Cowles Foundation
for Research in Economics
at Yale University

Cowles Foundation Discussion Paper No. 1933

TRUTHFUL EQUILIBRIA IN DYNAMIC BAYESIAN GAMES

Johannes Hörner, Satoru Takahashi and Nicolas Vieille

December 2013

Truthful Equilibria in Dynamic Bayesian Games

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

December 20, 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 be correlated and values interdependent. The focus is on equilibria in which players report truthfully. The characterization generalizes that for repeated games, reducing the analysis to static Bayesian games with transfers. With correlated types, results from mechanism design apply, yielding a folk theorem. With independent private values, the restriction to truthful equilibria is without loss, except for the punishment level; if players withhold their information during punishment-like phases, a "folk" theorem obtains also.

Keywords: Bayesian games, repeated games, folk theorem.

JEL codes: C72, C73

1 Introduction

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

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

[*]Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, johannes.horner@yale.edu.
[†]National University of Singapore, ecsst@nus.edu.sg.
[‡]HEC Paris, 78351 Jouy-en-Josas, France, vieille@hec.fr.

have been obtained by means of increasingly elaborate constructions, starting with Athey and Bagwell (2008) and culminating with Escobar and Toikka (2013).[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) 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.

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

[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 equilibrium payoff set.

[2]The one exception is the lowest equilibrium payoff in Renault, Solan and Vieille (2013), who also characterize Pareto-inferior "babbling" equilibria.

To sum up, our contribution is twofold. First, we provide a family of one-shot games with transfers that reduce the analysis from a dynamic infinite-horizon game to a static game. Unlike the one-shot game of FL and HSTV (special cases of ours), this one-shot game is Bayesian. Each player makes a report, then takes an action; the transfer is then determined. This reduction provides a bridge between dynamic games and Bayesian mechanism design. As explained below, its payoff function is not entirely standard, raising interesting new issues for static mechanism design. Nonetheless, well-known results can be adapted for a wide class of dynamic games. This is our second contribution: under either independent private values, or correlated types, the analysis of the one-shot game yields an equilibrium payoff set that is best possible, except for the definition of individual rationality.

Specifically, when types are independent (though still possibly affected by one's own action), and payoffs are private, all Pareto-optimal payoffs that are individually rational –in the sense of dominating the stationary minmax payoff– are limit equilibrium payoffs, provided monitoring satisfies standard identifiability conditions. Insisting on truthfulness has a cost in terms of individual rationality: as discussed below, the stationary minmax payoff does not generally coincide with the lowest minmax payoff in the dynamic game. But this is the only restriction imposed: leaving aside individual rationality, we show that the payoff set attained by truthful equilibria is actually equal to the limit set of *all* Bayes Nash equilibrium payoffs, whichever message sets one chooses. In other words, in the revelation game in which players commit to the map from reports to actions, but not to current or future reports, there is no loss of generality in restricting attention to truthful equilibria. In this sense, the revelation principle extends when players are patient enough. Beyond generalizing the results of Athey and Bagwell, as well as Escobar and Toikka, this characterization has some interesting consequences. For instance, when actions do not affect transitions, the invariant distribution of the Markov chain is a sufficient statistic for the Markov process, as far as this equilibrium payoff set is concerned, leaving individual rationality aside.

When types are correlated, then all feasible and individually rational payoffs can be obtained in the limit (again, under suitable identifiability conditions). The "spanning" condition familiar from mechanism design with correlated types must be stated in terms of *pairs* of states: more precisely, player $-i$'s current and *next* state must be sufficiently informative about player i's current and *previous* state.

In Section 6.4, we elaborate on individual rationality in the case of independent private values. The failure of truthful equilibria to attain payoffs as low as the minmax payoff in the dynamic game should come as no surprise: after all, the same holds for public equilibria in repeated games with imperfect public monitoring. In this special case, our characterization yields the same payoff set as Fudenberg, Levine and Maskin (1994, hereafter FLM). Yet there is a natural class of monitoring structures for which FLM's payoff set coincides with the set of all sequential equilibrium payoffs. Namely, this is the case when public signals have a product structure. Similarly, we can build on truthful equilibria to obtain an exact characterization of all Bayes Nash equilibrium payoffs (as $\delta \to 1$) when monitoring has a product structure. This requires considering equilibria that are truthful *except* during punishment-like phases, in which meaningful communication is suspended.

Hence, conclusive characterizations are obtained under independent private values as well as correlated types. This mirrors the state of affairs in static mechanism design. In fact, our results are obtained by applying familiar techniques to the one-shot game, developed by Arrow (1979) and d'Aspremont and Gérard-Varet (1979) for the independent case, and d'Aspremont, Crémer and Gérard-Varet (2003) in the correlated case.

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

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

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

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

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

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

There is also a literature on undiscounted zero-sum games with such a Markovian structure, see Renault (2006), which builds on ideas introduced in Aumann and Maschler (1995). Not surprisingly, the average cost optimality equation plays an important role in this literature as well. Because of the importance of such games for applications in industrial organization and macroeconomics (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 HSTV for an asymptotic analysis, and Pęski and Wiseman (2013) for the case in which the time lag between consecutive moves goes to zero. There are some similarities in the techniques used, although incomplete information introduces significant complications.[3]

More related are the papers by Escobar and Toikka, already mentioned, Barron (2013)

[3]Among others, HSTV (as before FLM) rely on the equilibrium payoff set being full-dimensional, an assumption that fails with independent private values, as explained.

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

2 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. We assume that there are at least two players. 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 $s_n^i \in S^i$;
2. Players simultaneously make reports $(m_n^i)_{i=1}^I \in \times_i M^i$, where M^i is a finite set. Depending 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 $[0, 1]$;[4]
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 $Y(\bar{a}) \subseteq Y$.[5] 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 state profile

[4] 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.

[5] 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.

s. This is consistent with perfect monitoring. Note that actions might affect transitions.[6] The irreducibility of the Markov chain is a strong assumption, ruling out among others the case of perfectly persistent states (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.[7,8]

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 state profile $t = s_{n+1}$ 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

[6] 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.

[7] 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.

[8] 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.

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, states are independently distributed. A formal 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.[9] 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.[10]

In fact, our main characterization result extends immediately to the case in which monitoring is private, rather than public; see Section 5.0.3 for a discussion. As we focus on public monitoring for the applications that are considered in Sections 6 and 7, we have refrained from such generality here.

Monetary transfers are not allowed. We view the stage game as capturing all possible interactions among players, and there is no difficulty in interpreting some actions as monetary transfers. In this sense, rather than ruling out monetary transfers, what is assumed here is limited liability.

The game defined above allows for public communication among players. In doing so, we follow most of the literature on Markovian games with private information, see Athey and Bagwell (2001, 2008), Escobar and Toikka, Renault, Solan and Vieille, etc.[11] As in static

[9]With this interpretation, pointed out by AS, interdependent values with observable payoffs reduce to private values *ex post*, as conditional on a player's entire information, a player's payoff does not depend on the other players' types. It would then be natural to allow for a second round of messages at the end of each period.

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

[11]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.

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 to truthful behavior.

Yet this is precisely the types of equilibria that we will focus on. 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

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.[12] 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; Renault, Solan and Vieille) 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: with interdependent values, the uninformed player's payoff depends on the informed player's type, so that he cannot resist adjusting his action to the message he receives. This might hurt the informed player, who cannot be statistically disciplined into truth-telling, given independent types.

In our dynamic environment as well, positive results will obtain as soon as we impose private values or relax independent types.

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). 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

[12]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, and Bressaud and Quas (private communication) have determined the optimal strategies for values of p up to $\sim .7323$.

	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.

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. To achieve lower equilibrium payoffs, truth-telling must be abandoned, at least during punishments. As follows from Theorem 4 below, it is indeed possible to drive player 2's payoff down to his minmax payoff of 0 in equilibrium, as $\delta \to 1$.

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 (omitted), and are as follows. If $s_n^1 = k > 0$, then $s_n^2 = 0$ and $s_{n+1}^1 = s_n^1 - 1$. If $s_n^1 = 0$, then $s_n^2 = 1$ and s_{n+1}^1 is drawn randomly (and uniformly) from S^1. In words, s_n^1 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, players $-i$'s states) be as rich as player i's, unlike in static mechanism design with correlated types (the familiar "spanning condition" of Crémer and McLean, 1988, generically satisfied if only if $|S^{-i}| \geq |S^i|$). Two states for one player can be enough to cross-check the reports of an opponent with many more states, provided that states in later rounds are informative enough.

Second, the long-term dependence of the stochastic process implies that one player's report should not always be evaluated on the fly. It is better to hold off until more evidence is collected. Note that this is not the same kind of delay as the one that makes review strategies effective, taking advantage of the central limit theorem to devise powerful tests even when signals are independently distributed over time (see Radner, 1986; Fang and Norman, 2006; Jackson and Sonnenschein, 2007). It is precisely because of the dependence 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 of the sequence? This question appears to be open in general. In the case of transitions that do not depend on actions, it has been raised by Blackwell and Koopmans (1957) and answered by Gilbert (1959): it is enough to consider the next $2|S^i| + 1$ values of the sequence $(s^{-i}_{n'})_{n' \geq n}$.[13]

At the very least, when types are correlated and the Markov chain exhibits time dependence, it is useful to condition player i's continuation payoff given his report about s^i_n on $-i$'s next private state, s^{-i}_{n+1}. Because this suffices to obtain sufficient conditions analogous to those invoked in the static case, we will limit ourselves to this conditioning.[14]

4 Truthful Equilibria

Given $M := \times_{i \in I} M^i$, 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 $(\mathfrak{m}^i, \mathfrak{a}^i) = (\mathfrak{m}^i_n, \mathfrak{a}^i_n)_{n \in \mathbb{N}}$ with $\mathfrak{m}^i_n : H^i_n \times S^i \to \Delta(M^i)$, and $\mathfrak{a}^i_n : H^i_n \times S^i \times M \to \Delta(A^i)$, which specify i's report and action as a function of his private information, his current state and

[13] 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).

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

the report profile in the current round.[15] 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, FL provide a convenient algorithm to describe and study a subset of equilibrium payoffs –perfect public equilibrium payoffs. A perfect public 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. Here, we briefly motivate its main ingredients.

FL's algorithm is remarkable in its parsimony: as its proof makes clear, the past, that is, the public history leading to a given period, can be summarized by some value of λ.

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

This parameter can be interpreted as the relative standings of the players, determining who should be punished and who should be rewarded. Meanwhile, the future is summarized by the transfer x, converting instantly the realized signal into a continuation payoff.

We must add parameters. Because players' incentives to reveal their state depend on their beliefs about the other players' states, we must keep track of these beliefs, in addition to the players' relative standings. This is where the focus on truthful equilibria is powerful: because the other players' last report discloses their last type, it subsumes their past reports. In addition, because their action choices are a function of the public history, this last report also pins down player i's belief about their last (mixed) action, which matters for his beliefs. Finally, because the private state profile follows a Markov process, there is no need for player i to keep track of his own private history, beyond his last state and realized action.

To summarize, player i's beliefs can be summarized by some (public) parameters: these correspond to the last report made by his opponents as well as to the last public signal; and by his private information, namely, his previous state, his previous choice of action and his current state. Hence, the Bayesian game will be parametrized by some $(m_{n-1}, y_{n-1}) \in M \times Y$, which is public, as well as by player i's private information $(s^i_{n-1}, a^i_{n-1}, s^i_n) \in (S^i)^2 \times A^i$.

This is why the natural choice for the message space is $M^i = (S^i)^2 \times A^i$. It allows player i 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; the one-shot deviation principle does not apply here. Players $-i$ cannot detect such a deviation, which is "on-schedule," to borrow Athey and Bagwell (2008)'s terminology. 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^i_{n-1}, 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.

Hence, we must enlarge the type space, and we must also enlarge the set of parameters that summarizes the past, to account not only for λ, but also for (m_{n-1}, y_{n-1}). Similarly, we cannot simply summarize the future by a transfer determined on the fly. This is the point of Example 3. Because tomorrow's report s^{-i}_{n+1} is informative about player i's report about s^i_n, it should be included as an argument of the transfer x^i. This is fortunate, as we have just argued that player i's type is rich, including both his previous type s^i_{n-1} and his current type s^i_n. Using both s^{-i}_n and s^{-i}_{n+1} (as well as y_n) as arguments of the transfer allows us to augment the set of correlated "signals" proportionately.

