$ and $(\bar{s}, s) \in S \times S$, $G_\beta((\bar{s}, s), x, \delta, \gamma)$ is the δ-discounted version of the game in which in stage n (i) the game stops with probability $\xi := (1-\delta)^\beta$ with final payoff $\gamma + (1-\delta)\theta_*(s)$, (ii) and the stage payoff in stage n is otherwise given by $r(s_n, a) + x(m_{n-1}, m_n)$, (iii) the initial state in stage 1 is s (and $\bar{s}$ is a "fictitious" stage 0 state).

Fix a positive direction $\lambda \in \Lambda$, and consider the MDP in which players cooperate to maximize the λ-weighted sum of discounted payoffs in $G_\beta((\bar{s}, s), 0, \delta, \gamma)$, with value $V_{\delta,\lambda}(\bar{s}, s)$ the value of this MDP. Observe that $V_{\delta,\lambda}$ does depend on γ, but the optimal policy is independent of γ (and of $(\bar{s}, s)$). The original argument of Blackwell does not directly apply (because transitions depend here on δ), yet it adapts and there is a fixed policy which is optimal for all δ close to 1. We denote this policy by ρ_λ. Note that ρ_λ maximizes $\lambda \cdot \mathbf{E}_{\mu_\rho}[(1-\xi) r(s, a) + \xi \theta_*(s)]$ over the set of all policies and therefore also $\mathbf{E}_{\mu_\rho}[\lambda \cdot r(s, a)]$.

We next specialize Athey Segal (2013) to the present setup. Define first the (VCG type) transfers $\bar{x} : S \to \mathbf{R}^I$ by

$$\lambda^i \bar{x}^i(s) := \sum_{j \neq i} \lambda^j r^j(s^j, \rho_\lambda(s)).$$

Hence, for each $s \in S$ and $i \in I$, the report $m^i = s^i$ maximizes the expectation of

$$\lambda^i (1-\delta) \left\{ r^i(s^i, \rho_\lambda(m^i, s^{-i})) + x^i(m^i, s^{-i}) \right\} + \delta V_{\delta,\lambda}(s'),$$

where the expectation is over $s' \in S$.

For $s \in S$, $T^i(s) := \mathbf{E}_{s,\rho_\lambda}\left(\sum_{n=1}^{\tau-1} \delta^{n-1} x^i(s_n)\right)$ is then the expected transfer to i until the stage τ at which the game ends.

Next, for $j \in J$, $\tilde{s}^j \in S^j$ and $\bar{s} \in S$, we define

$$\Delta T^j(\bar{s}, \tilde{s}^j) = \mathbf{E}_{s^{-j} \sim p_{\bar{s},\rho_\lambda(\bar{s})}} T^j(\tilde{s}^j, s^{-j}) - \mathbf{E}_{s \sim p_{\bar{s},\rho_\lambda(\bar{s})}} T^j(s).$$

Note that, thanks the irreducibility assumption, ΔT^j is bounded, uniformly over $\delta < 1$.

We finally define the transfers $x_{\delta,\lambda} : S \times S \to \mathbf{R}^I$ by the formula

$$\lambda^i x^i(\bar{s}, s) = \lambda^i \Delta T^i(\bar{s}, s^i) - \frac{1}{I-1} \sum_{j \neq i} \lambda^j \Delta T^j(\bar{s}, s^j),$$

so that $\lambda \cdot x_{\delta,\lambda}(\cdot) = 0$ holds by construction.

