TRUTHFUL EQUILIBRIA IN DYNAMIC BAYESIAN GAMES

By

Johannes Hörner, Satoru Takahashi and Nicolas Vieille

December 2013
Revised January 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1933R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Truthful Equilibria in Dynamic Bayesian Games

Johannes Hörner*, Satoru Takahashi† and Nicolas Vieille‡

January 19, 2015

Abstract

This paper characterizes an equilibrium payoff subset for dynamic Bayesian games as discounting vanishes. Monitoring is imperfect, transitions may depend on actions, types may be correlated and values may be 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 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.

Keywords: Bayesian games, repeated games, folk theorem.

JEL codes: C72, C73

1 Introduction

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

The contribution of this paper is as follows. First, we define a class of equilibria –truthful equilibria– in which players report their private information honestly in every period, on

*Yale, 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.

and off path, and do not condition their play on past private information. This class of strategies specializes to public strategies (and truthful equilibria to perfect public equilibria) in the case of complete information. Second, we relate truthful equilibrium payoffs (as the discount factor tends to one) to a static Bayesian game augmented with transfers. Here as well, this is a natural generalization of the results for repeated games (Fudenberg and Levine, 1994, in particular). Third, we prove that the restriction to truthful equilibria is without loss in some contexts (again, as $\delta \to 1$). In particular, we show that this is the case with independent private values, where truthful equilibria are only restrictive for obedience. That is, under usual identifiability conditions, truthfulness only limits how low a particular player's equilibrium payoff can be.

Relative to existing papers on games with persistent private information, the set of equilibrium payoffs we obtain is larger,[1] and tight under some conditions. Furthermore, the class of games considered is significantly more general. Our methods apply to games exhibiting:

- moral hazard (imperfect monitoring);

- endogenous serial correlation (actions affecting transitions);

- correlated types (across players) and interdependent values.

These features are all missing from the existing literature (with the exception of interdependent values, which is exhibited by the cheap-talk game of Renault, Solan and Vieille, 2013, and imperfect monitoring, in the class considered by Barron, 2013).

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.)

More precisely, 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

[1]The one exception is the lowest equilibrium payoff in Renault, Solan and Vieille (2013), who also characterize Pareto-inferior "babbling" equilibria, in a game that has interdependent values.

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.

As mentioned, our analysis is about *truthful* equilibria. In a truthful equilibrium, players truthfully reveal their type at the beginning of each round, after every history. In addition, players' action choices are public: they only depend on their current type and the public history. 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.

In Sections 2–5, we focus on the case of independent private values: payoffs only depend on a player's own private information (and the action profile), and this information evolves independently across players, conditional on the public information and one's own private action. We provide a family of one-shot games with transfers that reduce the analysis from a dynamic infinite-horizon game to a static game. Unlike the one-shot game of FL and HSTV (special cases of ours), this one-shot game is Bayesian. Each player makes a report, then takes an action; the transfer is then determined. This reduction provides a bridge between dynamic games and Bayesian mechanism design. As explained below, its payoff function is not entirely standard, raising interesting new issues for static mechanism design. Nonetheless, well-known results can be adapted for a wide class of dynamic games. Under independent private values (and also under correlated types), the analysis of the one-shot game yields an equilibrium payoff set that is best possible, except for the definition of individual rationality.

For such games, we prove a folk theorem: truthful equilibria might be restrictive in terms of individual rationality (lowest equilibrium payoff for a given player), but they do not restrict the set of equilibrium payoffs otherwise. Leaving aside individual rationality, we show that the payoff set attained by truthful equilibria is actually equal to the limit set of *all* Bayes Nash equilibrium payoffs, whichever message sets one chooses. In other words, in the revelation game in which players commit to the map from reports to actions, but not to current or future reports, there is no loss of generality in restricting attention to truthful equilibria. In this sense, the revelation principle extends, despite the absence of commitment, provided players are patient enough. Beyond generalizing the results of Athey and Bagwell (2001), as well as Escobar and Toikka (2013), this characterization has some interesting consequences. For instance, when actions do not affect transitions, the invariant distribution of the Markov chain is a sufficient statistic for the Markov process, as far as this equilibrium payoff set is concerned, leaving individual rationality aside.

In Section 5, we further concentrate on games in which monitoring has a product struc-

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

In Section 6, we state a general version of our main theorem, which provides a subset of limit equilibrium payoffs, whether types are correlated and values are private, or not. Conclusive characterizations are obtained under independent private values and correlated types. The paper focuses mostly on private independent values. The case of correlated types is relegated to the working paper. 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 2013 for instance) whose adaptation to incomplete information is inspired by the linking mechanism described in Fang and Norman (2006) and Jackson and Sonnenschein (2007). 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 precisely independent private values. Together with the irreducibility of the Markov chain, independence 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 does not suffice to adjust his long-run payoff, as such an adjustment affects all the different types identically (and so cannot give them strict incentives to use different strategies). On the other hand, we cannot focus on the flow payoff either to

provide such incentives, as with persistent types, a player's private information also enters his continuation payoffs. Hence, 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 regarding one's type are concerned. Furthermore, this transient component depends on the player's initial state, according to the future actions played. On the other hand, as far as obedience is concerned (playing the agreed upon action profile, given the public reports), the usual logic applies, since this action does not depend on the player's private information: changes in long-run payoffs according to the realized public signal provide adequate incentives.

Hence, 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 (2001, 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). Because of the importance of such games for applications in industrial organization and macroeconomics, there is an extensive literature on recursive formulations for fixed discount factors. In game theory, recent progress has been made in the case in which the state is observed, see Fudenberg and Yamamoto (2011) and HSTV for an asymptotic analysis, and Pęski and Wiseman (2014) for the case in which the state transition becomes infrequent as the time lag between consecutive moves goes to zero. There are some similarities in the

techniques used, although incomplete information introduces significant complications.[2]

More related are the papers by Athey and Bagwell (2001, 2008), Escobar and Toikka (2013), Barron (2013) and Renault, Solan and Vieille. All these papers assume that types are independent across players. Barron introduces imperfect monitoring in Escobar and Toikka (whose model generalizes most of the results of Athey and Bagwell), 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. As mentioned, none of these papers allow transitions to depend on actions. When specialized to the environments considered by Escobar and Toikka, our result provides a characterization of the asymptotic equilibrium payoff set in these environments, which in general is larger than the set that they identify.

Section 2 introduces the model and defines truthful equilibria. Mostly for pedagogical reasons, we start our analysis of independent private values with the special case in which monitoring is perfect and actions do not affect transitions (this is the environment of Escobar and Toikka). In Section 4, we drop these two restrictions but stick with independent private values. Section 5 indicates how one can obtain a true "folk theorem" by slightly relaxing the class of equilibria considered (and specializing to an environment in which there is any hope of achieving such a folk theorem –product monitoring). Section 6 defines the one-shot Bayesian game in full generality. Readers interested in the application to correlated values are referred to the working paper.

2 Model and Equilibrium

We consider dynamic games with imperfectly persistent incomplete information.

2.1 Extensive Form

The stage game is as follows. The finite set of players is denoted I. We assume that there are at least two players. Each player $i \in I$ has a finite set S^i of (private) states, or types,

[2]Among others, HSTV (as before FLM) rely on the equilibrium payoff set being full-dimensional, an assumption that fails with independent private values: When the players' types follow independent Markov chains and values are private, the players' limit equilibrium payoff must be independent of their initial type, given irreducibility and incentive-compatibility.

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. These reports are publicly observed;

3. The outcome of a public randomization device (p.r.d.) is observed. For concreteness, it is a draw from the uniform distribution on $[0, 1]$;[3]

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 $p(\cdot, \cdot \mid s_n, a_n) \in \Delta(S \times Y)$.

The stage-game payoff or *reward* of player i is a function $r^i : S \times A \to \mathbf{R}$, 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 $g^i : S \times A^i \times Y \to \mathbf{R}$, $r^i(s, a) = \mathbf{E}[g^i(s, a^i, y) \mid s, a]$. In that case, given (s, a^i, y), the realized reward does not convey additional information about a^{-i}, so that whether this reward is observed or not is irrelevant (for the updating of beliefs over a^{-i}, conditional on (s, a^i, y)) . We do not make this assumption, but assume instead that realized rewards are not observed. Hence, we assume that a player's private action, private state, the public signal and report profile are all the information available to him.

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

$$\sum_{n=1}^{+\infty} (1 - \delta)\delta^{n-1} r_n^i,$$

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

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

[3]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.

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

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

as well as

$$\pi_1(s) = \times_{i \in I} \pi_1^i(s^i),$$

for some transitions $\{p^i(\cdot \mid s^i, y)\}_{s^i, y} \subseteq \Delta(S^i)$, and distributions $\{p(\cdot \mid a)\}_a \subseteq \Delta(Y)$, $\pi_1^i \in \Delta(S^i)$, all $i \in I$, and (ii) rewards satisfy, for all $i \in I$, $s \in S$, $a \in A$, $r^i(s, a) = r^i(s^i, a)$. The first assumption guarantees that beliefs over state profiles are common knowledge throughout the game, on and off path. We assume *full support*: $\pi_1^i(s^i) > 0$, $p^i(t^i \mid s^i, y) > 0$ for all t^i, s^i and y, but allow $p(y \mid a) = 0$.

In Section 6, we extend our analysis to types that are not independent, and/or values that are not private. In the case of interdependent values, it matters whether players observe their payoffs or not. One can view privately observed payoffs as a special case of private values: simply define a player's private state as including his last realized payoff.[4] It would then be natural to allow for a second round of messages at the end of each period –and this second message could include both the realized payoff and the realized (private) action. In fact, our main characterization result extends immediately to the case in which monitoring is private, rather than public; see Section 6 for a discussion.

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 (as captured by the boundedness of the action simplex).