Last but not least, we must adjust the payoff function of the Bayesian game. To see why this must be so, consider the case in which there is only one player, and set $\lambda = 1$. Because the purpose of transfers is to align individual and collective interests, there is no need for them. The score should then simply be the value of the Markov decision process. But clearly, if actions affect transitions, then the optimal action is not the one that maximizes the flow payoff: it must also account for the impact of this action on future states. To take this into account in the stage game, we must somehow convert the future costs and benefits from a given action into current terms. This is the essence of dynamic programming: the continuation values summarizes these costs and benefits. Here, we consider the case of low discounting, so that the appropriate functional equation is the *average cost optimality equation*, formally described in the next section. The *relative value* is the right measure to convert these costs into current units. It will be added to the flow payoff, as a basic ingredient of our Bayesian game.

A strategy $(\mathfrak{m}^i, \mathfrak{a}^i)$ is *public* and *truthful* if $\mathfrak{m}^i_n(h^i_n, s^i_n) = (s^i_{n-1}, a^i_{n-1}, s^i_n)$ (or s^i_n in the IPV 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 $u : S \to \mathbf{R}$.[16] 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 in what follows, t stands for the "next" state profile ("$t$omorrow"'s state), 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. (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.

[16] As is well known, the unichain assumption cannot be relaxed.

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.[17] 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 $\tilde{\omega}_n = (\omega_{\text{pub},n-1}, m_n)$, with transition function denoted π_ρ, and with invariant measure μ_ρ.

Let $\theta_{\rho,r+x} : \Omega_{\text{pub}} \times M \to \mathbf{R}^I$ denote the relative values of the players, obtained when applying Lemma 1 to the latter chain (and to all players).[18]

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 the one-shot Bayesian game in which states s are drawn according to p (given $\bar{\omega}_{\text{pub}}$), players submit reports m, then choose actions a, and obtain the (random) payoff

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

where (y, t) are chosen according to $p_{s,a}$ and $\omega_{\text{pub}} = (m, y)$.[19]

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;[20] player i is informed of s^i,

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

[18] There is here a slight and innocuous abuse of notation: $\theta_{\rho,r+x}$ solves the equations $v + \theta(\bar{\omega}_{\text{pub}}, m) = r(s, \rho(s)) + \mathbf{E}[x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta(\omega_{\text{pub}}, m')]$, where $v = \mathbf{E}_{\mu_\rho}[r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)]$ is the long-run payoff under ρ.

[19] 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 (MDP), 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.

[20] 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 drawn according to $\rho^{-i}(\bar{m}_c)$; hence this belief assigns to $s \in S$ the probability $p_{\bar{s},\rho(\bar{s}))}(s \mid \bar{y})$. This is the case unless $\bar{y}$ is inconsistent with $\rho^{-i}(\bar{m}_c)$; if this is the case, use the same updating rule with some other arbitrary $\tilde{a}^{-i}$ such that $\bar{y} \in Y(\tilde{a}^{-i}, \bar{a}^i)$.

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

$$r^i(s, a) + x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t^{-i}) + \theta^i_{\rho, r+x}(\omega_{\text{pub}}, m'), \tag{1}$$

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

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

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, $\rho^i(s)$ is a (not necessarily unique) optimal mixed action (Obedience).*

Requiring in addition ρ to be pure, and $\rho^i(m_c)$ to be optimal even after a lie would yield a smaller set of admissible pairs, and hence a weakening of Theorem 2 below. Yet, this weakened version would suffice to deliver all results derived in Sections 6 and 7.

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

The truth-telling condition is slightly more delicate to interpret. Consider first an outcome $\bar{\omega} \in \Omega$ such that $\bar{s}^i = \bar{m}_c^i$ 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 $r(s, a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t) + \theta_{\rho, r+x}(\omega_{\text{pub}}, m')$. 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.[21]

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 $(\bar{s}^{-i}, \bar{a}^{-i}) =$

[21]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.

$(\bar{m}_c^{-i}, \rho^{-i}(\bar{m}_c))$, 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 did misreport his previous state ($\bar{m}_c^i \neq \bar{s}^i$). 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 (1) 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 **1** below) on feasible policies in the IPV case.

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

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

- $(\rho, x) \in \mathcal{C}_0$;
- $\lambda \cdot x(\cdot) \leq 0$;
- $v = \mathbf{E}_{\mu_\rho}\left[r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)\right]$, where $\mu_\rho \in \Delta(\Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S)$ is the invariant distribution under truth-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.

We note that Theorem 2 is also valid when $M^i = S^i$ and when the definition of an admissible pair is modified in an obvious way.

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 larger 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.[22]

Two variations to this theorem are worth mentioning. First, Theorem 2 can be adapted to the case in which some of the players are short-run, whether or not such players have private information (in which case, assume that it is independent across rounds). As this is a standard feature of such characterizations (see FL, for instance), we will be brief. Suppose that players $i \in LR = \{1, \ldots, L\}$, $L \leq I$ are long-run players, whose preferences are as before, with discount factor $\delta < 1$. Players $j \in SR = \{L+1, \ldots, I\}$ are short-run players, each representative of which plays only once. We consider a "Stackelberg" structure, common in economic applications, in which long-run players make their reports first, thereupon the short-run players do as well (if they have any private information), and we set $M^i = S^i$ for the short-run players. Actions are simultaneous. Let $m^{LR} \in M^{LR} = \times_{i=1}^{L} M^i$ denote

[22] 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.

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.

Second, signals can be private. That is, we may replace Step 5 in Section 2 by: A profile $y_n = (y_n^i) \in Y := \times_i Y^i$ of private signals and the next state profile $s_{n+1} = (s_{n+1}^i)_{i \in I}$ are drawn according to some joint distribution $p_{s_n, a_n} \in \Delta(S \times Y)$. We then re-define a message m^i as including: player i's state, action and signal in the last period, and player i's current state. Transfers are then assumed to depend on the past, current and next message profile, with the restriction, as with public monitoring, that player i's transfer does not depend on his own future message, only on player $-i$'s. The definition of admissibility remains the same, given the re-defined message space, and so does the statement of the theorem.

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

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 $\mathcal{H}_0$, 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 pair (ρ, x) such that (ρ, x, v) is feasible in $\mathcal{P}_0(\lambda)$.[23] Players are required to report truthfully their type and to play (on path) according to ρ, and we update the target to $w_{\bar{m},m,a} := v + \dfrac{1-\delta}{\delta} x(\bar{m}, m, a)$ 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.[24] 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 relative values into the definition of admissible pairs.

But this raises a difficulty. In any recursive construction such as the one outlined above, continuation relative 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 pair (ρ, x) involves the relative values $\theta_{\rho,r+x}$ induced by an indefinite play of (ρ, x). This difficulty is solved by adjusting the recursive construction in such a 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 ξ. 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

[23]If v is a boundary point, λ is an outwards pointing normal to Z at v.

[24]With $w_{\bar{m},m,a}$ serving as the target payoff vector in the next, second, step.

correlation device. The parameter ξ is chosen large enough compared to $1-\delta$, ensuring that target payoffs always remain within the set Z. Yet, ξ is chosen small enough so that the continuation relative values be approximately equal to $\theta_{\rho,r+x}$: in terms of relative 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 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 pairs $(s^{-i}, \alpha^{-i}, t^{-i})$, $(\tilde{s}^{-i}, \tilde{\alpha}^{-i}, \tilde{t}^{-i})$,*

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

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

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

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.[25] 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.[26] 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.

[25] 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.

[26] 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

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 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.[27]

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 $s^i \sim \tilde{s}^i$, 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 $p_{s^i,s^{-i},a} \neq p_{\tilde{s}^i,s^{-i},a}$.

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

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

[27]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.

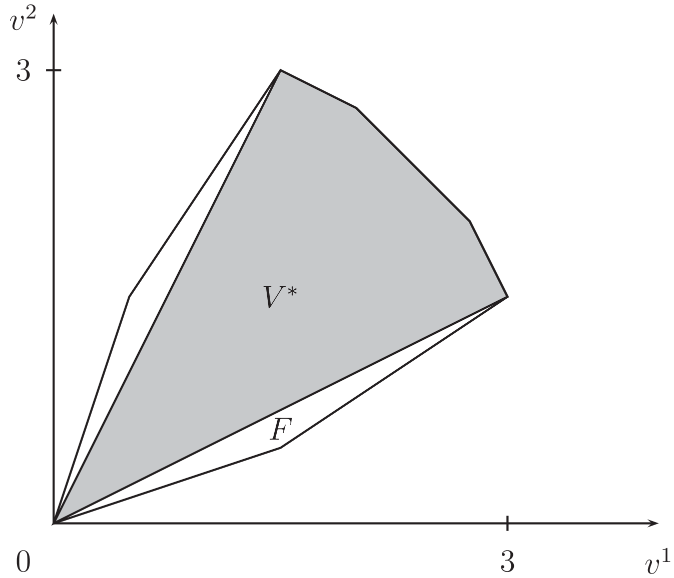

Figure 4: Incentive-compatible and feasible payoff sets in Example 4

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. In their environment with action-independent transitions and perfect monitoring, Escobar and Toikka show that all individually rational (as defined below) payoffs in $\mathrm{co}\,(Ext^{pu} \cup Ext^{po})$ are equilibrium payoffs (whenever this set has non-empty interior). Indeed, the following is easy to show.

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

In Example 4, this lower 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 AS, 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[28] $m^i_\rho : \Omega_{\text{pub}} \times S^i \to \Delta(M^i)$ 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

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

[28] 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.

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

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

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 that shows that this conclusion –that the invariant distribution is a sufficient statistic for the set of limit incentive-compatible payoffs– does not hold when values are interdependent).

6.3 Obedience and Individual Rationality

Recall that truth-telling incentives must be strict. This requires some minimal assumption on preferences. To motivate it, consider the case in which player i's types are i.i.d. over time. If the vector $r^i(s^i,\cdot)$ can be written as $cr^i(\bar{s}^i,\cdot)+d$, with $\bar{s}^i \neq s^i$, for some $c,d \in \mathbf{R}$, $c>0$, then it is clearly impossible to provide incentives for player i to strictly prefer revealing that his private state is s^i rather than $\bar{s}^i$.[29] Similarly, if $r^i(s^i,a)$ is independent of $a \in A$, player i does not care about the action profile played in that period, and strict incentives cannot be provided. This is a familiar result in repeated games, see Abreu, Dutta and Smith (1994): with constant or equivalent utility functions, it is impossible to make truth-telling incentives strict. It is necessary that, for at least two possibly mixed action profiles $\alpha,\bar{\alpha} \in A$, player i prefers α to $\bar{\alpha}$ in state s^i, and $\bar{\alpha}$ to α in state $\bar{s}^i$. Without the i.i.d. assumption, we have more leeway, as preferences are defined over infinite streams of actions. We directly state our assumption in terms of payoff asymmetry.

Assumption 1 *For all i, $s^i \neq \bar{s}^i \in S^i$, there exists $(a^{s^i,\bar{s}^i})_n, (a^{\bar{s}^i,s^i})_n \in A^{\mathbb{N}}$ such that*

$$\varliminf_{\delta\to 1} \mathbf{E}_{s_1=s^i} \sum_{n\geq 1} \delta^n \left(r^i(s_n, a_n^{s^i,\bar{s}^i}) - r^i(s_n, a_n^{\bar{s}^i,s^i}) \right) > 0 > \varliminf_{\delta\to 1} \mathbf{E}_{s_1=\bar{s}^i} \sum_{n\geq 1} \delta^n \left(r^i(s_n, a_n^{s^i,\bar{s}^i}) - r^i(s_n, a_n^{\bar{s}^i,s^i}) \right).$$

Assumption 1 implies the existence of $|S^i|$ lotteries over a set Γ^i of sequences $\{(a_n^k)_{n=1}^\infty : k = 1,\ldots,|S^i|(|S^i|-1)/2\}$ such that each type of player i has a strictly preferred lottery (as $\delta \to 1$) within that set, with no single lottery being the best one for two different types. (See Lemma 2 of Abreu, Dutta and Smith, 1994.) For simplicity we have stated Assumption 1 in terms of action profiles, but we could as well assume that there exist two distributions over sequences in $A^{\mathbb{N}}$ that have the stated property. Let us now turn to monitoring.

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

[29] It might make sense to identify $s^i, \bar{s}^i$ if utilities are equivalent, and ask player i to report the equivalence class. We do not pursue this here.

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. For simplicity, we make the assumption on all pairs of states and actions, although of course only those that are used in the construction matter.

Assumption 2 *For all $(s,a) \in S \times A$:*

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 **2** reduces to Assumptions **A1**–**A3** of Kandori and Matsushima (1998) in the case of repeated games (with the caveat that Kandori and Matsushima apply it to the relevant action profiles only).

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 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, in what follows, we use $\underline{v}^i$ as our benchmark, and let $\underline{\rho}_i$ denote a policy that achieves $\underline{v}^i$.