We denote by $\gamma_{\delta,\lambda}(\bar{s}, s)$ the payoff induced by ρ_λ in $G_\beta((\bar{s}, s), \delta, x_{\delta,\lambda}, \gamma)$. Athey Segal (2013) proves that the transfers $x_{\delta,\lambda}$ are IC in the dynamic game. To be specific, for each $\delta < 1$, $i \in I$, $\bar{s} \in S$ and $s^i \in S^i$, the truthful report $m^i = s^i$ maximizes the expectation of

$$(1-\delta)\left(r^i(s^i, \rho_\lambda(m^i, s^{-i})) + x^i(\bar{s}, (m^i, s^{-i})) + \delta \gamma_{\delta,\lambda}((m^i, s^{-i}, t)\right).$$

We next pick $\gamma \in \mathbf{R}^I$ to be the unique vector such that $\gamma = \mathbf{E}_{\mu[\rho_\lambda]}\left[\gamma_{\delta,\lambda}(\bar{s}, s)\right]$, which we will denote by $\gamma^*_{\delta,\lambda}$. With this choice, $\gamma^*_{\delta,\lambda}$ is also given by

$$\gamma^*_{\delta,\lambda} = \mathbf{E}_{\mu[\rho_\lambda]}\left[(1-\xi) r(s, \rho[\lambda](s)) + \xi \theta_*(s)\right].$$

In particular, $\lim_{\delta \to 1} \lambda \cdot \gamma^*_{\delta,\lambda} = \bar{k}(\lambda)$.

Finally, we define the δ-relative values $\theta_{\delta,\lambda} : S \times S \to \mathbf{R}^I$ by

$$\gamma_{\delta,\lambda}(\bar{s}, s) = \gamma^*_{\delta,\lambda} + (1-\delta)\theta_{\delta,\lambda}(\bar{s}, s).$$

Note that $\theta_{\delta,\lambda}$ is uniformly bounded for $\delta < 1$.

We denote by κ_λ a common bound on $\|\theta_{\delta,\lambda}\|$ and $\|x_{\delta,\lambda}\|$ for all $\delta < 1$.

Note that, given $\varepsilon > 0$, subtracting a constant vector from $x_{\delta,\lambda}$, one may assume that $\lambda \cdot x_{\delta,\lambda}(\cdot) \leq -\frac{\varepsilon}{2}$ and $\lambda \cdot \gamma^*_{\delta,\lambda} > \bar{k}(\lambda) - \varepsilon$ for all δ close to one.

When $\lambda \in \Lambda$ is a non-necessarily positive direction, the previous construction and the results are still valid, provided one restricts attention to those policies $\rho : \times_{i \in I(\lambda)} S^i \to A$.

D.4.2 Preliminaries: the parameters

Given Z, pick first $\eta > 0$ such that Z_η is contained in the interior of $\tilde{\mathcal{H}}$, and $\varepsilon_0 > 0$ such that $\max_{Z_\eta} \lambda \cdot z < \tilde{k}(\lambda) - 2\varepsilon_0$ for all directions $\lambda \in \Lambda$.

Apply Lemma **??** with $\varepsilon := \varepsilon_0$. This yields $\beta < 1$, κ_{IR} and $\bar{\delta}_{IR} < 1$ such that the conclusion of that lemma hold.

Next, pick $\varepsilon_{IR} < \varepsilon_0/\kappa_{IR}$, and set $\tilde{\Lambda} := \Lambda \setminus \cup_i B(-e^i, \varepsilon_{IR})$. For $\lambda \in \tilde{\Lambda}$, set $\varepsilon_\lambda := \varepsilon_0/\kappa_\lambda$. Since $\tilde{\Lambda}$ is compact, there is a finite subset Λ_f of $\tilde{\Lambda}$ such that $\cup_{\lambda \in \Lambda_f} B(\lambda, \varepsilon_\lambda)$ contains $\tilde{\Lambda}$. Set then $\kappa_{IC} := \max_{\lambda \in \Lambda_f} \kappa_\lambda$ and note that $\tilde{\Lambda}$ is compact.

Fix κ large enough, and pick $\beta_* < \beta$. Finally, $\varepsilon_1 < \varepsilon_0$, $\varepsilon := \varepsilon_1/\kappa$ to get $\bar{\zeta}$, and then fix $\bar{\delta}$ large enough (exact conditions to be computed). (on utilise $\kappa_0 \geq 3\|\theta\|$, $\frac{1-\delta}{\delta\xi_*} \leq 1$, $\|z - \gamma\| + \|x\| \leq \kappa_0/2$.) Attention: demander $\lambda x_{\delta,\lambda} < 0$, et id dans les directions $-e^i$.