The game defined above allows for public communication among players. In doing so, we follow most of the literature on dynamic Bayesian games, see Athey and Bagwell (2001, 2008), Escobar and Toikka (2013), Renault, Solan and Vieille (2013), etc.[5] As in static Bayesian mechanism design, communication is required for coordination even in the absence of strategic motives; communication allows us to characterize what restrictions on payoffs, if any, are imposed by non-cooperative behavior.

[4]This interpretation is pointed out by AS. See also Mezzetti (2004) for the "static" (two rounds) counterpart.

[5]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.

Throughout, when a period is fixed and understood, we index variables relative to the previous period with an upper bar ($\bar{s}, \bar{a}$, etc.). Also, when referring to the following period, we use either t instead of s (for "t"omorrow's state), or label the variable with a prime.

2.2 Truthful Equilibria

2.2.1 Definition

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

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

$$M^i = S^i.$$

This is *a priori* restrictive. Because players cannot commit, the revelation principle does not apply (see Bester and Strausz, 2001), and richer message sets might lead to larger sets of equilibrium payoffs. Let $M := \times_{i \in I} M^i$.

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

Formal definitions require additional notation. A public history at the start of round $n \geq 1$ is a sequence $h_{\text{pub},n} = (m_1, y_1, \ldots, m_{n-1}, y_{n-1}) \in H_{\text{pub},n} := (M \times Y)^{n-1}$. Player i's private history at the start of round n is a sequence $h_n^i = (s_1^i, m_1, a_1^i, y_1, \ldots, s_{n-1}^i, m_{n-1}, a_{n-1}^i, y_{n-1}) \in H_n^i := (S^i \times M \times A^i \times Y)^{n-1}$. (Here, $H_1^i = H_{\text{pub},1} := \{\emptyset\}$.) A (behavior) strategy for player i is a pair of sequences $(\mathfrak{m}^i, \mathfrak{a}^i) = (\mathfrak{m}_n^i, \mathfrak{a}_n^i)_{n \in \mathbf{N}}$ with $\mathfrak{m}_n^i : H_n^i \times S^i \to \Delta(M^i)$, and $\mathfrak{a}_n^i : H_n^i \times S^i \times M \to \Delta(A^i)$, which specify i's report and action as a function of his private information, his current state and the report profile in the current round. Recall however that a p.r.d. is assumed, although it is omitted from the notation. A strategy profile $(\mathfrak{m}, \mathfrak{a})$ defines a distribution over finite and infinite histories in the usual way.

Definition 1 (Truthful Equilibrium) *A strategy $(\mathfrak{m}^i, \mathfrak{a}^i)$ is truthful if $\mathfrak{m}_n^i(h_n^i, s_n^i) = s_n^i$ for all histories h_n^i, $n \geq 1$, and $\mathfrak{a}^i(h_n^i, s_n^i, m_n)$ depends on $(h_{\text{pub},n}, s_n^i, m_n)$ only.*

[6]For clarity, we maintain the notational distinction.

The first requirement imposes players to always report their current state truthfully, after all histories. The second requires actions to depend on public information (and current state) only.[7]

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

2.2.2 Limitations

To appreciate why truthful equilibria are restrictive, consider a two-player game with perfect monitoring in which player 1 has two equiprobable states $s = t, b$, which are *i.i.d.* over time, while player 2 has only one state.[8] Players have two actions, $\{T, B\}$ and $\{L, R\}$. Further, suppose that player 1's payoff from T (resp., B) exceeds his payoff from playing B if the state is t (resp., b), and that his actions are not observed. Hence, in any truthful equilibrium, player 1 must play T (resp., B) whenever his state is t (resp., b).

This means that we cannot drive player 2's payoff below what he can get from taking a best reply to player 1's action. If his best-reply is strict, then we could achieve a lower equilibrium payoff by considering a non-truthful equilibrium –player 1 simply does not announce his state, leaving player 2 guessing what he should do.

It is clear that the argument is more general. Even with *i.i.d.* states, it is not usually possible to have a player be indifferent over several actions in more than one particular state in a truthful equilibrium.[9]

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

Whether or not such randomization is easy to achieve when players do not reveal their type is irrelevant: What matters for minmaxing or statistical detection of deviations is that a player's action be unpredictable, whether this is because he deliberately randomizes over

[7] This generalizes the familiar notion of public strategies to Bayesian games.

[8] For the case of *i.i.d.* states, FL's algorithm can be adapted, see Section 8 of FLM.

[9] In repeated games, players have a unique state, so this problem does not arise.

actions, or because his type determining his pure action cannot be inferred from his report. Hence, mixed minmaxing is consistent with a player playing a pure action given his type for all of them but one, as long as he does not disclose his type.

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

2.3 The Revelation Game

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

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

The other steps are unchanged. Payoffs are defined as before. The definition of strategies and of equilibrium is as before, with the obvious restriction to reports. In line with the previous definition, an equilibrium of the revelation game is *truthful* if $\mathfrak{m}_n^i(h_n^i, s_n^i) = s_n^i$ for all $i \in I$, $h_n^i \in H_n^i$, $n \geq 1$ and states $s_n^i \in S^i$.

We will be interested in the set of equilibrium payoffs of the revelation game that can be achieved for *some* ρ. Because players only affect actions via messages, the revelation game dispenses with obedience –in particular, individual rationality. Hence, the set of truthful equilibrium payoffs of the original game is a subset of the set of truthful equilibrium payoffs of the revelation game.[10]

3 Perfect Monitoring, Action-Independent Transitions

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

[10] *A priori*, this is not obvious for the set of all equilibrium payoffs, because in non-truthful equilibria, actions may depend on states, and not just on reports. Nonetheless, our results below imply that this is the case.

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

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

When defining feasible payoffs, the restriction to deterministic policies rather than arbitrary strategies is clearly without loss. Given the public randomization device, F is convex.

3.1 A Superset of Bayes Nash Equilibrium Payoffs

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

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

If actions can be dictated, knowing the state profile can only help. But if $\lambda^i < 0$, this information would be used against i's interests. Not surprisingly, player i is unlikely to be forthcoming about this. This suggests distinguishing players in the set $I_+(\lambda) := \{i : \lambda^i > 0\}$ from the others. Suppose that players in $I_+(\lambda)$ truthfully disclose their private state, while the remaining players choose a reporting strategy that is independent of their private state.

Define

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

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 $I_+(\lambda) = \emptyset$). Note that $\mathbf{E}_{\mu,\rho}\left[\lambda \cdot r(s,a)\right]$ is the long-run payoff vector when players report truthfully and use the policy ρ. 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 a superset of the set of *all* equilibrium payoff vectors.

Let NE_δ (resp., NE_δ^R) denote the equilibrium payoffs in the original (resp., revelation) game, given $\delta \in [0, 1)$.

Proposition 1 *Assume IPV. The limit sets of Bayes Nash equilibrium payoffs are contained in V^*:*

$$\limsup_{\delta\to 1} NE_\delta \subseteq V^*, \limsup_{\delta\to 1} NE_\delta^R \subseteq V^*.$$

Proof. This Proposition is implied by Proposition 3, whose proof is in Appendix A. Here we provide a sketch (for $\limsup_{\delta\to 1} NE_\delta$) in the case in which the initial belief $(\pi_1^i)_{i\notin I_+(\lambda)}$ is equal to the ergodic distribution $(\mu^i)_{i\notin I_+(\lambda)}$. Fix $\lambda \in \Lambda$. Fix also $\delta < 1$. Consider the Bayes Nash equilibrium $\sigma = (\mathfrak{m}, \mathfrak{a})$ 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 π_1). This equilibrium need not be truthful or in pure strategies. Consider $i \notin I_+(\lambda)$. Along with σ^{-i} and π_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 would specify if this had been i's report.[11] The new profile $(\sigma^{-i}, \tilde{\sigma}^i)$ need no longer be an equilibrium of the game. Yet, thanks to the IPV assumption, it gives players $-i$ the same payoff as 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 partially observable Markov decision problem, which admits a solution within the class of deterministic policies (on the state space $\times_{i\in I_+(\lambda)} S^i \times \times_{i\notin I_+(\lambda)} \Delta(S^i)$).

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

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

[11] To be slightly more formal: in a given round, the randomization device selects a report for player i according to the conditional distribution induced by 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.

	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 1: Payoffs of Example 1

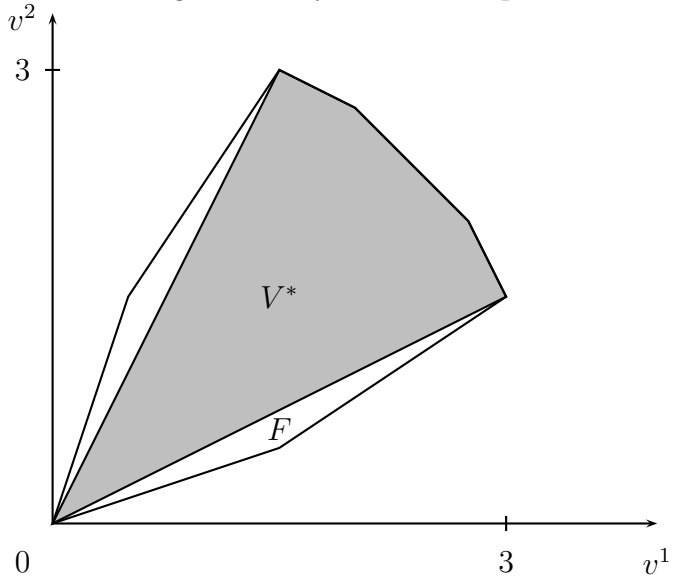

Figure 2: Incentive-compatible and feasible payoff sets in Example 1

As should be clear from the proof, Proposition 1 does not rely on $M^i = S^i$ and holds for any message space.

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

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

A set-theoretic 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, 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). It follows from our results and theirs that this is a subset of V^*. (In fact, the restriction to individually rational payoffs is not needed; it is not hard to show directly from the definition of V^* that $\mathrm{co}\,(Ext^{pu} \cup Ext^{po}) \subseteq V^*$.) In Example 1, this lower bound is tight, but this is not always the case.