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**–**2**, 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

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

where the minimum is over the set of (independent) strategies by players $-i$. We denote by $\underline{\sigma}_i$ 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. First, the state-independent pure-strategy minmax payoff rules out mixed strategies. Yet mixed strategies play a key role in some applications, *e.g.* the literature on tax auditing. More disturbingly, when $\underline{v}^i > w^i$, 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.[30]

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

[30] To see this, note that 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 $k(\lambda^e)$.

$-i$ use private strategies (depending on past action choices) that would allow for harder, coordinated punishments than those assumed in the definition of the mixed-strategy minmax payoff. Private histories may allow players $-i$ to correlate play unbeknownst to i. One special case in which they do coincide is when monitoring has a product structure, which rules out such correlation.[31] As this is the class of monitoring structures for which the standard folk theorem for repeated games is a characterization of (as opposed to a lower bound on) the equilibrium payoff set, we maintain this assumption throughout this section.

Definition 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 $y = (y^1, \ldots, y^I) \in Y$, all (s, a).

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

Assumption 3 *For all i, (s, a),*

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

Note that, given product structure, Assumption **3** 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 **1** and **3** 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."[32] Nonetheless, we show that a slight extension of the set of strategies considered so far, to allow for silent play during punishment-like phases, suffices.

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

[32] 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.

Unlike in repeated games, imposing product structure does not guarantee that the minmax strategy is stationary: players $-i$ draw inferences from the public signal y^i about player i's action, hence about his private state, which can be exploited to adjust the next punishment action. Our construction relies on an extension of Theorem 2, as well as an argument inspired by Gossner (1995), based on approachability theory (Blackwell, 1956). Roughly speaking, the argument is divided in two parts. First, one must extend Theorem 2 to allow for "blocks" of T rounds, rather than single rounds, as the extensive form over which the score is computed. This part is delicate; in particular, the directions $-e^i$ –for which such aggregation is necessary– cannot be treated in isolation, as $\Lambda \setminus \{-e^i\}$ would no longer be compact, a property that is important in the proof of Theorem 2. Second, considering such a block in which player i, say, is "punished" (that is, a block corresponding to the direction $-e^i$), one must devise transfers x at the end of the block, as a function of the public history, that makes players $-i$ willing to play the minmax strategy, or at least some strategy profile achieving approximately the same payoff to player i. The difficulty, illustrated by Example 2, is that typically there are no transfers making player i indifferent over a subset of actions for different types of his simultaneously; yet minmaxing might require precisely as much. To ensure that the distribution over action profiles during the punishment phase matches the theoretical one (computed using the realized actions taken by player i), we design a statistical test that a player $j \neq i$ can pass with very high probability (by conforming to the minmax strategy, for instance), independently of the other players' strategies; and that he is very likely to fail if the distribution of his realized signals departs too much from the one that his minmax strategy would yield.[33] When testing player j, it is critical to condition on player i's realized signal, so as to incentivize player j to be unpredictable.

7 Correlated Types

We now consider the case of correlated types, as defined by Assumption **5** 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

[33] This is where the IPV assumption and product monitoring are used. It ensures that player j's minmax strategy can be taken to be independent of his private information, hence adapted to the public information.

re-define the set of relevant policies Ξ as the set of policies that achieve extreme points of F.[34,35] These are simply the policies achieving the extreme points of the feasible (limit) 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 2'(a) *For all $\rho \in \Xi$, all s, $a = \rho(s)$:*

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

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

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

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

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 **2'(a)**–**2'(b)**, 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.*

[34]If multiple policies achieve the maximum, Assumption **2'(a)** has to be understood as asserting the existence of a selection of policies satisfying the stated requirement.

[35]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.

- *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 **4**–**5** below, which cannot hold otherwise. For the case in which states are independently distributed over time, the counterpart of Theorem 4 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 a player i and $\bar{m}, \bar{a}, \bar{y}$. 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;[36] 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 (s^{-i}, y, t^{-i}), 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

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

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

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

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

[36] 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.

$p_{s^j}(s^{-j})$ 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.

As is well known, Assumption **4** 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.[37] We start with the joint distribution

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

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

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

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

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

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

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

Assumption 5 *For all $\rho \in \Xi$, all $(\bar{m}, \bar{a}, \bar{y})$,*

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

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

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

Assumption **5** combines two assumptions: any deviation by a player is detectable ($\pi^{\bar{m},\bar{a},\bar{y}}(\cdot \mid \hat{c}) \notin \mathrm{co}\,\mathcal{R}^i(\bar{m}, \bar{a}, \bar{y})$), and unilateral deviations by two players are distinguishable (this is Assumption **5.2**). 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 **5.1** 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.[38]

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 **2'(a)–2'(b)**, **4–5**, the limit set of truthful equilibrium payoffs includes V^{**}.*

As in the static case, Assumptions **4–5** are generically satisfied if $|S^{-i}| \geq |S^i|$ for all i.[39] 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.

[38]See also Hörner, Takahashi and Vieille (2013). One easy way to understand this is in terms of the cone spanned by the vectors $\pi^{\bar{m},\bar{a},\bar{y}}(\cdot \mid c^i, \hat{c}^{-i})$ and pointed at $\pi^{\bar{m},\bar{a},\bar{y}}(\cdot \mid \hat{c})$. 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 $-p(\cdot \mid c^j)$.

[39]Generically, for Assumption **4**, it suffices that $|S^{-i}|^2 \geq |S^i|^2$ for all i, while Assumption **5** 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.

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 under alternative assumptions on the communication opportunities. At one extreme, one might like to know what can be achieved without communication; at the other extreme, how to extend the analysis to the case in which a mediator is available.

References

Abreu, D., P.K. Dutta and L. Smith (1994). "The Folk Theorem for Repeated Games: A NEU Condition," *Econometrica*, **62**, 939–948.

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*, **81**, 2463–2485.

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

Barron, D. (2013). "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. (1956). "An analog of the minimax theorem for vector payoffs," *Pacific Journal of Mathematics*, **6**, 1–8.

Blackwell, D. (1962). "Discrete dynamic programming," *Annals of Mathematical Statistics*, **33**, 719–726.

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*, **81**, 1887–1934.

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*, **52**, 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, when δ is close enough to one.

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 a feasible triple (v, ρ, x) in $\mathcal{P}(\lambda)$ such that $\max_{z \in Z_\eta} \lambda \cdot z < \lambda \cdot v$ and $\lambda \cdot x(\cdot) < 0$. Using the compactness of Λ, Lemma 7 below then follows.

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 $(v, \rho, x) \in \mathcal{S}_0$ feasible in $\mathcal{P}_0(\lambda)$ and s.t.* $\max_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.

We let $\kappa_0 \in \mathbf{R}$ be 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 $(v, \rho, x) \in \mathcal{S}_0$ and every $z \in Z_\eta$.[40]

We quote without proof the following classical result, which relies on the smoothness of the boundary 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 $w \in Z_\eta$.*

Let an admissible pair $(\rho, x) \in \mathcal{C}_0$, a player $i \in I$ and $(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i)$ be given. Given $s^i \in S^i$, we denote by $\gamma^i(\bar{\omega}_{\text{pub}}, (\bar{s}^i, \bar{a}^i, s^i) \to m^i)$ the highest (interim) payoff of player i in the decision problem $D^i(\bar{\omega}_{\text{pub}}, \bar{s}^i, \bar{a}^i)$, when his state is s^i and when reporting $m^i \in M^i$. Since $(\rho, x) \in \mathcal{C}_0$, truth-telling is uniquely optimal, so there exists $\nu_{\rho,x} > 0$ such that

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

whenever $m^i \neq (\bar{s}^i, \bar{a}^i, s^i)$. We set $\nu := \min_{(v,\rho,x) \in \mathcal{S}_0} \nu_{\rho,x} > 0$.

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

We let $\varepsilon_1 \in (0, \varepsilon_0)$ be arbitrary, set $\varepsilon := \varepsilon_1/2\kappa_0$ and then let $\bar{\zeta}$ be given by Lemma 8. Next, we pick $\beta \in (0,1)$, and let $\bar{\delta} < 1$ be large enough so that for all $\delta \geq \bar{\delta}$ (i) $(1-\delta) \leq \xi\delta$, (ii) $2\kappa_0\delta\xi \leq \varepsilon_0 - \varepsilon_1$ (iii) $\dfrac{\xi}{1-\xi} \leq \dfrac{\nu}{5\kappa_0}$, where $\xi = (1-\delta)^{\beta}$.

A.2 Strategies

We let the initial state profile be commonly known and equal to $s_1 \in S$. The p.r.d. is ignored in round 1.

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 (when supplemented with appropriate beliefs).

The play is partitioned into blocks of random duration. The durations of the successive blocks are i.i.d., and follow a geometric distribution of parameter ξ. 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 ξ and $1-\xi$.

With each block k is associated a direction $\lambda[k] \in \Lambda$, and the triple $(v[k], \rho[k], x[k]) \in \mathcal{S}_0$ associated to $\lambda[k]$ by Lemma 7. The direction $\lambda[k]$ 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 exact updating process is reminiscent of that of FLM. It is best described by introducing two "target" payoffs $w[k], z[k]$ (instead of one in FLM and HSTV), with $w[k] \in Z_\eta$ for all k.[41]

Given the public history up to round $\tau_{k+1} = n+1$, the target $w[k+1]$ is defined 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 $\lambda[k+1] \in \Lambda$ be one of the directions associated to $w[k+1]$ by Lemma 8, pick

[41] The modified target $z[k]$ will be obtained from $w[k]$ by adding a correcting term. The role of the correction is to align exactly the incentives in the discounted game with switching blocks with those in the "limit" optimization program $\mathcal{P}(\lambda)$, by adjusting for the fact that the relative values depend on the policy being implemented, and on the discount factor. There is no need for such a correction in repeated/stochastic games.

$(\rho[k+1], x[k+1], v[k+1]) \in \mathcal{S}_0$ using Lemma 7, and set

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

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

The construction is initialized with $w[1] = z_*$, $\theta[0] = 0$, and an arbitrary pair $(\omega_{\text{pub},0}, m_1) \in \Omega_{\text{pub}} \times M$, where $\omega_{\text{pub},0}$ is consistent with $\rho[1]$, and m_1 is consistent with $\omega_{\text{pub},0}$ and s_1.[42]

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 this 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 $w[k] \in Z_\eta$,[43] and note that $\|w[k] - z[k]\| \leq \kappa_0(1-\delta)$. By (7) and (8),

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

so that

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

Thus,

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

Set $\zeta := 2\frac{(1-\delta)}{\delta\xi}\kappa_0$. Note that

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

(where the last inequality uses $\kappa_0\delta\xi \leq \varepsilon_0 - \varepsilon_1$). The result follows from Lemma 8, by the choice of $\lambda[k]$ and since $\zeta < \bar{\zeta}$. ■

[42] Note that $w[1]$ pins down $\lambda[1]$, $(v[1], \rho[1], x[1])$ and $\theta[1]$.

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

We denote by $\mathcal{T}_n = (z_n, w_n, \rho_n, x_n, v_n, \theta_n)$ the family of all relevant variables in round n. Thus, $w_n = w[k]$, $\rho_n = \rho[k]$, etc., if $\tau_k \leq n < \tau_{k+1}$. We stress that $\mathcal{T}_n$ is known only after the output of the randomizing device has been observed in round n.

Under σ, a player always reports truthfully at the reporting step (even if a deviation from σ was observed in the past), and, at the action step, plays according to the mixed action ρ_n whenever he reported truthfully his current state.

Fix now a player i, and a private history $h_n^i \in H_n^i \times S^i \times M$ up to the action step in round n. Assume first that his currently reported state is correct: $m_{c,n}^i = s_n^i$. Since the belief of player i assigns probability one to $s_n^{-i} = m_{c,n}^{-i}$, his continuation payoff under σ is well-defined, and only depends on $\omega_{\text{pub},n-1}$, m_n and on $\mathcal{T}_n$.[44] We denote it by $\gamma_\sigma^i(\omega_{\text{pub},n-1}, m_n; \mathcal{T}_n)$.

If instead $m_{c,n}^i \neq s_n^i$, we let σ^i prescribe any action a^i which maximizes the discounted sum of the current payoff and of expected continuation payoffs, that is, the expectation of

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

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

Theorem 2 follows from **Q1** and **Q2** below.

Q1 For given $\mathcal{T} = (z, w, \rho, x, v, \theta)$, one has $\gamma_\sigma(\bar{\omega}_{\text{pub}}, m; \mathcal{T}) = z + (1-\delta)\theta(\bar{\omega}_{\text{pub}}, m)$ for every $(\bar{\omega}_{\text{pub}}, m) \in \Omega_{\text{pub}} \times M$.[45]

In particular, the expected payoff induced by σ is equal to z_*.

Q2 The profile σ is a sequential equilibrium.

A.3 Proof of Q1

The rationale behind the twisted recursive formula (9) is the simple observation below. We place ourselves right before the p.r.d. is observed in round $n+1$.