D.4.3 The updating process

The description is recursive. Consider first an even block $k+1$, starting in round $n := \tau[k+1]$. We define first $w[k+1]$ through the equality

$$\tilde{w}[k+1] = \xi_* w[k+1] + (1 - \xi_*) z[k],$$

where $\xi_* = (1-\delta)^{\beta_*}$, and

$$\tilde{w}[k+1] = \frac{1}{\delta} z[k] - \frac{1-\delta}{\delta} v[k] + \frac{1-\delta}{\delta} x_*(y_{n-1}, m_n).$$

Given $w[k+1]$, 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 pick one of the directions $\lambda \in \Lambda_f$ such that $\lambda[k+1] \in B(\lambda, \varepsilon_\lambda)$, and we set $\rho[k+1] = \rho_\lambda$, $x[k+1] = x_{\delta,\lambda}$ and $\theta[k+1] = \theta_{\delta,\lambda}$, as defined in Section D.4.1. Next,

$$z[k+1] := w[k+1] + (1-\delta)\left(1 + \frac{1-\delta}{\delta\xi_*}\right)\theta_*(m_n) - (1-\delta)\theta[k+1](m_n).$$

If instead $\lambda[k+1] \in B(-e^i, \varepsilon_3)$ for some $i \in I$, we set

$$z[k+1] := w[k+1] + (1-\delta)\left(1 + \frac{1-\delta}{\delta\xi_*}\right)\theta_*(m_n).$$

Consider now an odd block, $k+1$, which starts in round $n := \tau[k+1]$. The choice of $\lambda[k+1]$ is irrelevant. Assume first that the previous block k was regular. We define $w[k+1]$ through the equality

$$\tilde{w}[k+1] = \xi w[k+1] + (1 - \xi) z[k],$$

with $\xi = (1-\delta)^\beta$ and

$$\tilde{w}[k+1] := \frac{1}{\delta} z[k] - \frac{1-\delta}{\delta} v[k] + \frac{1-\delta}{\delta} x[k](\omega_{pub,n-1}, m_n).$$

Next, we set $z[k+1] := w[k+1]$.

If instead $\lambda[k] \in B(-e^i, \varepsilon_3)$ for some $i \in I$, we set

$$w[k+1] = z[k+1] := \frac{1}{\delta^T} z[k] - \frac{1-\delta^T}{\delta^T} \gamma[i] + (1-\delta) x[i](m_{\tau[k]}, y_{\tau[k]}, \ldots, y_{\tau[k+1]-1}).$$

The process is initialized as follows. Given $\bar{z} \in Z_\eta$, we set $z[1] := \bar{z} - (1-\delta)\mathbf{E}_p[\theta_*(s)]$ and $w[1] := \bar{z}$.

D.4.4 The strategies

We now complete the description of σ. Consider a block k and the full history $h_{\tau[k]}$ leading to it, and fix a player i.

If k is odd, σ^i has already been defined, except following a lie: in all rounds between $\tau[k]$ and $\tau[k+1]$ (including in round $\tau[k]$ if block $k-1$ was not regular), report of i is truthful following any private history and the action of i following h_n^i is given by ρ_* whenever $m_n^i = s_n^i$.

If now k is even, and block k is regular, player i behaves as in a transition block, except that actions are now dictated by $\rho[k]$.

Assume finally that block k is an i_*-punishment block. If the report of i in round $\tau[k]$ was truthful ($m^i_{\tau[k]} = s^i_{\tau[k]}$) player i plays $\sigma^i[m_{\tau[k]}, i_*]$ up to round $\tau[k+1] = \tau[k] + T$ (and forgets the history up to round $\tau[k]$). If instead player i lied about his state in the initial state of the punishment phase, player i plays a best reply to $\sigma^{-i}[m_{\tau[k]}, i_*]$ in the game $G((s^i_{\tau[k]}, m^{-i}_{\tau[k]}, \delta, x[i_*], T)$ (including following deviations from this best-reply).