3.2 The Average Cost Optimality Equation

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

Let be given an irreducible (or more generally unichain) transition function q over the finite set S with invariant measure μ, and a payoff function $u : S \to \mathbf{R}$.[12] 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.[13]

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

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

As mentioned, the lemma is standard in average cost dynamic programming, but a short direct proof is provided in the online appendix (appendix E).

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

[13] Lemma 1 defines the relative values for an exogenous Markov chain, or equivalently for an *arbitrary* policy. It is simply an "accounting" identity. The standard ACOE delivers more, as it provides a way of identifying the *optimal* policy: 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.

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.)

3.3 Characterization

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

Definition 2 *Fix $\lambda \in \Lambda$. Let*

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

where the supremum is taken over all $v \in \mathbf{R}^I$, $x : Y \to \mathbf{R}^I$ 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 function $r(a) + \sum_y p(y \mid a)x(y)$;*

(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 $\lim_{\delta \to 1} E_\delta = \mathcal{H}$.*

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

Because we insist on truthful equilibria, and because we need to incorporate the dynamic effects of actions on states, we must consider instead policies $\rho : S \to \times_{i\in I}\Delta(A^i)$ 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 state and current public outcome.

In what follows, the set of public outcomes in a given round is $\Omega_{\text{pub}} := S \times A$ (where the S-components stand for the reports). Let a policy $\rho : S \to \times_{i\in I}\Delta(A^i)$, and a map $x : S\times\Omega_{\text{pub}} \to \mathbf{R}^I$ be given. The vector $x(\bar{s}, \omega_{\text{pub}})$ is to be interpreted as transfers, contingent on previous reports $\bar{s}$, and on the current public outcome ω_{pub}.[14] 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 of pairs of reports (s_{n-1}, s_n). Let $\theta_{\rho,r+x} : S \times S \to \mathbf{R}^I$ denote the relative values of the players, obtained when applying Lemma 1 to the latter chain (and to all players).

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

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

(ii) each player i reports publicly some state $m^i \in S^i$, then chooses an action $a^i \in A^i$.

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

Given $\lambda \in \Lambda$, we denote by $\mathcal{P}_0(\lambda)$ the optimization program $\sup \lambda\cdot v$, where the supremum is computed over all payoff vectors $v \in \mathbf{R}^I$, policies $\rho : S \to \times_{i\in I}\Delta(A^i)$ and transfers $x : S \times \Omega_{\text{pub}} \to \mathbf{R}^I$ such that

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

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

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

[14] Conceptually, it might make sense to condition transfers on previous actions as well. This extension is not needed when transitions are action-independent.

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

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

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

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

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

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

Let TE_δ (TE_δ^R) denote the set of truthful equilibrium payoffs in the original (resp., revelation) game.

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

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

The same result holds true for the revelation game, weakening condition (a) in the definition of $\mathcal{H}_0(\lambda)$ by dropping the requirement that playing ρ be optimal. Let k_0^R denote the corresponding score.

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

Proposition 2 *For every $\lambda \neq -e^i$, $k_0(\lambda) = k_0^R(\lambda) = \bar{k}(\lambda)$.*
For $\lambda = -e^i$, $k_0(-e^i) \geq -\underline{v}^i$, and $k_0^R(-e^i) = \bar{k}(-e^i)$.

Set $V^{**} := \{v \in V^*, v \geq \underline{v}\}$. By Proposition 2, $V^{**} \subseteq \mathcal{H}_0$. Hence Theorem 2 implies the following.

Corollary 3 (Folk theorem, Special Case) *Assume that V^* has non-empty interior. Then*

$$\lim_{\delta\to 1} TE_\delta^R = V^*.$$

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

$$\liminf_{\delta\to 1} TE_\delta \supseteq V^{**}.$$

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

3.4 Proof overview

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

Let Z be a compact set with a smooth boundary contained in the interior of $\mathcal{H}_0$, and a discount factor $\delta < 1$ be given. Given a target payoff $z \in Z$, we construct recursively a truthful PBE candidate with payoff z.

Given a target payoff $z_n \in Z$ in round n, we set a direction $\lambda_n \in \mathbf{R}^I$ to be (heuristically) "the" normal vector to Z at z_n, and pick a feasible triple (v_n, ρ_n, x_n) in $\mathcal{P}_0(\lambda_n)$ such that $\lambda_n \cdot z_n < \lambda_n \cdot v_n$. The target payoff is publicly updated in round $n+1$ to

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

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

As in FLM (see also HSTV), the choice of λ_n and of (v_n, ρ_n, x_n) given z_n ensure that $z_{n+1} \in Z$, so that this recursive construction is well-defined. Moreover, the expected continuation payoff in round n (computed as of round 1) is $\mathbf{E}_\sigma[z_n]$. Fix now a history h_n up to round n along which all reports are truthful, with public part $h_{\text{pub},n}$. The choice of (ρ_n, x_n) and the updating formula (1) also ensure that truthful reporting followed by ρ_n is a truthful PBE outcome of the Bayesian game[15] with payoff $(1-\delta) r(s_n, a_n) + \delta z_{n+1}(h_{\text{pub},n}, \omega_{\text{pub},n}) + (1-\delta)\theta_{\rho_n, r+x_n}(s_n, s_{n+1})$, and the induced equilibrium payoff is $z_n(h_{\text{pub},n})$.[16]

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

Indeed, under this construction, the pair (ρ_n, x_n) is updated in round $n+1$, and the actual continuation relative values need not coincide with $\theta_{\rho_n, r+x_n}$. Whether a specific state reached in round $n+1$ is "good" relative to some other state depends on ρ_{n+1}, hence on λ_{n+1} and therefore on a_n through y_n. That is, even though player i's action choice has no influence on the distribution of s^i_{n+1}, it does affect the relative values of the different states in round

[15] The prior belief is unambiguously derived from the public history.

[16] That is, when computed under μ.

$n+1$. These changes in relative values would cancel in expectation, if states in round $n+1$ were drawn using the invariant measure μ. Yet, player i is computing expectations based on s_n^i. Hence, (conditional on $\omega_{\text{pub},n}$), player i's continuation payoff is not exactly equal to $z_{n+1}^i(h_{\text{pub},n}, \omega_{\text{pub},n})$ and state persistence thus affects the incentives faced by player i.[17]

In our construction, the above sketch is amended along the following lines. We first prove that any feasible triple (v, ρ, x) in $\mathcal{P}_0(\lambda)$ can be perturbed into some other triple, for which truth-telling incentives are strict. In other words, the value of $\mathcal{P}_0(\lambda)$ is unchanged when truth-telling incentives are required to be strict. We then divide the play into a sequence of phases of random duration. In effect, the p.r.d. chooses in each round with probability ξ, whether to start a new phase. When a new phase starts, target payoffs and policies are updated according to formulas derived from the FLM ones.

The switching probability ξ is set to be large compared to $(1-\delta)$, so that the expected contribution of a single phase to the overall payoff is small. Yet, ξ is set to be small, so that the expected duration of each phase is large.[18] The former property ensures that the recursive procedure is well-defined. The latter one ensures that, in any phase k under the plan $(v_{(k)}, \rho_{(k)}, x_{(k)})$, players perceive future payoffs as a small perturbation of the relative values $\theta_{\rho_{(k)}, r+x_{(k)}}$. Given that truth-telling incentives are strict in $\Gamma(\rho_{(k)}, x_{(k)})$, it thus remains optimal to report truthfully in the dynamic game.

4 Action-dependent Transitions, Imperfect Monitoring

We now generalize these results to the case in which monitoring is imperfect, and actions affect transitions. The environment is still of independent private values (IPV), which (cf. Section 2.1) requires that

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

and $\pi_1(s) = \times_{i \in I} \pi_1^i(s^i)$.[19] Proofs for this section are in Appendix B.

[17] The issue does not arise only when successive states are *i.i.d.* But then the dynamic game is truly the repetition of a one-shot Bayesian game, to which the results of FLM apply.

[18] To be clear, we pick $\xi(\delta)$ as a function of δ such that $\lim_{\delta \to 1} \xi(\delta) = 0$ and $\lim_{\delta \to 1} \dfrac{\xi(\delta)}{1-\delta} = +\infty$.

[19] For expositional simplicity, we assume here that states do not affect the signal distribution. There are important applications for which states do affect the distribution of signals. Theorem 6 below makes no restriction in this respect. See also the working paper for the analogs of the theorems developed in Sections 4–5 for the case in which states affect signal distributions.

4.1 The Superset Revisited

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

This example demonstrates that constant action choices no longer suffice to minimize or maximize a player's payoff, when his state is unknown to others and he fails to reveal it, even as $\delta \to 1$. Plainly, in the example, player 1's belief about the state of player 2 matters for the choice of an optimal action, and the chosen action matters for player 1's next belief. Hence, if we wish to describe player 1's choice as a Markov policy, we must augment the state space to account for player 1's belief. In the previous example, there is a binary sufficient statistic for this belief, namely, the last action chosen by player 1. Yet in general, the role of the belief is not summarized by such a simple statistic. It is necessary to augment the state space by (at least) an arbitrary summary statistic, which follows a Markov chain as well. The next result establishes that finite representations suffice, under our assumptions.[20]

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

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

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

[20] This is closely related to the literature on finite-state controllers in POMDP, see Yu and Bertsekas (2008).

initial distribution π_1.

Proposition 3 *Assume IPV. Then* $\limsup_{\delta\to 1} NE_\delta(\pi_1) \subseteq V_1^*$, *for all* π_1.[21]

4.2 Characterization

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

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

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

with $\omega_{\text{pub}} = (m, k, y)$.