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

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

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

[45]The payoff vector $\gamma_\sigma(\omega_{\text{pub}}, m; \mathcal{T})$ 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.

Recall that $\tilde{w}_{n+1}$ is given by (7) and is measurable w.r.t. the public information available before the random device is observed in round $n+1$. The expectation in (10) is taken with respect to the random output of the p.r.d., and the equality is an algebraic identity based on the updating formulas (7) and (8).

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} below. With probability $1-\xi$, $\mathcal{T}_{n+1} = \mathcal{T}_n$ and with probability ξ, z_{n+1} is given by (9). Thus,

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

■

We now place ourselves at the action step in round n.

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

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

Here, the expectation is taken over a_n and $h_{\text{pub},n+1}$,[46] and is computed given the belief held by player i at h_n^i, assuming all players play σ from h_n^i on. More concisely, we write

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

Proof. Since $(v_n, \rho_n, x_n) \in \mathcal{S}_0$, and since $\theta_n = \theta_{\rho_n, r+x_n}$, one has

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

[46] $h_{\text{pub},n+1}$ is the public information available at the action step in round $n+1$, and therefore also includes the outcome of the p.r.d. in round $n+1$.

Omitting again all arguments, we rewrite (11) as

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

Adding $z_n^i - v_n^i$ on both sides yields

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

where the last equality holds by 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 $\rho_n^i(m_{c,n})$ is replaced by any action a^i in its support (resp. not in its support). This implies that, when an arbitrary action a_n^i is substituted to $\rho_n^i(m_{c,n})$ when taking expectations, the conclusion of the lemma still holds with equality or a weak inequality $\leq$, depending on whether a^i belongs to the support of $\rho_n^i(m_{c,n})$ or not.

In probabilistic terms, Lemma 11 amounts to

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

where $\mathcal{H}_{\mathrm{pub},n}$ is the σ-algebra corresponding to the public information available *after* the p.r.d. is observed in round n.[47]

Of course, the continuation payoffs $\gamma_\sigma(\omega_{\mathrm{pub},n-1}, m_n; \mathcal{T}_n)$ also satisfy the recursive equation

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

Since all quantities are bounded, this implies that

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

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

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

A.4 Proof of Q2

Fix a player i. Note first that, by construction, player i has no profitable one-step deviation at the action step following a lie ($m^i_{c,n} \neq s^i_n$). Using the remark following the proof of Lemma 11, this is also true following a truthful report.

Let now $h^i_n \in H^i_n \times S^i$ be an arbitrary private history of player i at the reporting step in round n. We will prove that truth-telling is optimal. The objective of player i is to pick the report m^i_n that maximizes

$$\mathbf{E}\left[(1-\delta)r^i_n + \delta\gamma^i_\sigma(\omega_{\text{pub},n}, m_{n+1}; \mathcal{T}_{n+1})\right],$$

where the expectation is computed given the belief of player i, when facing σ^{-i}, as a function of m^i_n. By **Q1**, the expectation is also equal to

$$\mathbf{E}\left[(1-\delta)r^i_n + \delta\left(z^i_{n+1} + (1-\delta)\theta^i_{n+1}(\omega_{\text{pub},n}, m_{n+1})\right)\right],$$

which using Lemma 10 is equal to

$$\mathbf{E}\left[(1-\delta)r^i_n + \delta\tilde{w}^i_{n+1} + (1-\delta)\theta^i_n\right].$$

As in the proof of Lemma 11, the latter expectation is also given by

$$(1-\delta)\mathbf{E}\left[r^i(s_n, \rho(s^{-i}_n, m^i_n)) + x^i_n(\omega_{\text{pub},n-1}, \omega_{\text{pub},n}, s^{-i}_{n+1}) + \theta^i_n(\omega_{\text{pub},n}, s_{n+1})\right] + \mathbf{E}[z^i_n - (1-\delta)v^i_n]. \tag{13}$$

(Beware that we are taking expectations at the reporting step in round n: $\mathcal{T}_n$ is not known at h^i_n, hence the expectation over z^i_n.)

We thus need to prove that the expectation in (13) is maximal when reporting truthfully. Given an untruthful report m^i_n, we will compare the expectation, denoted $\mathbf{E}_{\text{lie}}$, when reporting m^i_n to the expectation $\mathbf{E}_{\text{truth}}$ when reporting truthfully. 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 $(v_{n-1}, \rho_{n-1}, x_{n-1}) \in \mathcal{S}_0$ (and since the belief of player i at h^i_n is deduced from ρ_{n-1}), the conditional $\mathbf{E}_{\text{truth}}$ exceeds the conditional $\mathbf{E}_{\text{lie}}$ by at least $(1-\delta)\nu$.

With probability ξ, the play switches to a new block in round n. Conditional on switching, lying may possibly improve both expectations on the right-hand side of (13). Yet, the gain in $(1-\delta)\mathbf{E}[r^i_n + x^i_n + \theta^i_n]$ is of at most $2\kappa_0(1-\delta)$, and the gain in $\mathbf{E}[z^i_n - (1-\delta)v^i_n]$ is, given (9) and the choice of $\bar{\delta}$, of at most $3\kappa_0(1-\delta)$.[48]

[48]Indeed, since $x^i(\cdot, m_c)$ is independent of m^i_c, the report m^i_n does not influence $\tilde{w}_{n+1}$.

Hence,

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

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

B First Proofs for Independent Private Values

Proof of Lemma 3. Note that the equality $\lambda \cdot \mathbf{E}_{\mu_\rho}[r(s,a)] \leq \bar{k}(\lambda)$ trivially holds for each constant policy $\rho = a \in A$ and $\lambda \in \Lambda$.

Let $\lambda \in \Lambda$. Suppose first that $\lambda \leq 0$. Consider the vector $v \in Ext^{po}$ that maximizes $\lambda \cdot v$, and the corresponding policy ρ. This policy implements the distribution $\mathbf{E}_{\mu_\rho}[\rho(s)]$ in $\Delta(A)$. Consider the constant 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 $\bar{k}(\lambda) \geq \lambda \cdot v$. Suppose next that $\lambda^i > 0$ for some $i \in I$. 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 $\hat{\lambda} \geq 0$, $\hat{\lambda} \in \Lambda$, with $\hat{\lambda}^i = 0$ whenever $\lambda^i < 0$. Such a vector is achieved by a policy that only depends on $(s^i)_{i \notin J}$, because of private values. Hence again, $\bar{k}(\lambda) \geq \lambda \cdot v$.

This implies $\text{co}\,(Ext^{pu} \cup Ext^{po}) \subset \{v \in \mathbf{R}^i : \lambda \cdot v \leq \bar{k}(\lambda)\}$, as desired. ■

B.1 Proof of Proposition 1

Given a policy $\rho : S \to \times_{i \in I} \Delta(A^i)$, we denote by p_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 $\times_{j \neq i} p^j(s^j \mid \bar{s}^j, \rho^j(\bar{s}), \bar{y})$.

Fix a weakly truthful pair (ρ, x), with $\rho : S \to \times_{i \in I} \Delta(A^i)$ 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

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

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

The next lemma is the long-run analog of a key step in the proof of Proposition 2 in AS. The logic of the proof is identical.

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

$$\tilde{x}^i(\bar{\omega}_{\text{pub}}, s)(= \tilde{x}^i(\bar{\omega}_{\text{pub}}, s^i)) = \theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, s^i) - \mathbf{E}_{\tilde{s}^i \sim p_\rho(\cdot|\bar{\omega}_{\text{pub}})}[\theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, \tilde{s}^i)]. \tag{14}$$

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

Proof.

We first argue that $\theta^i_{\rho,\tilde{x}}(\cdot) = \tilde{x}^i(\cdot)$ (up to an additive constant, as usual). It suffices to prove that $\tilde{x}^i$ solves the system

$$\tilde{x}^i(\bar{\omega}_{\text{pub}}, s^i) = \tilde{x}^i(\bar{\omega}_{\text{pub}}, s^i) + \mathbf{E}_{(\omega_{\text{pub}}, t^i) \sim p_\rho(\cdot|\omega_{\text{pub}}, s^i)}\left[\tilde{x}^i(\omega_{\text{pub}}, t^i)\right], \text{ for all } (\bar{\omega}_{\text{pub}}, s^i).$$

But this follows from the fact that

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

Next, fix $i \in I$, and $(\bar{\omega}^i, \bar{s}^i, \bar{a}^i)$. Since (ρ, ξ) is weakly truthful, for each $s^i \in S^i$, the expectation of

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

is maximized for $m^i = s^i$.[49]

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

$$r^i(s^i, \rho(s^{-i}, m^i)) + \tilde{x}^i(\bar{\omega}_{\text{pub}}, m^i) + \theta^i_{\rho,r}(\omega_{\text{pub}}, t) + \theta^i_{\rho,\tilde{x}}(\omega_{\text{pub}}, t) \tag{16}$$

is maximized for $m^i = s^i$ as well. Fix $m^i \in M^i$. Using $\theta^i_{\rho,\tilde{x}} = \tilde{x}^i$, and the definition of $\tilde{x}^i$, the expression in (16) is equal to

$$r^i(s^i, \rho(s^{-i}, m^i)) + \theta^i_{\rho,\xi}(\bar{\omega}_{\text{pub}}, m^i) + \theta^i_{\rho,r}(\omega_{\text{pub}}, t) + \theta^i_{\rho,\xi}(\omega_{\text{pub}}, t^i) - \mathbf{E}_{t^i \sim p_\rho(\cdot|\omega_{\text{pub}})}\theta^i_{\rho,\xi}(\omega_{\text{pub}}, t^i), \tag{17}$$

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

Next, observe that by definition of $\theta^i_{\rho,\xi}$, one has

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

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

Thus, the expectations of (16) and (15) differ by a constant, so that the weak truthfulness of $(\rho, \tilde{x})$ follows from that of (ρ, ξ). ■

---

[49] For concreteness, 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 according to $p_{s,\rho(s^{-i},m^i)}$, and $\omega_{\text{pub}} = (s^{-i}, m^i, y)$.

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

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

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

Proof. It is enough to check that, for any $\bar{\omega}_{\text{pub}}$ and $j \neq i$, 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 the proof of Proposition 2 in AS). ∎

Proof of Proposition 1. Let (ρ, x) be weakly truthful. Since λ is not a unit vector, there exists a solution to the system $\lambda^i + \sum_{j \neq i} \lambda^j \mu_{ji} = 0$ $(i \in I)$. Apply Lemma 12 and Corollary 6 with this choice of μ_{ij}. Then $(\rho, \bar{x})$ is weakly truthful and $\lambda \cdot \bar{x}(\cdot) = 0$. ∎

Proof of Lemma 4. Focus on a player i and fix a pair (ρ, x) with[50] $\rho : S \to A$ and $x = \Omega_{\text{pub}} \times \Omega_{\text{pub}} \to \mathbf{R}^I$.

Consider the MDP, deduced from the game, in which player i only chooses in round n which state m_n^i to report, players $-i$ report truthfully, actions are set to $a_n = \rho(s_n^{-i}, m_n^i)$, and the reward is set to $r^i(s_n^i, a_n) + x^i(\omega_{\text{pub},n-1}, \omega_{\text{pub},n})$.

This MDP $\mathcal{M}$ is best viewed as an MDP in which (i) the state space is $\Omega_{\text{pub}} \times S^i$ with elements $(\bar{\omega}_{\text{pub}}, s^i)$ (interpreted as the public outcome in the previous round and the current state of player i), (ii) the action set is $M^i = S^i$, (iii) transitions (still denoted p) are deduced from p and ρ in the obvious way, and (iv) the reward induced by action m^i in state $(\bar{\omega}_{\text{pub}}, s^i)$ is $r_\rho^i((\bar{\omega}_{\text{pub}}, s^i), m^i) + x_\rho^i((\bar{\omega}_{\text{pub}}, s^i), m^i)$ where

$$r_\rho^i((\bar{\omega}_{\text{pub}}, s^i), m^i) = \mathbf{E}_{s^{-i} \sim p(\cdot|\bar{\omega}_{\text{pub}})} r^i(s^i, \rho(s^{-i}, m^i)),$$

and

$$x_\rho^i(\bar{\omega}_{\text{pub}}, s^i, m^i) = \mathbf{E}_{s^{-i} \sim p(\cdot|\bar{\omega}_{\text{pub}}), y \sim p_{s,\rho(s^{-i},m^i)}(\cdot|\bar{\omega}_{\text{pub}})} x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}),$$

as in Section 6.2.