Fix any private history h_n^i of player i ending with a truthful report $m_n^i = s_n^i$, with round n not belonging to a punishment block. This private history may be inconsistent with σ (if *e.g.* earlier public signals are inconsistent with the mixed action prescribed by σ). Yet, we let the belief of player i assign probability 1 to reports of $-i$ being truthful.[41] That is, observable deviations are interpreted as deviations in the action phase.

The continuation payoff of player i at h_n^i only depends on $\omega_{\text{pub},n-1}$, s_n and on the current value $\mathcal{T}_n$ of the auxiliary family $(z_n, w_n, \rho_n, x_n, v_n)$ used in round n (see the proof of Theorem 2), and we denote by $\gamma^i(\omega_{\text{pub}}, s; \mathcal{T})$ this continuation payoff.

[41] Here, players only report their current state, hence successive reports of player j cannot be inconsistent.

This observation also applies at the *first* round of a punishment block, and we also denote by $\gamma(\omega_{\mathrm{pub},n}, s_n; \mathcal{T}_n)$ $(n = \tau[k])$ the continuation payoff induced by σ at that round.

We complete the definition of σ as in Theorem 2. Given a private history h_n^i ending with a lie $m_n^i \neq s_n^i$, where n does not belong to a punishment phase, we let σ^i prescribe at h_n^i an action which maximizes the expectation of $(1-\delta)r^i(s_n^i, a_n) + \delta\gamma^i(\omega_{\mathrm{pub},n}, s_{n+1}; \mathcal{T}_{n+1})$.

In order to check that σ is indeed well-defined, we simply need to check that $w[k] \in Z_\eta$ for all even k.

Lemma 17 *For k even, $w[k] \in Z_\eta$. For k odd, one has $w[k] \in Z_{\eta'}$, where $\eta' := \eta + (1-\delta)\kappa_0\left(1 + \dfrac{1}{\delta\xi_*}\right)$.*

The proof is similar to that of Lemma **??**, but the proof is slightly more involved.

Proof. The proof goes by induction over k. Consider block $k+1$ where k is odd, and assume that $\|w[k] - z[k]\| \leq (1-\delta)\kappa_0$. Then

$$w[k+1] - z[k] = \frac{1}{\xi_*}\left(\tilde{w}[k+1] - z[k]\right) = \frac{1-\delta}{\delta\xi_*}\left(z[k] - v_* + x_*\right).$$

Hence $\|w[k+1] - w[k]\| \leq \frac{1-\delta}{\delta\xi_*}\kappa_0 + \|w[k] - z[k]\|$. Since $w[k] \in Z_\eta$, $w[k+1] \in Z_{\eta'}$ as desired.

Consider next block $k+1$ with even k. Assume $w[k] \in Z_{\eta'}$ and $\|w[k] - z[k]\| \leq (1-\delta)\kappa_0$. If block k was a punishment block, then

$$w[k+1] - z[k] = \frac{1-\delta^T}{\delta^T}\left(z[k] - \gamma[i]\right) + (1-\delta)x[i].$$

Hence

$$\begin{aligned}
\|w[k+1] - w[k]\| &\leq \frac{1-\delta^T}{\delta^T}\|z[k] - \gamma[i]\| + (1-\delta)\|x[i]\| + \|z[k] - w[k]\| \\
&\leq \frac{1-\delta^T}{\delta^T}\left(\|z[k] - \gamma[i]\| + \|x[i]\|\right) + \|z[k] - w[k]\|.
\end{aligned}$$