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

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

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

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

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

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

[21] A more precise statement holds. For each $\eta > 0$, there is $\bar{\delta} < 1$ such that, for each discount factor $\delta \geq \bar{\delta}$ and each initial distribution $\pi_1 \in \times_{i\in I}\Delta(S^i)$, $NE_\delta(\pi_1)$ is included in the η-neighborhood $V_{1,\eta}^*$ of V_1^*.

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

$$\theta^i_{\vec{a},r}(s^i) - \theta^i_{\vec{b},r}(s^i) \neq \theta^i_{\vec{a},r}(\tilde{s}^i) - \theta^i_{\vec{b},r}(\tilde{s}^i). \tag{2}$$

When successive states are *i.i.d.*, **A1** is equivalent to the assumption of no-two-equivalent states made in Section 3.3. However, when **A1** is specialized to the case of action-independent states, it neither implies nor is implied by this assumption.[22]

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

A2 For all $a \in A$,

1. For all $i \neq j$, $p(\cdot \mid a) \notin \operatorname{co}\{Q^i(a) \cup Q^j(a)\}$.
2. For all $i \neq j$, $\operatorname{co}(p(\cdot \mid a) \cup Q^i(a)) \cap \operatorname{co}(p(\cdot \mid a) \cup Q^j(a)) = \{p(\cdot \mid a)\}$.

For $i \in I$, we set $\underline{v}^i := \min_{a^{-i} \in A^{-i}} \max_{\rho^i : S^i \to A^i} \mathbf{E}_{(s,a) \sim \mu_{(\rho^i, a^{-i})}} [r^i(s,a)]$.[23] Proposition 4 and Theorem 4 parallel the results of Section 3.3.

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

Theorem 4 (Folk theorem) *Assume that IPV and Assumption **A1**. If V_1^* has nonempty interior, then, for any π_1,*

$$\lim_{\delta \to 1} TE^R_\delta(\pi_1) = V_1^*.$$

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

$$\liminf_{\delta \to 1} TE_\delta(\pi_1) \supseteq V_1^{**}.$$

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

[22] Yet, all results below still hold when **A1** is weakened and it is only required that (2) holds for some *sequences* $\vec{a} = (a_n)_n$ and $\vec{b} = (b_n)$ in A –at the cost of a slight extension of the notion of relative value, and of notational complexity. The weakened assumption is strictly weaker than both **A1** and the no-two-equivalent-states assumption.

[23] This definition of minmax can be strengthened by considering extended policies for players $-i$. All results remain valid with this change.

5 Product Monitoring

This section strengthens the assumption **A2** on monitoring and considers a slightly larger class of equilibria. By doing so, we obtain an exact characterization of the asymptotic (Nash) equilibrium payoff set.

The reason why previous theorems failed to be characterizations is because of the minmax payoff. As mentioned, there are many examples in which the state-independent pure-strategy minmax payoff $\underline{v}^i$ coincides with the "true" minmax payoff

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

where the minimum is over the set of (independent) strategies by players $-i$. 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 auditing, corruption, etc. (starting with Becker and Stigler, 1974). More disturbingly, when $\underline{v}^i > w^i$, it can happen that $V_1^{**} = \emptyset$. Theorem 4 becomes meaningless, as the corresponding equilibria no longer exist. On the other hand, the set

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

is never empty.[24]

As is also well known, even when attention is restricted to repeated games, there is no reason to expect the punishment level w^i to equal the mixed-strategy minmax payoff commonly used (that lies in between w^i and $\underline{v}^i$), as w^i might only be obtained when players $-i$ use private strategies (depending on past action choices) that would allow for harder, coordinated punishments than those assumed in the definition of the mixed-strategy minmax payoff. 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.[25] 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.

[24]To see this, note that the state-independent *mixed* minmax payoff lies (weakly) 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 $k(\lambda^e)$.

[25]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.

Definition 3 *Monitoring has product structure if there are finite sets* $(Y^i)_{i=1}^I$ *such that* $Y = \times_i Y^i$*, and*

$$p(y \mid a) = \times_i p^i(y^i \mid a^i),$$

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

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

A3 For all i, a,

$$p(\cdot \mid a) \notin \operatorname{co} Q^i(a).$$

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

We maintain the non-degeneracy assumption introduced in Section 4.2. In the appendix C (with additional details in online Appendix F), 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 5 *Assume that monitoring has the product structure, and that Assumptions* ***A1*** *and* ***A3*** *hold. If* W *has non-empty interior, the set of (Nash, sequential) equilibrium payoffs converges to* W *as* $\delta \to 1$*.*

Because minmaxing requires unpredictability, and as explained in Section 2.2, unpredictability might be inconsistent with truthful equilibria, this requires using strategies that are not truthful, at least during "punishments."[26] Nonetheless, we show that a slight extension of the set of strategies considered so far, to allow for silent play during punishment-like phases, suffices.

Unlike in repeated games, imposing product structure does not guarantee that the minmax strategy is stationary: players $-i$ draw inferences from the public signal y^i about player i's action, hence about his private state, which can be exploited to adjust their action. Our construction relies on an extension of Theorem 2, as well as an argument inspired by Gossner

[26]We use quotation marks as there are no clearly defined punishment phases in recursive constructions (as in Abreu, Pearce and Stacchetti, 1990, or here), unlike in the standard proof of the folk theorem under perfect monitoring.

(1995), based on approachability theory (Blackwell, 1956). Roughly speaking, the argument is divided in two parts. First, we extend Theorem 2 to allow for "blocks" of T rounds, rather than single rounds, as the extensive form over which the score is computed. 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 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 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.[27] When testing player j, it is critical to condition on player i's realized signal, so as to incentivize player j to be unpredictable.

6 Dropping the IPV Assumption

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

Example 3—*A Silent Game.* This game follows Renault (2006). This is a zero-sum two-player game in which player 1 has two private states, s^1 and $\hat{s}^1$, and player 2 has a single

[27] 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.

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

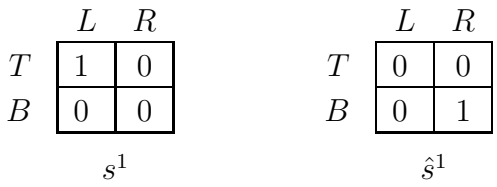

Figure 3: Player 1's reward in Example 3

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

Pick M^1 such that $\#M^1 \geq 2$, so that player 1 can disclose his state if he wishes. Will he? If player 1 reveals 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 1/4 by choosing reports and actions at random. In fact, this is the (uniform) value for $p = 1$ (Aumann and Maschler, 1995). When $p < 1$, player 1 can get more than this by trading off the higher expected reward from a given action with the information that it gives away (for instance, he can play T (B) with probability close to but above 1/2 when the state is s^1 ($\hat{s}^1$) so as to leave some uncertainty, yet repeat some benefits). He has no interest in giving this information away for free through informative reports. Truthful equilibria do not exist: all equilibria are babbling.

Just because we may focus on the silent game does not make it easier. Its (limit) value for arbitrary $p > 2/3$ is still unknown.[28] 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.

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

Example 4— ***Waiting for Evidence.*** There are two players. Player 1 has $K+1$ types, $S^1 = \{0, 1, \ldots, K\}$; player 2 has only two types, $S^2 = \{0, 1\}$. Transitions do not depend on actions (omitted), and are as follows. If $s_n^1 = k > 0$, 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 related points. First, in order for player $-i$ to statistically discriminate between player i's states, while simultaneously guaranteeing a given interim payoff to each player's type, 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_n^i, s_{n-1}), summarized by the distribution of an initial segment of the sequence? This question appears to be open in general. In the case of transitions that do not depend on actions, it has been raised by Blackwell and Koopmans (1957) and answered by Gilbert (1959): it is enough to consider the next $2\#S^i + 1$ values of the sequence $(s_{n'}^{-i})_{n' \geq n}$.[29]

At the very least, when types are correlated and the Markov chain exhibits time dependence, it is useful to condition player i's continuation payoff given his report about s_n^i on $-i$'s next private state, s_{n+1}^{-i}. Because this suffices to obtain sufficient conditions analogous

[29]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).

to those invoked in the static case, we limit ourselves to this conditioning in this section.[30]

6.1 A General Theorem

We now return to the general model, without restricting attention to either private values or independent types. In this section, $M^i := S^i \times A^i \times S^i$ for all i. This has to be interpreted as player i's state yesterday, his action yesterday, and his state today. In the spirit of Myerson (1986), we wish to allow player i to disclose all information that is relevant to his preferences and beliefs; in this case, with correlated types, his belief about $-i$'s type profile depends on the action he has taken, his type yesterday and his current type. Off path, none of these are known, and a player shouldn't find it impossible to disclose his beliefs if he happened to deviate in the previous round.

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

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

Let a policy $\rho : S \to \Delta(A)$, and a map (interpreted as transfer) $x : \Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S \to \mathbf{R}^I$ 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.This requirement will not be systematically stated, but it is assumed throughout. Note that, compared to IPV, we have added the current state profile t^{-i} as an argument of player i's transfer, given that this profile is statistical evidence about player i's state, as explained in Example 4.

Assuming states are truthfully reported and actions chosen according to 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).[31]

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

[31] 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.

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 $p(\cdot \mid s,a)$ and $\omega_{\text{pub}} = (m,y)$, and m' is the truthful report tomorrow determined by t. Here, one should interpret $\bar{\omega}_{\text{pub}}$ as the public information from yesterday.

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;[32] player i is informed of s^i, then submits a report $m^i \in M^i$;

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

$$r^i(s,a) + x^i(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t^{-i}) + \theta^i_{\rho,r+x}(\omega_{\text{pub}}, m'), \tag{3}$$

where a^{-i} is drawn according to $\rho^{-i}(s^{-i}, m_c^i)$, the pair (y,t) is drawn according to $p(\cdot \mid s,a)$, 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 4 *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).*

[32] 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 $p(s \mid \bar{y}, \bar{s}, \rho(\bar{s}))$. 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 $\tilde{a}^{-i}$ that is consistent with $\bar{y}$.