Note that the map x_ρ^i depends on s^i only through $[s^i]$. Conversely, let $x_\rho^i((\bar{\omega}_{\text{pub}}, [s^i]), m^i)$ be any such map. By definition of the equivalence classes $[s^i]$, and since all actions are potentially played, there exists a map $x^i : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \to \mathbf{R}$ such that

$$x^i((\bar{\omega}_{\text{pub}}, [s^i]), m^i) \leq \mathbf{E}\left[x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})\right]$$

for all $(\bar{\omega}_{\text{pub}}, s^i), m^i$, and with equality iff $[m^i] = [s^i]$.[51]

[50] We will assume that, up to a perturbation of ρ which is implemented by the p.r.d., each action profile $a \in A$ is played with positive probability in each $s \in S$.

[51] We refer to the literature on scoring rules for details.

Under our irreducibility assumption, there is an equivalent LP formulation (see Puterman, Ch. 8.8) of $\mathcal{M}$, in which player i chooses the invariant joint distribution of states and actions. Namely, agent i chooses $\pi \in \hat{\Pi}^i_\rho$ to maximize

$$\mathbf{E}_\pi \left[r^i_\rho + x^i_\rho \right], \tag{18}$$

where $\hat{\Pi}^i_\rho \subset \Delta(\Omega_{\text{pub}} \times S^i \times M^i)$ is the set of joint distributions π such that

$$\pi(\omega_{\text{pub}}, t^i, M^i) = \sum_{(\bar{\omega}_{\text{pub}}, s^i), m^i} p_{(\bar{\omega}_{\text{pub}}, s^i), m^i}(\omega_{\text{pub}}, t^i)\pi(\bar{\omega}_{\text{pub}}, s^i, m^i), \text{ for all } (\omega_{\text{pub}}, t^i) \in \Omega_{\text{pub}} \times S^i, \tag{19}$$

that are induced by some stationary reporting strategy in $\mathcal{M}$. (This is equation (3) from Section 6.2.)

Thus, (ρ, x) is weakly truthful iff truth-telling is an optimal strategy in $\mathcal{M}$, that is, iff (18) is maximized by the (truth-telling) distribution $\pi_*((\bar{\omega}_{\text{pub}}, s^i), m^i) := \mu_\rho(\bar{\omega}_{\text{pub}}, s^i)1_{s^i = m^i}$.

We use a duality-based approach. Consider the zero-sum game between player i (who picks $\pi \in \hat{\Pi}^i_\rho$) and the designer who picks[52] $x^i_\rho : (\Omega_{\text{pub}} \times [S^i]) \times M^i \to [-M, M]$ to minimize the reward

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

This game has a value in pure strategies, and it is clear that any optimal strategy π^i_ρ for i is such that

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

That is, any optimal strategy of player i lies in Π^i_ρ, as defined in Section 6.2.

Note now that $\mathbf{E}_\pi[x^i_\rho]$ is independent of $\pi \in \Pi^i_\rho$. Thus, if π_* does *not* maximize $\mathbf{E}_\pi[r^i_\rho]$ over Π^i_ρ, then it cannot possibly maximize $\mathbf{E}_\pi[r^i_\rho + x^i_\rho]$ over $\Pi^i_\rho \subseteq \hat{\Pi}^i_\rho$.

Conversely, assume that π_* maximizes $\mathbf{E}_\pi[r^i_\rho]$ over Π^i_ρ, and let x^i_ρ be an optimal strategy of the designer in the game. Thus, π_* achieves $\max_{\Pi^i_\rho} \mathbf{E}\left[r^i_\rho + x^i_\rho\right]$ which, by the optimality of x^i_ρ, is equal to $\max_{\hat{\Pi}^i_\rho} \mathbf{E}\left[r^i_\rho + x^i_\rho\right]$. This concludes the proof. ■

[52]Pick as M any common upper bound on the Lagrangian coefficients in 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 $(\bar{\omega}_{\text{pub}}, m^i)$,

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

C Proof of Theorem 3

C.1 A Quick Overview

We slightly rephrase Theorem 3 by writing $V^{**} = \mathcal{H}_1 := \{v \in \mathbf{R}^I : \lambda \cdot v \le k_1(\lambda) \text{ for all } \lambda\}$, where $k_1(-e^i) = -\bar{v}^i$ and $k_1(\lambda) = \bar{k}(\lambda)$ otherwise.

We will work under Assumption **1**' below, similar to Assumption **1**, and will comment on the adjustments to be made under the latter one.

Assumption 1' There exists a policy $\rho_* : S \to \Delta(A)$, transfers $\tilde{x}_* : S \to \mathbf{R}^I$ such that the following holds with $\theta_* := \theta_{\rho_*, r+\tilde{x}_*}$. For each player $i \in I$, any two $\bar{s}^i \neq \tilde{s}^i \in S^i$, and any $s^{-i} \in S^{-i}$, one has

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

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

The proof is a variant of the proof of Theorem 2, and we will skip many technical details. We will construct a sequential equilibrium σ which implements a given payoff. The play is divided into an infinite sequence of blocks of random length. Odd blocks serve as transition blocks, and the even blocks are the main ones. The durations of the successive blocks are independent random variables, which follow geometric distributions, with parameter $(1-\delta)^{\beta_*}$ and $(1-\delta)^{\beta}$ for odd and even blocks respectively. We will have $\beta > \beta_*$, so that the expected duration of the main blocks is much larger. As in the proof of Theorem 2, the end of a block is "decided" by the p.r.d.; so this is revealed only *after* reports have been submitted in the current round.

Under σ, players always report their true state, and they play a fixed policy $\rho[k] : M \to A$ while in block k.[53] For k odd, $\rho[k]$ is set to ρ_*. For k even, the policy $\rho[k]$ is computed in the first round, τ_k, of block k, based on the publicly available information.

The updating formulas at the end of block k rely on "transfers" $x[k]$. These transfers will here be obtained as the sum of two components, $x_o : S \times Y \to \mathbf{R}^I$ and $x_t : S \times S \to \mathbf{R}^I$, which only depend on the policy $\rho[k]$ that is being implemented, and on a direction $\lambda[k] \in \Lambda$. The transfers x_o and x_t provide respectively the incentives for playing obediently $\rho[k]$ and for reporting truthfully. For x_o, we will rely on Assumption **2** and refer to Kandori and Matsushima (1998). For x_t, we will closely follow AS.

[53]Even following a lie.

We first explain how to choose together a policy ρ and transfers x, as a function of the direction λ.

C.2 The Design of x_t

The construction of x_t is different for odd and even blocks. For odd blocks, we simply set $x_t := \tilde{x}_*$. For later use, we note that, thanks to **Assumption 1'**, truth-telling is strictly optimal for player i in the decision problem with payoff

$$r^i(s^i, \rho_*(m^i, s^{-i})) + x_t^i(m^i, s^{-i}) + \theta_*^i(t),$$

as soon as the belief of player i over s^{-i} has full support.

For even blocks, we rely on AS. Because a player may be indifferent between different reports, transfers cannot be defined independently of the discount factor unlike in the proof of Theorem 2, and the design of x_t takes into account the perturbations on reporting incentives created by the transitions between blocks.

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

Assume first that $\lambda \in \Lambda$ is a positive, non-coordinate direction and consider the MDP in which players cooperate to maximize the λ-weighted sum of discounted payoffs in the game $G_\beta((\bar{\omega}_{\text{pub}}, s), 0, \delta, \gamma)$. The value $V_{\delta,\lambda}(\bar{\omega}_{\text{pub}}, s)$ of this MDP does depend on γ and on $(\bar{\omega}_{\text{pub}}, s)$, but there is a fixed policy ρ_λ which is optimal for all δ close to 1.[54] For δ large, ρ_λ maximizes $\lambda \cdot \mathbf{E}_{\mu_\rho}\left[(1-\tilde{\xi})r(s,a) + \tilde{\xi}\theta_*(s)\right]$ with $\tilde{\xi} = \frac{\delta\xi}{1-\delta(1-\xi)}$, over the set of all policies ρ, and therefore also $\mathbf{E}_{\mu_\rho}[\lambda \cdot r(s,a)]$ when taking the limit $\delta \to 1$, hence $\rho_\lambda \in \Xi$.

We now focus on the constrained version of the game $G_\beta((\bar{\omega}_{\text{pub}}, s), x, \delta, \gamma)$ (still denoted in the same way) in which players only choose which state to report, and actions following the report profile m are set to $\rho_\lambda(m)$.

Claim 7 *There exists $x_t : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$, with $\lambda \cdot x_t(\cdot) = 0$, such that truth-telling is a (sequential) equilibrium of $G_\beta((\bar{\omega}_{\text{pub}}, s), x_t, \delta, \gamma)$ (for every $(\bar{\omega}_{\text{pub}}, s)$.)*

[54] The original proof of Blackwell (1962) does not directly apply (because transitions depend here on δ), yet it adapts immediately.

Proof. The result follows by specializing AS (as before, a key step in the proof of Proposition 2) to our setup. Define first VCG transfers $\bar{x} : S \to \mathbf{R}^I$ as

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

Thus, truth-telling is an equilibrium of $G_\beta((\bar{\omega}_{\text{pub}}, s), \bar{x}, \delta, \gamma)$, but the "budget-balance" requirement is not met.

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

Next, for $j \in I$, $\tilde{s}^j \in S^j$ and $\bar{\omega}_{\text{pub}} \in \Omega_{\text{pub}}$, we define

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

Thanks to the irreducibility assumption, ΔT^j is bounded, uniformly over $\delta < 1$.

We finally define the transfers $x_t : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ by the formula

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

so that $\lambda \cdot x_t(\cdot) = 0$ holds by construction.

By the same argument as in AS, truth-telling is an equilibrium of $G_\beta((\bar{\omega}_{\text{pub}}, s), x_t, \delta, \gamma)$. To be more specific, denote by $\gamma_\delta(\bar{\omega}_{\text{pub}}, s)$ the payoff induced by ρ_λ in $G_\beta((\bar{\omega}_{\text{pub}}, s), \delta, x_t, \gamma)$. For each $\delta < 1$, $i \in I$, $\bar{\omega}_{\text{pub}} \in \Omega_{\text{pub}}$ and $s^i \in S^i$, the truthful report $m^i = s^i$ maximizes the expectation of

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

■

At this point, we have thus assigned to each positive, non-coordinate direction λ a fixed policy $\rho_\lambda \in \Xi$ which maximizes $\mathbf{E}_{\mu_\rho}[\lambda \cdot r(s, \rho(s))]$, and δ-dependent transfers x_t for which the conclusion of Claim 7 holds. When $\lambda \in \Lambda$ is a non-coordinate, but not necessarily positive direction, this construction and the conclusions are still valid, provided one restricts attention to the policies $\rho : \times_{i \in I(\lambda)} S^i \to A$.

For $\lambda = +e^i$, we let again $\rho_{+e^i} : S^i \to A$ be a (fixed) policy that maximizes the payoff of player i in $G_\beta((\omega_{\text{pub}}, s), 0, \delta, \gamma)$ (for all δ close to 1). Again, the policy ρ_{+e^i} maximizes $\mathbf{E}_{\mu_\rho}[r^i(s, \rho(s))]$, and the conclusions of Claim 7 trivially hold with $x_t(\cdot) = 0$.

Finally, assume $\lambda = -e^i$, and pick constant pure policies $\rho^{-i} = a^{-i} \in A^{-i}$ and a policy $\rho^i : S^i \to A^i$ that minmaxes player i in $G_\beta((\bar{\omega}_{\text{pub}}, s), 0, \delta, \gamma)$ that is, which achieves $\min_{a^{-i} \in A^{-i}} \max_{\rho^i : S^i \to A^i} \mathbf{E}\left[(1-\tilde{\xi})r(s,a) + \tilde{\xi}\theta_*(s)\right]$. Again, one can pick these policies independently of δ, provided δ is close enough to 1, and $\mathbf{E}_{\mu_\rho}[r^i(s^i, \rho(s^i))] = \bar{v}^i$. The conclusions of Claim 7 trivially hold with $x_t(\cdot) = 0$.

Transfers x_t are independent of the final payoff γ, which we now set equal to

$$\gamma := \mathbf{E}_{\mu_{\rho_\lambda}}\left[(1-\tilde{\xi})r(s, \rho_\lambda(s)) + \tilde{\xi}\theta_*(s)\right].$$

Observe that $\lim_{\delta \to 1} \lambda \cdot \gamma = k_1(\lambda)$,[55] and γ satisfies the fixed-point property $\gamma = \mathbf{E}_{\mu_{\rho_\lambda}}[\gamma_\delta(\bar{\omega}_{\text{pub}}, s)]$, where $\gamma_\delta(\bar{\omega}_{\text{pub}}, s)$ is the expected payoff (under truth-telling and policy ρ_λ) in the game $G_\beta((\bar{\omega}_{\text{pub}}, s), x, \delta, \gamma)$.