Picking $\bar{w}[k]$ a point in Z_η with $\|w[k] - \bar{w}[k]\| \leq \eta' - \eta$, this implies

$$\begin{aligned}
\|w[k+1] - \bar{w}[k] &\leq \frac{1-\delta^T}{\delta^T}\kappa_0 + (1-\delta)\kappa_0 + \eta' - \eta \\
&\leq \zeta := \left(\frac{1-\delta^T}{\delta^T} + 2(1-\delta) + \frac{1-\delta}{\delta\xi_*}\right)\kappa_0.
\end{aligned}$$

On the other hand,

$$\lambda[k] \cdot (w[k+1] - \bar{w}[k]) = \lambda[k] \cdot (w[k+1] - z[k]) + \lambda[k] \cdot (z - w[k]) + \lambda[k] \cdot (w[k] - \bar{w}[k]).$$

Since $\lambda[k] \cdot (z[k] - \gamma[i]) \leq -\bar{\varepsilon} \times \frac{1-\delta^T}{\delta^T}$ and $\lambda[k] \cdot x[i] \leq 0$, this yields

$$\lambda[k] \cdot (w[k+1] - \bar{w}[k]) \leq -\bar{\varepsilon}\frac{1-\delta^T}{\delta^T} + 2\kappa_0(1-\delta) + \frac{1-\delta}{\delta\xi_*}\kappa_0 \leq -\tilde{\varepsilon}\xi/\kappa_0.$$

(use here $\bar{\varepsilon}$ comes from date, $\tilde{\varepsilon} < \bar{\varepsilon}$, δ such that $(\bar{\varepsilon} - \tilde{\varepsilon})\frac{1-\delta^T}{\delta^T} \geq 2(1-\delta)(\tilde{\varepsilon} + \kappa_0) + \frac{1-\delta}{\delta\xi_*}(\tilde{\varepsilon} + 1))$. Hence $w[k+1] \in Z_\eta$, and note that $z[k+1] = w[k+1]$.

The computations in the case where k was a regular block are similar. ■

The rest of the proof is essentially identical to that of Theorem 2.

E Proofs for the correlated case

Proof of Lemma 6. Assumptions **2**'–**1**' are the exact counterparts of Assumptions **2**–**1**, so that the result follows exactly as in the proof of Theorem 3. ■

Proof of Theorem 5. Given Lemma 6, we may focus on the reporting step, and then augment the resulting transfers with those ensuring that players do not want to deviate at the action step (on and off-path for all $\lambda \neq \pm e^i$, and for $j \neq i$ in case $\lambda = \pm e^i$; on path only if $\lambda = \pm e^i$ and $j = i$).

At the reporting step, we must distinguish as usual between coordinate and non-coordinate directions. It suffices to consider non-coordinate directions with only two non-zero coordinates λ^i, λ^j. Fix $\rho \in \Xi$ throughout. Because of detectability ($\pi^{\bar{m},\bar{y},\bar{a}}(\cdot \mid \hat{c}) \notin \mathrm{co}\,\mathcal{R}^i(\bar{m},\bar{y},\bar{a})$, implied by Assumption **6.1**), there exists transfers x^i that ensure that truthful reporting by player i is strictly optimal. Because of weak identifiability (invoking **6.2** if $\mathrm{sgn}(\lambda^i) = \mathrm{sgn}(\lambda^j)$ and **6.1** otherwise), we can apply Lemma 2 of Kosenok and Severinov (2008) –which relies on the results of d'Aspremont, Crémer and Gérard-Varet– and conclude that these transfers can be chosen so that $\lambda \cdot x(\cdot) = 0$.

For direction $\lambda = \pm e^i$ (considering an arbitrary $\rho \in \Xi$ if $\lambda = e^i$, and $\rho = \underline{\rho}^i$ if $\lambda = -e^i$), we set $x^i = 0$ and use Assumption **5** to conclude that there exists transfers x^j, $j \neq i$, so that player j has incentives to tell the truth. Given that $\lambda^j = 0$ for all $j \neq i$, we have $\lambda \cdot x(\cdot) = 0$. ■