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 $p(\cdot \mid \bar{y}, \bar{s}, \bar{a})$. 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.[33]

The admissibility requirement in Definition 4 is demanding, however, in that it requires in addition truth-telling to be optimal for player i at any outcome $\bar{\omega}$ such that $(\bar{s}^{-i}, \bar{a}^{-i}) = (\bar{m}_c^{-i}, \rho^{-i}(\bar{m}_c))$, 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 4 puts no restriction on player i's behavior if he lies in Step 1 ($m^i \neq (\bar{s}^i, \bar{a}^i, s^i)$). The second part of Definition 4 is equivalent to saying that $\rho^i(s)$ is one best-reply to $\rho^{-i}(s)$ in the complete information game with payoff function given by (3) when $m = (\bar{s}, \bar{a}, s)$.

The requirement that truth-telling be uniquely optimal reflects an important difference between our approach to Bayesian games and the traditional approach of Abreu, Pearce and Stacchetti (1990) 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.

[33] 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.

Truth-telling incentives must be strict for the approximation involved not to affect them.[34]

We denote by $\mathcal{C}_2$ the set of admissible pairs (ρ, x).

For given weights $\lambda \in \Lambda$, we denote by $\mathcal{P}_2(\lambda)$ the optimization program $\sup \lambda \cdot v$, where the supremum is taken over all triples (v, ρ, x) such that

- $(\rho, x) \in \mathcal{C}_2$;
- $\lambda \cdot x(\cdot) \leq 0$;
- $v = \mathbf{E}_{\mu_\rho}[r(s,a) + x(\bar{\omega}_{\text{pub}}, \omega_{\text{pub}}, t)]$, where $\mu_\rho \in \Delta(\Omega_{\text{pub}} \times \Omega_{\text{pub}} \times S)$ is the invariant distribution under truth-telling and ρ, so that v is the long-run payoff induced by (ρ, x).

The three conditions mirror those of Definition 2 for the case of repeated games. The first condition (admissibility) and the third condition are the counterparts of the Nash condition in Definition 2(i); the second condition is the "budget-balance" requirement imposed by Definition 2(ii). We denote by $k_2(\lambda)$ the value of $\mathcal{P}_2(\lambda)$ and set $\mathcal{H}_2 := \{v \in \mathbf{R}^I, \lambda \cdot v \leq k_2(\lambda) \text{ for all } \lambda \in \Lambda\}$.

Theorem 6 *Assume that $\mathcal{H}_2$ has non-empty interior. Then, given π_1,*

$$\mathcal{H}_2 \subseteq \liminf_{\delta \to 1} TE_\delta(\pi_1).$$

This result (proved in Appendix D) is simple enough. For instance, in the case of "standard" repeated games with public monitoring, Theorem 6 generalizes FLM, yielding the folk theorem with the mixed minmax under their assumptions.

To be clear, there is no reason to expect Theorem 6 to provide a characterization of the entire limit set of truthful equilibrium payoffs. One might hope to achieve a larger set of payoffs by employing finer statistical tests (using the serial correlation in states, for instance), just as one can achieve a bigger set of equilibrium payoffs in repeated games than the set of PPE payoffs, by considering statistical tests (and private strategies). Example 4 makes plain that using only the current report of $-i$ as evidence for player i's truthfulness is *ad hoc*. Allowing for more signals/reports comes at an obvious cost in terms of the simplicity of the characterization.

Nonetheless, as we have shown in Sections 3–5, variants of this theorem suffice to establish "folk theorems" under IPV. Similarly, with correlated types, one can use arguments based

[34]Fortunately, this requirement is not demanding, as it is implied by standard full-rank conditions in the correlated case, and by our non-degeneracy condition in the IPV case.

on Crémer and McLean (1988) and Kosenok and Severinov (2008) to derive a folk theorem with appropriate full rank assumptions. See the working paper for details. But Example 3 illustrates the difficulties that arise under the ominous combination of independent types and common values.

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

Second, signals can be private. That is, we may replace Step 2 of the decision problem $D^i_{\rho,x}$ by: A profile $y_n = (y^i_n) \in Y := \times_i Y^i$ of private signals and the next state profile $s_{n+1} = (s^i_{n+1})_{i \in I}$ are drawn according to some joint distribution $p_{s_n, a_n} \in \Delta(S \times Y)$. 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 is unchanged, given the re-defined message space, and so does the statement of the theorem.

In a sense, this more general formulation is more natural, as the current one already reduces the program to a one-player decision-theoretic problem, in which each player reports his private information; he might as well report the signal he observed, and his realized reward, in case of known-own payoffs. This variation mirrors Kandori and Matsushima (1998)'s extension of FLM to private monitoring; the issues they raise regarding the possibility of a

folk theorem in truthful strategies under imperfect monitoring apply here as well.

7 Conclusion

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

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

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

References

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.

Becker, G.S. and G.J. Stigler (1974). "Law Enforcement, Malfeasance, and Compensation of Enforcers," *Journal of Legal Studies*, **3**, 1–18.

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

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.

Bressaud, X. and A. Quas (2014). "Asymmetric Warfare," arXiv:1403.1385.

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.

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.

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 (2011). "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.

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 (2014). "On the Limit Perfect Public Equilibrium Payoff Set in Repeated and Stochastic Games," *Games and Economic Behavior*, **85**, 70–83.

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

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

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.

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

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 (2014). "A Folk Theorem for Stochastic Games with Infrequent State Changes," *Theoretical Economics*, forthcoming.

Pęski, M. and J. Toikka (2014). "Value of Persistent Information," working paper, MIT.

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.

Yu, H. and D.P. Bertsekas (2008). "On near optimality of the set of finite-state controllers for average cost POMDP," *Mathematics of Operations Research*, **33**, 1–11.

A Proofs for Section 3

Here and in what follows, we focus on the statements involving truthful equilibria, *i.e.*, accounting for the obedience constraints. The corresponding results for the revelation game are immediate corollaries.

A.1 Proof of Theorem 2

We let Z be a compact convex set included in the interior of $\mathcal{H}_0$. Given $z \in Z$, we construct a truthful PBE σ with payoff z. Under σ, the play is divided into a sequence of phases of random duration. During any given phase, the players (report truthfully and) follow a policy $\rho_\lambda : S \to A$ that depends on a direction $\lambda \in \Lambda$. Players are incentivized to report truthfully and to follow the prescribed policy by means of "transfers," which are implemented via adjustments in the continuation payoff, updated at the beginning of each phase.

A.1.1 Preliminaries

We pick $\eta > 0$ small enough so that the η-neighborhood $Z_\eta := \{z \in \mathbf{R}^I, d(z, Z) \leq \eta\}$ is also included in the interior of $\mathcal{H}_0$. Since k_0 is lower semi-continuous, there exists $\varepsilon_0 > 0$ such that $\max_{z \in Z_\eta} \lambda \cdot z + 2\varepsilon_0 < k_0(\lambda)$ for all $\lambda \in \Lambda$.

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

Lemma 2 *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 any vector $w \in \mathbf{R}^I$ which satisfies $\|w - z\| \leq \zeta$ and $\lambda \cdot w \leq \lambda \cdot z - \varepsilon\zeta$ for some $\zeta < \bar{\zeta}$, belongs to Z_η.*

The equilibrium construction relies on Lemma 3 below.

Lemma 3 *There is a finite set $\mathcal{S}$ of triples (v, ρ, x) such that the following holds. For every direction $\lambda \in \Lambda$, there is an element (v, ρ, x) of $\mathcal{S}$ such that (i) (v, ρ, x) is feasible in $\mathcal{P}_0(\lambda)$ with strict truth-telling incentives and (ii) $\max_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.*

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

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

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

We next define $\rho_* : S \to \Delta(A)$ as $\rho_*(s) := \dfrac{1}{|I|} \sum_{i \in I} \rho_i(s^i)$ and $x_t : S \to \mathbf{R}^I$ as $x_t^i(s) := \dfrac{1}{|I|} x^i(s^i)$. We then define $x_{\text{ob}} : \Omega_{\text{pub}} \to \mathbf{R}^I$ as $x_{\text{ob}}^i(s, a) = 0$ if $a^i = \rho_*^i(s)$ and set $x_{\text{ob}}^i(s, a)$ to be a large negative constant otherwise, and set $x_* := x_t + x_{\text{ob}}$.[35] Since transitions are action-independent, for each i and $s \in S$, the expectation of the sum

$$r^i(s^i, (\tilde{a}^i, \rho_*^{-i}(\tilde{s}^i, s^{-i}))) + x_*^i((\tilde{s}^i, s^{-i}), (\tilde{a}^i, \rho_*^{-i}(m^i, s^{-i}))) + \theta^i_{\rho_*, r + x_*}(t)$$

of current payoffs, transfers x_*^i and continuation relative values $\theta^i_{\rho_*, r+x_*}$ has a strict maximum for $\tilde{s}^i = s^i$ and $\tilde{a}^i = \rho_*^i(s)$.

Let a direction $\lambda \in \Lambda$ be given, and subtract a constant to $x_*(\cdot)$ in order that $\lambda \cdot x_*(\cdot) < 0$. The long-run payoff associated with (ρ_*, x_*) is $v_* := \mathbf{E}_{\mu, \rho_*(s)} [r(s, a) + x_*(s, a)]$. The triple (v_*, ρ_*, x_*) is then feasible in $\mathcal{P}_0(\lambda')$ for all λ' close enough to λ, with strict truth-telling incentives.

Let now (v, ρ, x) be a feasible triple in $\mathcal{P}_0(\lambda)$ such that $\lambda \cdot x(\cdot) < 0$ and $\lambda \cdot v > k_0(\lambda) - \varepsilon_0$. For $\varepsilon > 0$, we denote by $(\rho_\varepsilon, x_\varepsilon)$ the pair obtained when letting the p.r.d. choose between (ρ, x) and (ρ_*, x_*) with probabilities $1 - \varepsilon$ and ε respectively. The long-run payoff associated to the pair $(\rho_\varepsilon, x_\varepsilon)$ is $v_\varepsilon := (1 - \varepsilon)v + \varepsilon v_*$.