Finally, we define the δ-relative values $\theta_\delta : \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ by means of the equality

$$\gamma_\delta(\bar{\omega}_{\text{pub}}, s) = \gamma_\delta + (1-\delta)\theta_\delta(\bar{\omega}_{\text{pub}}, s).$$

By the irreducibility property, θ_δ is uniformly bounded for $\delta < 1$.

We also define $\tilde{\theta}_\delta$ by the equality $\theta_\delta(\bar{\omega}_{\text{pub}}, s) = (1-\xi)\tilde{\theta}_\delta(\bar{\omega}_{\text{pub}}, s) + \xi\theta_*(s)$, with the interpretation that θ_δ stands for the "ex ante" relative value, before the p.r.d. is observed, and $\tilde{\theta}_\delta$ is the "ex post" relative value, once the p.r.d. has chosen not to stop the game G_β in the current round. Therefore, the continuation/final payoff is equal to $\gamma + (1-\delta)\theta_\delta(\bar{\omega}_{\text{pub}}, s)$ prior to the p.r.d. and is next equal to $\gamma + (1-\delta)\tilde{\theta}_\delta(\bar{\omega}_{\text{pub}}, s)$ or to $\gamma + (1-\delta)\theta_*(s)$, depending on the outcome of the p.r.d.

C.3 The Design of x_o

The same construction of x_o will apply to even and odd blocks.[56] Let a policy $\rho : S \to A$, and a non-coordinate direction $\lambda \in \Lambda$ be given.[57]

By Lemma 1 of Kandori and Matshushima (1998), Assumption **2** ensures that for any pair $\{i, j\}$ such that $\lambda^i, \lambda^j \neq 0$, and any $d > 0$, there exist $\hat{x}^h_{i,j} : S \times Y \to \mathbf{R}$, $h = i, j$, such that

$$\lambda^i \hat{x}^i_{i,j}(\cdot) + \lambda^j \hat{x}^j_{i,j}(\cdot) = 0, \tag{20}$$

[55] Note that γ depends on δ through $\tilde{\xi}$.

[56] Except that we do not have to require transfers to be balanced for the latter.

[57] We will apply the following with $\rho = \rho_*$ or $\rho = \rho_\lambda$.

and the following holds. For all s, $a = \rho(s)$, and for all $\hat{a}^h \neq a^h$, all $\hat{s}^h$,

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

By subtracting the constant $\mathbf{E}[\hat{x}^i_{i,j}(s,y) \mid s, a]$ from all values $\hat{x}^i_{i,j}(s,y)$ (which does not affect (20), since (20) must also hold in expectations), we may assume that, for our fixed choice of a, one has $\mathbf{E}[\hat{x}^h_{i,j}(s,y) \mid s, a] = 0$, for $h = i, j$ and $s \in S$.

We then set $\hat{x} = \sum_{i \neq j} \hat{x}_{i,j}$. 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 $(\hat{s}^i, \hat{a}^i)$ such that $a^i \neq \hat{a}^i$.

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.

Formally, consider the MDP in which player i faces truth-telling and ρ^{-i}, chooses a report $m^i \in M^i = S^i$ and an action $a^i \in A^i$, and he gets the reward $r^i(s, a^i, \rho^{-i}(s^{-i}, m^i)) + \hat{x}^i(s^{-i}, m^i, y)$. Then we may pick $d > 0$ such that every stationary optimal policy specifies $\hat{\rho}^i(s^{-i}, m^i) = \rho^i(s^{-i}, m^i)$. Equivalently, it is uniquely optimal in the decision problems $\mathcal{D}^i_{\rho,\hat{x}}$ to obey ρ^i at the action step (even after an incorrect report). Note that the incentives for obedience are strict, so that they still hold when $\theta_{\rho,r+\hat{x}}$ is slightly perturbed. Note also that because of our normalization of $\hat{x}^i$, the private values in this MDP are still equal to $\theta^i_{\rho,r+\hat{x}}$ if player i sets $m^i = s^i$.

The argument is similar for the case of coordinate directions. For $\lambda = +e^i$, we will use a policy $\rho : S^i \to A$ which maximizes $\mathbf{E}_{\mu_\rho}[r^i(s, \rho(s))]$. For $\lambda = -e^i$, we will use a policy ρ which achieves the $\min\max$ in the definition of $\bar{v}^i$. In both cases, we set $x^i_o = 0$ and we use Assumption **2** and follow Kandori and Matsushima (1998, Cases 1 and 2, Theorem 1) to design x^j_o, $j \neq i$.

To summarize the previous and current sections, we have established, for fixed $\lambda \in \Lambda$ and $\delta < 1$, the existence of a policy $\rho_\lambda \in \Xi$, of transfers $x(\bar{\omega}_{\text{pub}}, m, y) = x_t(\bar{\omega}_{\text{pub}}, m) + x_o(m, y)$, and of θ_δ such that the properties below hold for all $\delta < 1$ large enough:

E1 $\lambda \cdot x(\cdot) = 0$;

E2 $\lim_{\delta\to 1} \lambda \cdot \gamma = k_1(\lambda)$, where $\gamma := \mathbf{E}_{\mu_{\rho_\lambda}}[(1-\tilde{\xi})r(s,\rho_\lambda(s)) + \tilde{\xi}\theta_*(s)]$;

E3 For each $\bar{\omega}_{\text{pub}} \in \Omega_{\text{pub}}$, $i \in I$ and $s^i \in S^i$, truth-telling and $\rho^i_\lambda(m)$ maximize the expectation of

$$(1-\xi)\left(r^i(s,a) + x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}) + \delta\theta^i_\delta(\omega_{\text{pub}}, t)\right) + \xi\theta^i_*(s),$$

which is then equal (conditional on s^{-i}) to $\theta^i_\delta(\bar{\omega}_{\text{pub}}, s)$;[58]

E4 θ_δ and x are uniformly bounded, over $\delta < 1$.

The conditions **E1–E3** are somewhat similar to saying that $(\rho_\lambda, x, \gamma)$ is feasible in a discounted analogue of the program $\mathcal{P}(\lambda)$ which would be twisted to reflect the probability ξ of switching. Note that truth-telling incentives are not strict.

We have also constructed $x_* : M \to Y \to \mathbf{R}^I$ such that (ρ_*, x_*) is an admissible pair, such that incentives to play $\rho_*(m)$ at the action step are strict, even following a lie.[59]

C.4 The Equilibrium Construction

C.4.1 The Parameters

The construction involves various parameters. Pick first the exponents $0 < \beta_* < \beta < 1$. Given Z, pick $\eta > 0$ such that Z_η is contained in the interior of $\mathcal{H}_1$, and $\varepsilon_0 > 0$ such that $\max_{Z_\eta} \lambda \cdot z < k_1(\lambda) - 2\varepsilon_0$ for all directions $\lambda \in \Lambda$.

Next, we use a compactness argument similar to that of Lemma 7. Given $\lambda \in \Lambda$, subtract $\frac{\varepsilon_0}{3}\lambda$ from the map x that was associated to λ and ρ_λ in Sections C.2 and C.3, and rewrite **E1** and **E2** as $\lambda \cdot x(\cdot) < 0$ and $\lim_{\delta\to 1} \lambda \cdot \gamma > k_1(\lambda) - \frac{\varepsilon_0}{2}$. This ensures that ρ_λ, the maps x and θ_δ may be chosen to be locally independent of λ. Since Λ is compact, this ensures that the transfers x may then be picked from a finite set of maps $\mathcal{X}$ as λ varies through $\Lambda.\theta_\delta$ and $x \in \mathcal{X}$, valid for all $\delta < 1$.

As in the proof of Theorem 2, we fix κ_1 large enough, let $\varepsilon_1 \in (0, \varepsilon_0)$, set $\varepsilon := \varepsilon_1/2\kappa_1$, and then let $\bar{\zeta}$ be given by Lemma 8 applied with ε. Given these values, we finally let $\bar{\delta}$ be close enough to one, so that a finite number of inequalities hold for all $\delta \geq \bar{\delta}$. Since most

[58] Here, the expectation is taken over s^{-i}, a, y, t, under the assumption that (i) s^{-i} is drawn according to the belief held by player i (knowing that players $-i$ used ρ_λ in the previous round) and (ii) players $-i$ report truthfully and play ρ^{-i}.

[59] In fact, it is even optimal to report truthfully *ex post*. This property is not used here, but will be used in the proof of Theorem 4.

computations will be omitted, we omit the exact conditions on κ_1 and $\bar{\delta}$ under which the computations below are valid.[60]

C.4.2 The Updating Process

The description accounts for the difference between odd and even blocks. Consider block $k+1$, starting in round $n+1 := \tau_{k+1}$. If $k+1$ is even, we define first $w[k+1]$ through the equality

$$\tilde{w}_{n+1} = \xi_* w[k+1] + (1-\xi_*) z[k], \tag{21}$$

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

$$\tilde{w}_{n+1} = \frac{1}{\delta} z[k] - \frac{1-\delta}{\delta} v_* + \frac{1-\delta}{\delta} x_*(\omega_{\mathrm{pub},n}). \tag{22}$$

Given $w[k+1]$, pick $\lambda[k+1] \in \Lambda$ so that the conclusion of Lemma 8 holds. Set $\rho[k+1] = \rho_\lambda$, $x[k+1] = x$ and $\theta[k+1] = \theta_\delta$. Next,

$$z[k+1] := w[k+1] + (1-\delta)\left(1 + \frac{1-\delta}{\delta\xi_*}\right)\theta_*(m_{n+1}) - (1-\delta)\theta[k+1](\bar{\omega}_{\mathrm{pub},n}, m_{n+1}). \tag{23}$$

If $k+1$ is odd, we define $w[k+1]$ by

$$\tilde{w}_{n+1} = \xi w[k+1] + (1-\xi) z[k], \tag{24}$$

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

$$\tilde{w}_{n+1} = \frac{1}{\delta} z[k] - \frac{1-\delta}{\delta} v[k] + \frac{1-\delta}{\delta} x[k](\omega_{\mathrm{pub},n}, m_{n+1}). \tag{25}$$

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

The process is initialized as follows. Given a target payoff $\bar{z} \in Z$, we set $z[1] := \bar{z} - (1-\delta)\mathbf{E}_\pi[\theta_*(s)]$ and $w[1] := \bar{z}$.

C.4.3 The Strategies

Under σ, players report truthfully their current state in every round, and play $\rho(m_n) = \rho_*(m_n)$ or $\rho(m_n) = \rho[k](m_n)$ if n belongs respectively to an odd or an even block.

The continuation payoff of player i at the action step in round n, following a truthful report $m_n^i = s_n^i$ thus only depends on $\omega_{\mathrm{pub},n-1}$, on s_n and on the current value $\mathcal{T}_n$

[60]The computations are similar to those in Theorem 2, but not identical, hence the relevant conditions on κ_1 and $\bar{\delta}$ are not the same.

of the auxiliary family $(z_n, w_n, \rho_n, x_n, v_n)$ (see the proof of Theorem 2), and will be denoted $\gamma^i(\omega_{\mathrm{pub},n-1}, s_n; \mathcal{T}_n)$. Using (omitted) arguments similar to Lemmas 10 and 11, one has $\gamma(\omega_{\mathrm{pub},n-1}, s_n; \mathcal{T}_n) = z_n + (1-\delta)\theta_*(s_n)$ or $\gamma(\omega_{\mathrm{pub},n-1}, s_n; \mathcal{T}_n) = z_n + (1-\delta)\tilde{\theta}_\delta(m_{n-1}, s_n)$, if round n belongs to an odd or an even block respectively.

That σ is well-defined follows from the next lemma.