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

Observe that, since transitions are action-independent, one has $\theta_{\rho_\varepsilon, r+x_\varepsilon}(\bar{s}, s) = (1-\varepsilon)\theta_{\rho, r+x}(\bar{s}, s) + \varepsilon\theta_{\rho_*, r+x_*}(s)$ for all $(\bar{s}, s)$. Using once again the assumption that transitions are action-independent, this is easily seen to imply that the triple $(v_\varepsilon, \rho_\varepsilon, x_\varepsilon)$ is feasible in $\mathcal{P}_0(\tilde{\lambda})$ for all $\tilde{\lambda}$ close to λ, with strict truth-telling incentives.

In addition $\lambda \cdot v_\varepsilon > k_0(\lambda) - \varepsilon_0 > \sup_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0$ for ε small enough. The result follows, since Λ is compact and the left-most and right-most expressions are continuous in λ. ■

We let κ be a common bound on v, x, and $\theta_{\rho, r+x}$ for $(v, \rho, x) \in \mathcal{S}$, and on $z \in Z$ and r. We pick an arbitrary $\varepsilon_1 \in (0, \varepsilon_0)$, and set $\varepsilon := \varepsilon_1/4\kappa$. We let then $\bar{\zeta}$ be obtained via Lemma 2 given ε.

We assume that $\bar{\delta} < 1$ is high enough so that the conditions (i–iv) are met for all $\delta \geq \bar{\delta}$: (i) $\xi := \sqrt{1-\delta} < \frac{1}{3}$, (ii) $\frac{1-\delta}{\delta\xi} < 1$, (iii) $\zeta := 4\kappa\frac{1-\delta}{\delta\xi} < \bar{\zeta}$ and (iv) $6\kappa\xi < \varepsilon_0 - \varepsilon_1$.

A.1.2 Strategies

For simplicity, we assume that the initial state s_1, together with a fictitious state s_0 for round 0, is drawn according to μ. Let $z \in Z$ be the desired equilibrium payoff. The play is divided into a sequence of phases. The durations of the successive phases form a sequence of *i.i.d.* random variables. The initial round $\tau_{(k)}$ of phase k, $k \geq 1$, is set as follows: $\tau_{(1)} = 0$; in each round n, the p.r.d. decides with probability ξ whether to start a new phase.[36]

In round $\tau_{(k+1)}$, a target payoff $z_{(k+1)} \in \mathbf{R}^I$, a direction $\lambda_{(k+1)} \in \Lambda$, and a triple $(v_{(k+1)}, \rho_{(k+1)}, x_{(k+1)}) \in \mathcal{S}$ are updated based on past public play, together with an auxiliary target $w_{(k+1)} \in \mathbf{R}^I$.

We first update $w_{(k+1)}$ according to

$$\xi w_{(k+1)} + (1-\xi) z_{(k)} = \frac{1}{\delta} z_{(k)} - \frac{1-\delta}{\delta} v_{(k)} + \frac{1-\delta}{\delta} x_{(k)}(m_{\tau_{(k+1)}-2}, \omega_{\text{pub}, \tau_{(k+1)}-1}). \tag{4}$$

Next, we apply Lemma 2 with $z = w_{(k+1)}$ to get $\lambda_{(k+1)}$ and we next apply Lemma 3 with $\lambda_{(k+1)}$ to get $(v_{(k+1)}, \rho_{(k+1)}, x_{(k+1)}) \in \mathcal{S}$. Finally, we update $z_{(k+1)}$ to

$$z_{(k+1)} := w_{(k+1)} + (1-\delta)\left(\left(1 + \frac{1-\delta}{\delta\xi}\right)\theta_{(k)}(m_{\tau_{(k+1)}-1}, m_{\tau_{(k+1)}}) - \theta_{(k+1)}(m_{\tau_{(k+1)}-1}, m_{\tau_{(k+1)}})\right), \tag{5}$$

where $\theta_{(k)}$ is a short-hand notation for $\theta_{\rho_{(k)}, r+x_{(k)}}$.

[36] Thus, the duration $\Delta_k := \tau_{(k+1)} - \tau_{(k)}$ of phase k is such that $\Delta_k - 1$ follows a geometric distribution with parameter $1 - \xi$.

Updating thus takes place after the outcome of the p.r.d. is observed in round $\tau_{(k+1)}$. The left-hand side in (4) accounts for the random duration of the phases. The auxiliary variable $w_{(k+1)}$ and the extra term in (5) (when compared to FLM) serve to adjust continuation relative values along the play, as will be apparent.

The construction is initialized with $w_{(1)} = z$, which is used to define $\lambda_{(1)}$, and $(v_{(1)}, \rho_{(1)}, x_{(1)})$ and $z_{(1)}$ using (5) (with $\theta_{(0)} := 0$). That this recursive construction is well-defined follows from Lemma 4 below.

Lemma 4 *For all k (and all public histories), one has $w_{(k)} \in Z_\eta$.*

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

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

Observe also that

$$\begin{aligned}
w_{(k+1)} - w_{(k)} &= w_{(k+1)} - z_{(k)} + z_{(k)} - w_{(k)} && (6)\\
&= \frac{1-\delta}{\delta\xi}\left\{z_{(k)} - v_{(k)} + x_{(k)}(m_{\tau_{(k+1)}-2}, \omega_{\mathrm{pub},\tau_{(k+1)}-1})\right\} + z_{(k)} - w_{(k)} && (7)\\
&= \frac{1-\delta}{\delta\xi}\left(w_{(k)} - v_{(k)} + x_{(k)}(m_{\tau_{(k+1)}-2}, \omega_{\mathrm{pub},\tau_{(k+1)}-1})\right) + \left(1 + \frac{1-\delta}{\delta\xi}\right)(z_{(k)} - w_{(k)}) && (8)
\end{aligned}$$

so that

$$\begin{aligned}
\lambda_{(k)} \cdot (w_{(k+1)} - w_{(k)}) &\leq -\frac{1-\delta}{\delta\xi}\varepsilon_0 + 6\kappa(1-\delta)\\
&\leq -\frac{1-\delta}{\delta\xi}\varepsilon_1 + \frac{1-\delta}{\delta\xi}\ (\varepsilon_1 - \varepsilon_0 + 6\kappa\delta\xi)\\
&\leq -\varepsilon\zeta.
\end{aligned}$$

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

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

Under σ, each player i reports truthfully $m_n^i = s_n^i$ at the report step. At the action step, player i plays $\rho_n^i(m_n)$ if he reported truthfully $m_n^i = s_n^i$. In the (off-path) event $m_n^i \neq s_n^i$, player i plays a best-reply to $\rho^{-i}(m_n)$ in the complete information game with payoff $r(s_n, a) + x_n(m_{n-1}, (m_n, a)) + \mathbf{E}_{p(\cdot|s_n^i, m_n^{-i})}\theta_n(m_n, s_{n+1})$ (where $s_n^{-i} = m_n^{-i}$).

A.1.3 Equilibrium Properties

Given a round n, we denote by γ_n the expected continuation payoff under σ, conditional on the public history at round n (up to and including the outcome of the p.r.d.).[37]

Lemma 5 *One has* $\gamma_n = z_n + (1-\delta)\theta_n$.

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

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

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

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

$$z_n + (1-\delta)\theta_n = (1-\delta)r(s_n, \rho_n(s_n)) + \delta \mathbf{E}\left[z_{n+1} + (1-\delta)\theta_{n+1} \mid h_{\text{pub},n}\right]. \tag{9}$$

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

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

$$\begin{aligned}
& (1-\xi)\left(z_{(k)} + (1-\delta)\theta_{(k)}(s_n, s_{n+1})\right) + \xi\left(z_{(k+1)} + (1-\delta)\theta_{(k+1)}(s_n, s_{n+1})\right) \qquad (10)\\
= \;& (1-\xi)\left(z_{(k)} + (1-\delta)\theta_{(k)}(s_n, s_{n+1})\right) + \xi\left(w_{(k+1)} + (1-\delta)\left(1 + \frac{1-\delta}{\delta\xi}\right)\theta_{(k)}(s_n, s_{n+1})\right)\\
= \;& \frac{1-\delta}{\delta}\theta_{(k)}(s_n, s_{n+1}) + \left(\frac{1}{\delta}z_{(k)} - \frac{1-\delta}{\delta}v_{(k)} + \frac{1-\delta}{\delta}x_{(k)}(s_{n-1}, \omega_{\text{pub},n})\right),
\end{aligned}$$

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

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

$$\begin{aligned}
& (1-\delta)r(s_n, \rho_n(s_n)) + \delta \mathbf{E}\left[z_{n+1} + (1-\delta)\theta_{n+1} \mid h_{\text{pub},n}\right] \qquad (11)\\
= \;& z_{(k)} + (1-\delta)\left\{\mathbf{E}_{\rho_n(s_n), p(\cdot|s_n)}\left[r(s_n, a_n) + x_{(k)}(s_{n-1}, \omega_{\text{pub},n}) + \theta_{(k)}(s_n, s_{n+1})\right] - v_{(k)}\right\}\\
= \;& z_{(k)} + (1-\delta)\theta_{(k)}(s_{n-1}, s_n) = z_n + (1-\delta)\theta_n,
\end{aligned}$$

as desired – where the last equality uses the ACOE. ■

[37] That is, denote by $\mathcal{H}_{\text{pub},n}$ the (round n, public information) algebra on plays, by $\mathbf{P}_\sigma$ the probability distribution over plays induced by σ, and by $\mathbf{E}_\sigma$ the expectation operator under $\mathbf{P}_\sigma$. Then $\gamma_n := (1-\delta)\mathbf{E}_\sigma\left[\sum_{u=0}^{+\infty}\delta^u r_{n+u} \mid \mathcal{H}_{\text{pub},n}\right]$ –in particular, γ_n is computed under the "assumption" that $m_n = s_n$.