Lemma 13 *One has $w[k] \in Z_\eta$ for k even.*

Proof. The proof goes by induction. Note first that $\|w[k] - z[k]\| \le (1-\delta)\kappa_1$ for all k, and assume that $w[k] \in Z_\eta$ for some even k.[61] With the obvious adjustments to the proof of Lemma 9, one has

$$\begin{aligned}
\|w[k+2] - w[k]\| &\le \|w[k+2] - w[k+1]\| + \|w[k+1] - z[k]\| + \|z[k] - w[k]\| \\
&\le \frac{1-\delta}{\delta\xi_*}\kappa_1 + 2\frac{1-\delta}{\delta\xi}\kappa_1 + (1-\delta)\kappa_1 \le 3\frac{1-\delta}{\delta\xi}\kappa_1.
\end{aligned}$$

Set $\zeta = 3\dfrac{1-\delta}{\delta\xi}\kappa_1$, and note that

$$\begin{aligned}
\lambda[k] \cdot (w[k+2] - w[k]) &\le -\varepsilon_0 \times \frac{1-\delta}{\delta\xi} + \|w[k+2] - w[k+1]\| + \|z[k] - w[k]\| \\
&\le -\varepsilon_0 \times \frac{1-\delta}{\delta\xi}\kappa_1 + \frac{1-\delta}{\delta\xi_*}\kappa_1 + (1-\delta)\kappa_1 \le -\varepsilon\zeta.
\end{aligned}$$

The result follows as in Lemma 9. ■

C.4.4 Equilibrium Properties

We will argue briefly that player i has no profitable one-step deviation. We first place ourselves at the action step in round n, with $\tau_k \le n < \tau_{k+1}$. At that step, his overall continuation payoff (assuming no deviation from round $n+1$ on, and taking expectations at round $n+1$) is given by

$$(1-\delta)r_n^i + \delta\left(z_{n+1}^i + (1-\delta)\theta_{n+1}^i(\omega_{\mathrm{pub},n}, s_{n+1})\right).$$

Assume first that k is odd. Algebraic manipulations (akin to Lemmas 10 and 11) show that, for a fixed action $a_n^i \in A^i$, the expected continuation payoff is (up to a term independent of a_n^i) equal to

$$(1-\delta)\mathbf{E}\left[r^i(s_n^i, \rho_*^{-i}(m_n), a_n^i) + x_o^i(m_n, y_n) + \theta_*^i(s_{n+1})\right],$$

[61] One has $w[2] \in Z_\eta$ since $w[1] \in Z$ and δ is close to one.

where the expectation is taken over y_n and s_{n+1}. Whether $m_n^i = s_n^i$ or not, the design of x_o ensures that the expectation is maximal for $\rho^i(m_n)$.

Assume next that k is even. By similar manipulations, the expected continuation payoff is[62] (again up to a term which does not depend on a_n^i) equal to

$$(1-\delta)\mathbf{E}\left[r^i(s_n^i, \rho_{\lambda_n}^{-i}(m_n), a_n^i) + x_o^i(m_n, y_n) + \delta\theta_\delta^i(\omega_{\text{pub},n}, s_{n+1})\right].$$

Since $\delta\theta_\delta$ is arbitrarily close to $\theta_{\rho_{\lambda_n}, r+x_t+x_0}$, the design of x_o again ensures that the expectation is uniquely maximized for $\rho_{\lambda_n}^i$, provided δ is large enough.

We now place ourselves at the reporting step of round n, with $\tau_k < n \leq \tau_{k+1}$. If k is odd, the truth-telling property follows using the same argument as in Theorem 2.[63] Assume now k even. The expected continuation payoff is, up to a term which does not depend on m_n^i, (at most, with equality if $m_n^i = s_n^i$) equal[64] to the expectation of

$$(1-\xi)\left(r^i(s_n, \rho_{\lambda_{n-1}}(s_n^{-i}, m_n^i)) + x^i(\omega_{\text{pub},n-1}, \omega_{\text{pub},n}) + \delta\theta_\delta^i(\omega_{\text{pub},n}, s_{n+1}\right) + \xi\theta_*^i(s_n)$$

which, by **E3**, is maximized for $m_n^i = s_n^i$. This implies the result.

We now comment on the difference between Assumption **1** and **1'**. Instead of playing a fixed ρ_*, which simplified the description, the play in odd blocks is now replaced by the following. In the first period of an odd block, a player is selected at random (using the p.r.d.); if player i is selected and reported $s^i \in S^i$, we use the p.r.d. to determine which one of the sequences $(a_n^k) \in \Gamma^i$ is played, according to the distribution that makes truthful reporting strictly optimal. (This requires the discount factor to be above a certain threshold.) Yet this sequence is only played for a random duration: a random time τ is determined using the p.r.d., at which a new player is selected at random (this could be player i again); this random time follows a geometric distribution with parameter $(1-\delta)^{\beta_{**}}$ with β_{**} much smaller than β_*, the random duration of the odd block; β_{**} and the minimum discount factor are chosen so that strict incentives to report truthfully follow from Assumption **1**. Note that, given IPV, conditional on not being selected, the report of player i is irrelevant. On the other hand, conditional on being selected, it is strictly optimal to report truthfully, given Assumption **1**. This (random) strategy profile defines a relative value function θ_*, to be used in the definition of x_t and x_o in even blocks.

[62] We stress again that x_o and θ_δ are the transfer and relative rent associated to ρ_{λ_n} in Sections C.2 and C.3.

[63] In fact, truth-telling is even *ex post* optimal in any such round.

[64] Indeed, if the p.r.d. switches to a new block, the continuation "relative value" is at most $\theta_*(s_n)$, with equality if $m_n^i = s_n^i$.

D Proof of Theorem 4

The proof of Theorem 4 consists in adding a layer of complexity to the proof of Theorem 3 to deal with negative, unit directions. We will extensively refer to the latter to avoid duplications. We will work under the assumption that the distribution of signals is independent of the current states. The proof for the general case is more cumbersome, but does not involve additional insights.

D.1 Alternative Scores

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

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

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

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

where $h_{\text{pub},T+1}$ is the public history in the T rounds. Information and play is as in the infinite horizon game.

Denote by C a uniform bound on $\theta_{\rho,r}$, when 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.
- $\lambda \cdot x(\cdot) \leq 0$ and $\lambda \cdot \theta(\cdot) \leq C$.

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

Proposition 2 *One has $\mathcal{H}_2 = W$.*

Proposition 2 follows from Lemmas 14 and 15 below.[65]

[65] Only the inclusion $\mathcal{H}_2 \supseteq W$ is relevant for the proof.

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

Proof. Fix $\lambda \in \Lambda$ with $\lambda \neq -e^i$ for all $i \in I$. Let a weakly truthful pair (ρ, x) be given with $\lambda \cdot x(\cdot) = 0$, set $v := \mathbf{E}_{\mu_\rho}[r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})]$, and $\theta := \theta_{\rho, r+x}$. Given an integer $T \in \mathbf{N}$, define $x_T : \Omega_{\text{pub}}^T \to \mathbf{R}^I$ 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 : (i) each player i reports truthfully $m_n^i = s_n^i$ in all rounds, irrespective of past play, (ii) in each round n, player i plays $\rho^i(m_n)$ if $m_n^i = s_n^i$, and any action a^i which maximizes the expectation of

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

otherwise. Denote by $\tilde{\gamma}_T(\sigma_T)$ the expected payoff of σ_T in $G(T, x, \theta)$.

Since (ρ, x) is weakly truthful, it is easily checked that σ_T[66] is a sequential equilibrium in $G(T, x, \theta)$, hence $\lambda \cdot \tilde{\gamma}_T(\sigma_T) \leq \tilde{k}_T(\lambda)$. On the other hand, by the irreducibility assumption, one has $\lim_{T\to+\infty} \tilde{\gamma}_T(\sigma_T) = \mathbf{E}_{\mu_\rho}\left[r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}})\right] = v$, so that $\lambda \cdot v \leq k_2(\lambda)$. Using Lemma 5, this shows that $\bar{k}(\lambda) \leq k_2(\lambda)$, as desired.

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

$$\lambda \cdot \tilde{\gamma}_T(\tilde{\sigma}_T) \leq \lambda \cdot \gamma_T(\sigma_T) + \frac{C}{T}, \tag{26}$$

where $\gamma_T(\sigma_T)$ is the payoff induced in the T-round game $G(T, 0, 0)$ with no final payoffs. Denote by $v_T(\lambda) := \sup_\sigma \lambda \cdot \gamma_T(\sigma)$ the value of the λ-weighted T-round game, where the supremum is taken over $\sigma : \times_{i \in I(\lambda)} S^i \to A$. By the irreducibility assumption, $\lim_{T\to+\infty} v_T(\lambda) = \bar{k}(\lambda)$. Let now $T \to +\infty$ in (26) to get $k_2(\lambda) - \varepsilon \leq \bar{k}(\lambda)$. The result follows. ∎

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

[66] When supplemented with appropriate beliefs.

Proof. In all games $G(T,x,\theta)$ considered for this lemma, the reports will be "babbling," that is, a player sends the same report independently of his type in any given period. We set $\theta=0$. Given $k\in\mathbf{N}$, let $A^j_k := \{\alpha^j\in\Delta(A^j) : k\alpha^j(a^j_l)\in\mathbf{N}$ for all $a^j_l\in A^j\}$. The set A^j_k consists of those mixed action profiles that assign rational probabilities with denominator k. For any $\alpha^j\in\Delta(A^j)$, there exists $\alpha^j_k\in A^j_k$ such that $d(\alpha^j_k,\alpha^j)\leq |A^j|/k$; similarly, for all $\alpha^{-i}\in\times_{j\neq i}\Delta(A^j)$, $d(\alpha^{-i}_k,\alpha^{-i})\leq|A^{-i}|/k$, for some $\alpha^{-i}_k\in A^{-i}_k := \times_{j\neq i}A^j_k$. [67] We write Σ^j_k for the strategies of j with values in A^j_k, and we let

$$\underline{\sigma}_k = \arg\min_{\sigma^{-i}\in\Sigma^{-i}_k}\max_{\sigma^i}\limsup_T \frac{1}{T}\mathbf{E}_\sigma\left[\sum_{n=1}^T g^i(s^i_n,a^i_n,y_n)\right]$$

be minmax strategies when players $-i$ are constrained to strategies in Σ^{-i}_k. Given the product structure, these strategies may be taken measurable with respect to the history of signals of player i, and we write $h^i_{\mathrm{pub},n}\in H^i_{\mathrm{pub},n}=(Y^i)^{n-1}$ for such public histories. We write $w^i(k)$ for the limiting expected payoff of player i under $\underline{\sigma}_k$. Using the irreducibility assumption, it follows from standard arguments that $\lim_{k\to+\infty}w^i(k)=w^i_i$.[68]

Given $T\in\mathbf{N}$, we also write $w^i(k,T)$ for the highest expected payoff of player i over the first T rounds, when facing $\underline{\sigma}^{-i}_k$. Given a realized public history $h^i_{\mathrm{pub},T+1}\in H^i_{\mathrm{pub},T+1}$ and $\alpha^{-i}_k\in A^{-i}_k$, we let

$$T(\alpha^{-i}_k)=\{n=1,\ldots,T : \underline{\sigma}^{-i}_k(h^i_{\mathrm{pub},n})=\alpha^{-i}_k\}$$

denote the rounds at which $\underline{\sigma}^{-i}_k$ prescribes α^{-i}_k,[69] and $f[\alpha^{-i}_k]\in\Delta(Y)$ be the empirical distribution of signals observed in those stages. For $j\neq i$ and $y^j\in Y^j$, we also denote by $f[\alpha^{-i}_k](y^j)$ the empirical frequency of y^j over the stages in $T(\alpha^{-i}_k)$. We now let

$$D^j(h_{\mathrm{pub},T+1})=\sum_{\alpha^{-i}_k\in A^{-i}_k}\frac{|T(\alpha^{-i}_k)|}{T}\sum_y\left|f[\alpha^{-i}_k](y)-f[\alpha^{-i}_k](y^{-j})\mathbf{P}[y^j\mid\alpha^j_k]\right|,$$

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

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

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

[67] Throughout the lemma, we use the Euclidean distance.
[68] And that min and max are indeed achieved, so that $\underline{\sigma}_k$ is well-defined.
[69] Here, $h^i_{\mathrm{pub},n}$ refers to an initial segment of $h^i_{\mathrm{pub},T+1}$.

Claim 8 *Given $\varepsilon > 0$ and $\phi > 0$, there exists T_0 such that, if $T \geq T_0$,*

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

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

In words, if player j uses $\underline{\sigma}_k^j$, he is very likely to pass the test τ_ϕ^j no matter players $-j$'s strategy profile. The proof of Claim 8 relies on approachability theory, see Gossner (1995) for details.

Given $\varepsilon > 0$, we let $\phi < \dfrac{2\varepsilon}{\bar{r}(I-1)}$, and let T_0 be given by Claim 8 applied with $\varepsilon/2\bar{r}$ and ϕ.[70] Given $T \geq T_0$, we pick $M > 0$ such that

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

or equivalently, $M > T\dfrac{\bar{r}}{\varepsilon}$. That is, M is a punishment sufficiently large (for failing the test) that getting the worst reward for T rounds followed by a probability of failing the test of up to ε exceeds the payoff from the highest reward for T rounds followed by a probability of failing the test of at least 2ε.

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

$$x^j(h_{\text{pub},T+1}) = \begin{cases} -M & \text{if } \tau_\phi^j(h_{\text{pub},T+1}) = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

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

Claim 9 *For every sequential equilibrium σ of $G(T, x, \theta)$, one has*

$$\frac{1}{T}\mathbf{E}_{s_*,\sigma}\left[\sum_{n=1}^{T} g^i(s_n^i, a_n^i, y_n)\right] \leq w^i(k, T) + 2\varepsilon.$$

Proof. By the condition on M, one has $\mathbf{P}_\sigma\left[\tau_\phi^j(h_T) = 0\right] < 2\varepsilon_1$ in all equilibria of $G(T, x, \theta)$. Take any strategy profile in $G(T, x, \theta)$ such that $\mathbf{P}_\sigma\left[\tau_\phi^j(h_T) = 0\right] < \dfrac{\varepsilon}{2\bar{r}}$ for all $j \neq i$. On the event $\cap_{j \neq i}\{\tau_\phi^j(h_T) = 1\}$, one has for all $j \neq i$,

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

[70] Here, $\bar{r}$ is a uniform bound on all payoffs in the game.

which implies, by repeated substitution,

$$\sum_y \sum_{\alpha_k^{-i}} \frac{|T(\alpha_k^{-i})|}{T} \left| f[\alpha_k^{-i}](y) - f[\alpha_k^{-i}](y^i) \times_{j \neq i} \mathbf{P}[y^j \mid \alpha_k^j] \right| < (I-1)\phi. \tag{28}$$

We have that

$$\begin{aligned} \frac{1}{T}\sum_{n=1}^{T} g^i(s_n^i, a_n^i, y_n) \;&\leq\; \frac{1}{T}\sum_{\alpha_k^{-i}} \left| \sum_{n \in T(\alpha_k^{-i})} \left(g^i(s_n^i, a_n^i, y_n) - \sum_{\tilde{y}^{-i}} g^i(s_n^i, a_n^i, (\tilde{y}^{-i}, y_n^i)) \mathbf{P}[\tilde{y}^{-i} \mid \alpha_k^{-i}] \right) \right| \\ &+ \; \frac{1}{T}\sum_{\alpha_k^{-i}} \sum_{n \in T(\alpha_k^{-i})} \sum_{\tilde{y}^{-i}} g^i(s_n^i, a_n^i, (\tilde{y}^{-i}, y_n^i)) \mathbf{P}[\tilde{y}^{-i} \mid \alpha_k^{-i}]. \end{aligned}$$

By (28), the first sum is bounded by $\varepsilon/2$ on the event $\cap_{j \neq i}\{\tau_\phi^j(h_T) = 1\}$, and by $\bar{r}$ on its complement, which is of probability at most $\dfrac{\varepsilon}{2\bar{r}}$. The expectation of the second sum under an arbitrary profile σ does not depend on σ^{-i} and is equal to the payoff induced by $(\sigma^i, \underline{\sigma}_k^{-i})$. This implies the result. ■

Claim 9 implies $\tilde{k}_T(-e^i) \geq -w^i(k,T) - \varepsilon$ for all large T. Letting first $T \to +\infty$, then $k \to +\infty$, we get $k_2(-e^i) \geq -w_i^i - \varepsilon$, hence $k_2(\lambda) \geq -w_i^i$ since ε is arbitrary. The reverse inequality is obvious. ■

D.2 The Strategies

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

As in Theorem 3, the play is divided into an infinite sequence of blocks, with odd blocks serving as transition blocks. Even blocks are now either "regular," or devoted to the punishment of a single player. The behavior in odd and in regular even blocks is identical to that in Theorem 3. In contrast, the duration of an punishment even block is fixed and set equal to $(1-\delta)^{-\beta}$ rounds.

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

D.2.1 Punishment Blocks

The equilibrium behavior in punishment blocks relies on an elaborate version of Lemma 15, which we now introduce. Given $T \in \mathbf{N}$, $x : M \times Y^T \to \mathbf{R}^I$, $\delta < 1$ and $m \in M$, we denote by

$G(m, \delta, x, T)$ a discounted T-round version of $G(T, x, \theta_*)$ without communication and initial state $m \in M$. That is, in each round $n = 1, \ldots, T$, players observe their private states (s_n^i) choose actions (a_n^i), and $(y_n, s_{n+1}) \in Y \times S$ is drawn according to p_{s_n, a_n}.

The payoff vector is

$$\frac{1-\delta}{1-\delta^{T+1}} \left\{ \sum_{n=1}^{T} \delta^{n-1} r(s_n, a_n) + \delta^T x(m, \vec{y}) + \delta^T \theta_*(s_{T+1}) \right\}, \tag{29}$$

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

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

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

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

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

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

Plainly, x and γ can then be chosen such that $|\gamma^i[i] - w_i^i| < \varepsilon_2$ and $x^i[i](\cdot) > \frac{\varepsilon_2}{2}$.

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

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

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

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

Pick now $k \in \mathbf{N}$ such that $|w_i^i - w^i(k)| < \varepsilon$, then $\bar{T}$ such that $|w^i(k) - w^i(k, T)| < \varepsilon$ for all $T \geq \bar{T}$.[71]

We follow closely Lemma 15. We take ϕ and T_0 as specified after Claim 8. We let $\delta_2 < 1$ be such that $T = (1-\delta)^{-1/2} \max(\bar{T}, T_0)$ for all $\delta \geq \delta_2$ and $\delta_3 < 1$ such that the normalized δ-discounted sum of payoffs in the first $T = (1-\delta)^{-1/2}$ stages differs from the arithmetic mean by at most ε, for all $\delta \geq \delta_3$.

[71] Recall the definition of $w^i(k, T)$ from Lemma 15. Note however that the definition has to be amended, to reflect the fact that the initial state profile is m and no longer s_*. We still use the same notation.

Let $\delta \geq \max(\delta_1, \delta_2, \delta_3)$ be arbitrary. Define as before $x^i[i](\cdot) = 0$, and, for $j \neq i$, $x^j[i]$ as in (27) with $M = \kappa T$. Pick an arbitrary equilibrium $\sigma[i,m]$ of $G(m, \delta, x[i], T)$. It follows, as in Claim 9, that the (discounted) payoff $\tilde{w}^i_m$ of player i under $\sigma[i,m]$ does not exceed $w^i(k,T) + 2\varepsilon \leq w^i_i + 4\varepsilon$.

Observe also that $\tilde{w}^i_m \geq w^i_i - \varepsilon$ provided δ is close enough to one. Hence $\|\tilde{w}^i_m - \tilde{w}^i_{m'}\| < 5\varepsilon$. For all $j \in I$, define $x^j_m[i]$ by adding to $x^j[i]$ the quantity $\max_{m'} \left(\tilde{w}^j_{m'}[i] - \tilde{w}^j_m[i]\right)$ properly normalized. The added constant does not affect incentives, but ensures that the new equilibrium payoff vector, $\gamma[i]$, is independent of $m \in S$.

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

■

D.2.2 The Parameters

As in Theorem 4, given Z, pick first $\eta > 0$ such that Z_η is contained in the interior of $\mathcal{H}_2$, and $\varepsilon_0 > 0$ such that $\max_{Z_\eta} \lambda \cdot z < k_2(\lambda) - 2\varepsilon_0$ for all directions $\lambda \in \Lambda$. Let κ_R be obtained when applying Lemma 16 with $\varepsilon := \varepsilon_0$.

Pick $\varepsilon_R < \varepsilon_0/\kappa_R$, and set $\tilde{\Lambda} := \Lambda \setminus \cup_i B(-e^i, \varepsilon_R)$. Replicating with the compact set $\tilde{\Lambda}$ the same compactness argument as in Section C.4.1, we may assume wlog that the transfers x are picked from a finite set of maps $\mathcal{X}$ as λ varies through $\tilde{\Lambda}.\theta_\delta$ and $x \in \mathcal{X}$, valid for all $\delta < 1$. Pick next $\beta_* \in (0, 1/2)$.

As in the proof of Theorems 2 and 3, we fix κ_2 large enough, let $\varepsilon_1 \in (0, \varepsilon_0)$, set $\varepsilon := \varepsilon_1/2\kappa_1$, and then let $\bar{\zeta}$ be given by Lemma 8 applied with ε. Given these values, we finally let $\bar{\delta}$ be close enough to one, so that a finite number of inequalities hold for all $\delta \geq \bar{\delta}$. Again, we omit the exact conditions on κ_2 and $\bar{\delta}$ under which the computations below are valid.

D.2.3 The Updating Process

We follow Section C.4.2. Consider a block $k+1$, starting in round $n+1 := \tau_{k+1}$. If $k+1$ is even, we define first $w[k+1]$ and $\tilde{w}_{n+1}$ by (21) and (22), and we pick $\lambda[k+1] \in \Lambda$ so that the conclusion of Lemma 8 holds.

If $\lambda[k+1] \in \tilde{\Lambda}$, so that block $k+1$ is regular, we define $(\rho[k+1], x[k+1], v[k+1], \theta[k+1])$ as in Section C.2, and let $z[k+1]$ be defined by (23).

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

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

Assume now that $k+1$ is odd. If block k was regular, we define $w[k+1]$ and $\tilde{w}_{n+1}$ by means of (24) and (25), and 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 in Theorem 4.

D.2.4 The Strategies

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

In any block which is *not* a punishment block, the strategy of player i is defined as in the proof of Theorem 4.

That σ is well-defined follows from the next lemma.

Lemma 17 *One has $w[k] \in Z_\eta$ for k even.*

Proof. We proceed as in Lemma 13. Assume that $w[k] \in Z_\eta$ for some even k. It suffices to deal with the case where block k is a i_*-punishment block, for some $i_* \in I$. From the updating formula, it follows that

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

so that

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

Denote by ζ the right-hand side.

On the other hand,

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

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

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

hence $w[k+2] \in Z_\eta$. ■

D.2.5 The Equilibrium Property

Fix a player $i \in I$. As in Theorem 3, the construction of the strategy profile σ ensures that the continuation payoff of player i at the action step of a given round n is given by $\gamma^i(\omega_{\text{pub},n-1}, s_n; \mathcal{T}_n) = z_n + (1-\delta)\theta_*(s_n)$ or $z_n + (1-\delta)\tilde{\theta}_\delta(\omega_{\text{pub},n-1}, s_n)$ whenever $m_n^i = s_n^i$ and round n is not part of a punishment block. In addition, the continuation payoff in the first stage of a i_*-punishment block is $\gamma^i[i_*]$, again if $m_n^i = s_n^i$.

For use below, we make the following observation. Fix $m \in M$, $i \in I$, $s^i \in S^i$, and consider the variant $\tilde{G}_i(m, \delta, x[i_*], T)$ of $G(m, \delta, x[i_*], T)$ in which the initial state of i is s^i instead of m^i.[72] Thanks to the irreducibility property, the highest payoff of i in $\tilde{G}_i(m, \delta, x[i_*], T)$ when facing $\sigma^{-i}[i_*, m]$ differs from the payoff $\gamma^i[i_*]$ induced by $\sigma[i_*, m]$ in $G(m, \delta, x[i_*], T)$ (and therefore from the payoff induced by $\sigma[i_*, (m^{-i}, s^i)]$ in $G((m^{-i}, s^i), \delta, x[i_*], T))$ by at most $(1-\delta)\bar{\kappa}$, where $\bar{\kappa}$ is a constant that only depends on κ_2 and on the primitives of the game. Thus, misreporting at the beginning of a punishment block does not benefit much.

That player i cannot profitably deviate at the action step of a given round n follows as in Theorem 3, unless n is part of a punishment block, in which case it follows from the sequential rationality of σ in that block.

That player i cannot profitably deviate by lying in a regular block also follows as in Theorem 3. On the other hand, players babble in punishment blocks.

We now place ourselves at the reporting step of a round n in a transition block. There are two cases: either n is the first stage of the transition block, following a i_*-punishment block; or it is not. In the former case, the belief of i is derived from the public history and the strategies $\sigma[i_*]$; in the latter, it is derived using ρ_*. In both cases, the belief of i over s_n^{-i} has full support, and the optimality of truth-telling follows along the lines of Theorem 3, using (i) the *ex post* optimality of truth-telling under ρ_* and (ii) the fact that misreporting in the first round of a punishment block has only a minor impact (of the order of $(1-\delta)$, see above) on the continuation payoff of player i.

E Proofs for the Correlated Case

Proof of Lemma 6. Assumptions **2'(a)–2'(b)** are the counterparts of Assumption **2** (specialized to the action profiles that are played), so that the result follows exactly as in the proof of Theorem 3. ■

[72] But final transfers are still given by $x[i_*](m, \vec{y})$.

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 ($\pi^{\bar{m},\bar{a},\bar{y}}(\cdot \mid \hat{c}) \notin \mathrm{co}\,\mathcal{R}^i(\bar{m}, \bar{a}, \bar{y})$, implied by Assumption **5.1**), there exists transfers x^i that ensure that truthful reporting by player i is strictly optimal. Because of weak identifiability (invoking **5.2** if $\mathrm{sgn}(\lambda^i) = \mathrm{sgn}(\lambda^j)$ and **5.1** 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 **4** 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 $\lambda \cdot x(\cdot) = 0$. ■