Corollary 7 *σ is a truthful PBE with expected payoff z.*

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

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

$$z_n^i - v_n^i + (1-\delta)\left\{\mathbf{E}_{(\beta^i(m_n),\rho_n^{-i}(m_n))}\left(r^i(s_n^i,a_n) + x_n^i(m_{n-1},\omega_{\text{pub},n})\right) + \mathbf{E}_{p(\cdot|s_n)}\theta_n^i(m_n,s_{n+1})\right\}. \tag{12}$$

We thus need to check that the expectation (over s_n^{-i}) in (12) is maximized when reporting truthfully.[38] Conditional on the p.r.d. choosing *not* to switch to a new block in round n, the expected continuation payoff of player i given h_n^i, is equal (up to the constant term $z_n^i - v_n^i$) to the interim expected payoff of i in the game $\Gamma(\rho_n, x_n)$ when reporting m_n^i and playing β^i (given $(m_{n-1}^i, s_{n-1}^{-i})$ and s_n^i, and multiplied by $1-\delta$). From the strict optimality of truth-telling in the game $\Gamma(\rho_n, x_n)$, it follows that any incorrect report $m_n^i \neq s_n^i$ leads to a loss of the order of $1-\delta$ (compared to truth-telling).

Conditional on the p.r.d. choosing to switch to a new block, lying may improve the expectation of (12) by an amount of the order of at most $1-\delta$. Since the probability of switching is only ξ, truth-telling is optimal (for δ close to 1).

■

A.2 Proof of Proposition 2

We here prove Proposition 2. In this section, and in later sections as well, we find it convenient to use the notion of a truthful pair. Given here $\rho : S \to \times_{i\in I}\Delta(A^i)$ and $x : \Omega_{\text{pub}} \to \mathbf{R}^I$, the pair (ρ, x) is *truthful* if it is optimal for each $i \in I$ to report truthfully in $\Gamma(\rho, x)$, assuming players $-i$ do so, and ρ is played at the action step. The pair (ρ, x) is (*strictly*) *ex post* truthful if it is (strictly) ex post optimal to report truthfully in $\Gamma(\rho, x)$. That is, for each i

[38] Following a truthful report $m_n^i = s_n^i$, the optimality of ρ_n in $\Gamma(\rho_n, x_n)$ ensures that obedience is optimal in round n.

and $s \in S$, the expectation (over t) of

$$r^i(s^i, \rho(\tilde{s}^i, s^{-i})) + x^i((\tilde{s}^i, s^{-i}), \rho(\tilde{s}^i, s^{-i})) + \theta^i_{\rho, r+x}(t)$$

has a (strict) maximum for $\tilde{s}^i = s^i$.

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

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

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

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

$$\bar{k}(\lambda) + \theta_\lambda(s_+) = \max_{a \in A} \{r_\lambda(s_+, a)\} + \mathbf{E}_{p(\cdot|s_+)} \theta_\lambda(t_+) \text{ for all } s_+ \in S_+. \tag{13}$$

Pure optimal policies $\rho : S_+ \to A$ are characterized by the property that $\rho(s_+)$ achieves the maximum in (13) for each $s_+ \in S_+$.

We let $\rho_\lambda : S_+ \to A$ be an arbitrary optimal policy. We construct transfers $x : S \times \Omega_{\text{pub}} \to \mathbf{R}^I$ such that $(\bar{k}(\lambda), \rho_\lambda, x)$ is feasible in $\mathcal{P}_0(\lambda)$, thereby showing $k_0(\lambda) \geq \bar{k}(\lambda)$.[39]

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

Claim 8 *There exists $x_1 : S \to \mathbf{R}^I$ such that (ρ_λ, x_1) is (ex post) truthful.*

Proof. For $i \notin I_+$, it suffices to set $x^i_1 = 0$ as the reports by i are ignored. Fix now $i \in I_+(\lambda)$. For $s \in S$, define $x^i_1(s)$ by the equation

$$\lambda^i x^i_1(s) := r_\lambda(s_+, \rho_\lambda(s_+)) - \lambda^i r^i(s^i, \rho_\lambda(s_+)).$$

[39] In this section, we only deal with the policy ρ_λ, and will drop the reference to ρ_λ when writing relative values.

Observe that $\lambda^i\theta^i_{r+x_1} : S_+ \to \mathbf{R}$ satisfies (13) as well. Hence, $\lambda^i\theta^i_{r+x_1} = \theta_\lambda$ up to an additive constant.

Since $\rho_\lambda(s_+)$ achieves the maximum in (13), it follows that (ρ_λ, x_1) is truthful. ■

Note that $x_1 = 0$ if $\lambda = e^i$ for some i, so that $\lambda \cdot x_1(\cdot) = 0$. We set $x_t = x_1$ in that case. From now on, we assume that λ is not a coordinate vector and adapt arguments from AS.

For $i \in I$, define first $x^i_2 : S \times S^i \to \mathbf{R}$ by $x^i_2(\bar{s}, s^i) := \mathbf{E}_{s^{-i}\sim p(\cdot|\bar{s}^{-i})}[x^i_1(s)]$. Plainly, (ρ_λ, x_2) is truthful as well. The relative values $\theta^i_{x_2} : S \times S^i \to \mathbf{R}$ solve

$$\gamma + \theta^i_{x_2}(\bar{s}, s^i) = x^i_2(\bar{s}, s^i) + \mathbf{E}_{s^{-i}\sim p(\cdot|\bar{s}^{-i}),t^i\sim p(\cdot|s^i)}\theta^i_{x_2}(s, t^i) \tag{14}$$

where $\gamma = \mathbf{E}_{(\bar{s},s^i)\sim\mu}[x^i_2(\bar{s}, s^i)]$.

Define next $x^i_3 : S \times S^i \to \mathbf{R}$ as

$$x^i_3(\bar{s}, s^i) := \theta^i_{x_2}(\bar{s}, s^i) - \mathbf{E}_{\tilde{s}^i\sim p(\cdot|\bar{s}^i)}\theta^i_{x_2}(\bar{s}, \tilde{s}^i).$$

Claim 9 *The pair (ρ_λ, x_3) is truthful.*

Proof. One has $\mathbf{E}_{s^i\sim p(\cdot|\bar{s}^i)}[x^i_3(\bar{s}, s^i)] = 0$ for each $\bar{s}$, hence the equality

$$x^i_3(\bar{s}, s^i) = x^i_3(\bar{s}, s^i) + \mathbf{E}_{s^{-i}\sim p(\cdot|\bar{s}^{-i}),t^i\sim p(\cdot|s^i)}\left[x^i_3(s, t^i)\right]$$

holds. That is, $x^i_3 = \theta^i_{x_3}$.

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

$$r^i(s^i, \rho_\lambda(m)) + x^i_3(\bar{s}, m^i) + \theta^i_{r+x_3}(m, t^i) \tag{15}$$

is equal to

$$r^i(s^i, \rho_\lambda(m)) + \theta^i_{x_2}(\bar{s}, m^i) + \theta^i_r(m, t^i) + \theta^i_{x_3}(m, t^i)$$

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

$$r^i(s^i, \rho_\lambda(m)) + \theta^i_{x_2}(\bar{s}, m^i) + \theta^i_r(m, t^i) + \theta^i_{x_2}(m, t^i) - \mathbf{E}_{\tilde{t}^i\sim p(\cdot|m^i)}\left[\theta^i_{x_2}(m, \tilde{t}^i)\right]. \tag{16}$$

In view of (14), the expectation of (16) (and therefore of (15)) when $s^{-i} \sim p(\cdot \mid \bar{s}^{-i})$ and $t^i \sim p(\cdot \mid s^i)$, is equal to the expectation of

$$r^i(s^i, \rho_\lambda(m)) + x^i_2(\bar{s}, m^i) + \theta^i_{r+x_2}(m, t^i).$$

Since (ρ_λ, x_2) is truthful, so is (ρ_λ, x_3). ■

Claim 10 *Let $\mu_{ij} \in \mathbf{R}$ be arbitrary. For $i \in I$, set*

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

Then (ρ_λ, x_4) is truthful.

Proof. Fix $i \in I$, $\bar{s} \in S$, $s^i \in S^i$. For given $s^{-i} = m^{-i}$ and $t \in S$, the sum

$$r^i(s^i, \rho_\lambda(m)) + x_4^i(\bar{s}, m) + \theta_{r+x_4}^i(m, t) \tag{17}$$

is equal, thanks to $\theta_{x_3}^j = x_3^j$, to

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

In this latter expression, and for fixed $j \neq i$, $x_3^j(\bar{s}, m^j)$ is independent of m^i and $\mathbf{E}_{t^j \sim p(\cdot|s^j)} \left[x_3^j(m, t^j) \right] = 0$. Since (ρ_λ, x_3) is truthful, so is (ρ_λ, x_4). ■

Since λ is not a coordinate vector, the system $\lambda^i + \sum_{j \neq i} \lambda^j \mu_{ji} = 0$ $(i \in I)$ has a solution (μ_{ij}). With this choice, $\lambda \cdot x_4(\cdot) = 0$ and we set $x_t(\cdot) := x_4(\cdot)$.

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

The triple $(\bar{k}(\lambda), \rho_\lambda, x_t + x_{\rho_\lambda})$ is feasible in $\mathcal{P}_0(\lambda)$.

To conclude, we provide a short proof of the reverse inequality $k_0(\lambda) \leq \bar{k}(\lambda)$ for all $\lambda \in \Lambda$ (which is not needed for deriving Corollary 3). The proof uses the same idea as the proof of Proposition 1. Let (v, ρ, x) be feasible in $\mathcal{P}_0(\lambda)$. We modify ρ and x by letting the p.r.d. pick a fictitious report $\tilde{s}^j \sim \mu^j$ for all $j \notin I_+$ and let actions and transfers be determined by ρ and x, using these fictitious reports and the actual reports of players $i \in I_+$. Denote by $\tilde{\rho} : S_+ \to \times_{i \in I} \Delta(A^i)$ the modified policy and by $\tilde{x}$ the modified transfers. Since (v, ρ, x) is feasible and thanks to the private values assumption, all players $j \notin I_+$ are weakly worse off in $\Gamma(\tilde{\rho}, \tilde{x})$ while players $i \in I_+$ are unaffected. Since $\lambda \cdot \tilde{x}(\cdot) \leq 0$, this implies $\lambda \cdot v \leq \mathbf{E}_{\mu, \tilde{\rho}}[\lambda \cdot r(s, a)] \leq \bar{k}(\lambda)$.

B Proofs for Section 4

B.1 Proof of Proposition 3

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

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

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

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

$$\lim_{\delta \to 1} v_\delta(s_+, \pi_-) = \bar{k}_1(\lambda), \text{ for all } (s_+, \pi_-). \tag{18}$$

We will prove (18) by approximating $\mathcal{M}_\lambda$ with MDPs with a finite state space and using results from the theory of such MDPs. We introduce some piece of notation, to be used later as well. Given a finite subset K^i of $\Delta(S^i)$, a map $\phi^i : K^i \times Y \to K^i$ and $\eta > 0$, the pair (K^i, ϕ^i) is an η-*approximation* of $\Delta(S^i)$ if

$$\|\phi^i(k^i, y) - p^i(\cdot \mid k^i, y)\|_\infty < \eta, \tag{19}$$

for every $k^i \in K^i$ and $y \in Y$. Intuitively, (19) entails that the exact posterior on the next state given a prior k^i on the current state s^i and a signal y, is η-close to $\phi^i(k^i, y)$. That is, the map ϕ^i is a good approximation of the evolution of beliefs over states.[40]

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

We will without further notice assume that all η-approximations below satisfy the following *communication property*: for any two $k, \tilde{k} \in K$, there exists an integer $N \in \mathbf{N}$, action profiles $a_1, \ldots, a_N$, and signals $y_1, \ldots, y_N$, such that (i) $p(y_n \mid a_n) > 0$ for each n, and (ii)

[40] Note though that this interpretation is valid only if y is uninformative about s^i. Note also that we do not require that K^i be a "large" subset of $\Delta(S^i)$.

the sequence (k_n) defined by $k_1 = k$ and $k_{n+1} = \phi(k_n, y_n)$ is such that $k_{N+1} = \tilde{k}$.[41] Given an η-approximation (K, ϕ) of Δ_-, we define $\mathcal{M}_\phi$ to be the MDP with finite state space $S_+ \times K$, action set A, and transitions deduced from $p(\cdot \mid a) \in \Delta(Y)$ and ϕ. Finally, the stage reward function is (the restriction of) r_λ. Thus, $\mathcal{M}_\phi$ differs from $\mathcal{M}_\lambda$ only through the transition function, and we think of $\mathcal{M}_\phi$ as a finite state approximation of $\mathcal{M}_\lambda$. The MDP $\mathcal{M}_\phi$ is communicating.[42]

Let $\varepsilon > 0$ be arbitrary and let $\bar{r}$ be an upper bound on $\|r\|$. Since the transition function $p(\cdot \mid s, a)$ is aperiodic and irreducible, there exists a constant $c \in (0, 1)$ such that for each $(a_s)_{s \in S}$, and any two distributions π and $\tilde{\pi}$ in $\times_{i \in I} \Delta(S^i)$, one has $\| \sum_{s \in S} p(\cdot \mid s, a_s) (\pi_s - \tilde{\pi}_s) \|_\infty \leq c \|\pi - \tilde{\pi}\|_\infty$. Pick $\eta < \varepsilon(1-c)/\bar{r}$, and an η-approximation (K, ϕ) of Δ_-.

In both MDPs $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$, strategies map past public signals (y_n) and past (and current) states $(s_{+,n})$ of players in I_+ into an action profile. We prove in Lemma 6 below that any strategy induces approximately the same payoff in $\mathcal{M}_\lambda$ and in $\mathcal{M}_\phi$. Given a strategy σ, we denote by $\gamma_\delta(\cdot, \sigma)$ and $\gamma_{\delta,\phi}(\cdot, \sigma)$ the payoff induced in $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$ respectively, as a function of the initial state.

Lemma 6 *For every discount factor $\delta < 1$, any $s_+ \in S_+$, $\pi_- \in \Delta_-$ and $k \in K$, one has*

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

Proof. Fix σ and an arbitrary play $h_\infty = (s_{+,n}, y_n, a_n)_n$. Given a player $i \in I_-$ and a round $n \in \mathbf{N}$, let $\pi_n^i \in \Delta(S^i)$ and k_n^i be the i-th component of the state in $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$ along h_∞.[43] Along h_∞, the payoff difference in $\mathcal{M}_\lambda$ and $\mathcal{M}_\phi$ is

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

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

[41] The existence of communicating η-approximations is easy to establish. Not all η-approximations are communicating.

[42] Using the full-support assumption and the communicating property of (K, ϕ).

[43] That is, π_n^i is the conditional distribution of s_n^i given past signals, while k_n^i is obtained by repeated applications of ϕ^i.

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

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

Proof. By Lemma 6, one has both

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

Since ε is arbitrary, this implies the convergence of v_δ as $\delta \to 1$, with $|\lim_\delta v_\delta - v_\phi| \leq \varepsilon$. ■

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

Proof. Plainly, the policy ρ_* may also be either viewed as a strategy in $\mathcal{M}_\lambda$, or as a policy in the initial game with state space S, independent of the states of players $i \in I_-$. Under both "interpretations," the payoff induced by ρ_* is of course equal to $\gamma_\delta(\cdot, \rho_*)$. According to the first interpretation, Lemma 6 applies for each δ, and $\limsup_{\delta\to 1} \|\gamma_\delta(\cdot, \rho_*) - \gamma_{\delta,\phi}(\cdot, \rho_*)\|_\infty \leq \varepsilon$. According to the second interpretation, ρ_* induces a Markov chain over $S \times K$. Let $E = S \times K_E$ be an arbitrary ergodic set[44] for this Markov chain, with invariant measure $\mu_E \in \Delta(S \times K_E \times A)$. Given an initial state $(\bar{s}, \bar{k}) \in E$, one has $\lim_{\delta\to 1} \gamma_{\delta,\phi}((\bar{s}, \bar{k}), \rho_*) = v_\phi$ (by the choice of ρ_*), while $\lim_{\delta\to 1} \gamma_\delta((\bar{s}, \bar{k}), \rho_*) = \mathbf{E}_{\mu_E}[\lambda \cdot r(s, a)]$. Combining these results, one gets

$$v_\phi \leq \mathbf{E}_{\mu_E}[\lambda \cdot r(s, a)] + \varepsilon. \tag{20}$$

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

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

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

[44] That all ergodic sets are product sets follows from the full support assumption.

B.2 Proof of Proposition 4 and Theorem 4

B.2.1 An overview

To unify notations, we set $\tilde{k}_1(-e^i) = -\underline{v}^i$ for $i \in I$, and $\tilde{k}_1(\lambda) = \bar{k}_1(\lambda)$ otherwise, so that $V_1^{**} = \{z \in \mathbf{R}^I, \lambda \cdot z \leq \tilde{k}_1(\lambda) \text{ for all } \lambda\}$. We observe that $\bar{k}_1(\cdot)$ is lower semi-continuous, and that $\bar{k}_1(-e^i) \geq -\underline{v}^i$. Thus, $\tilde{k}_1(\cdot)$ is lower semi-continuous as well.

We will prove the following strengthening of Proposition 4.

Lemma 7 *For every $\lambda \in \Lambda$ and $\varepsilon > 0$, there exists a triple $(v, \rho_{\text{ext}}, x)$, which is feasible in $\mathcal{P}_1(\lambda)$, with strict truth-telling incentives, and such that $\lambda \cdot v > \tilde{k}_1(\lambda) - \varepsilon$.*

Lemma 7 readily implies $k_1(\lambda) \geq \tilde{k}_1(\lambda)$. The following subsections are devoted to the proof of Lemma 7.

In the meantime, we deduce Theorem 4 from Lemma 7. We let Z be a compact set included in the interior of V_1^{**}. Since Z is compact, there exists $\eta > 0$ such that the η-neighborhood Z_η of Z is also included in the interior of V_1^{**}. Thus, for all $\lambda \in \Lambda$, there is $\varepsilon > 0$ such that $\max_{z \in Z_\eta} \lambda \cdot z + \varepsilon < \tilde{k}_1(\lambda)$. Hence, by compactness of Λ and since $\tilde{k}_1$ is lower semi-continuous, there is $\varepsilon_0 > 0$ such that

$$\forall \lambda \in \Lambda, \max_{Z_\eta} \lambda \cdot z + 2\varepsilon_0 < \tilde{k}_1(\lambda). \tag{21}$$

Lemma 8 *There exists a finite set $\mathcal{S}$ of triples $(v, \rho_{\text{ext}}, x)$ such that the following holds. For every direction $\lambda \in \Lambda$, there is an element $(v, \rho_{\text{ext}}, x)$ of $\mathcal{S}$ such that*

1. *$(v, \rho_{\text{ext}}, x)$ is feasible in $\mathcal{P}_1(\lambda)$ with strict truth-telling incentives.*
2. *$\max_{z \in Z_\eta} \lambda \cdot z + \varepsilon_0 < \lambda \cdot v$.*

Proof. Given $\lambda \in \Lambda$, apply Lemma 7 with $\varepsilon = \varepsilon_0$. Plainly, by adding a small constant to x, we may assume that in addition $\lambda \cdot x(\cdot) < 0$ for fixed λ, hence $(v, \rho_{\text{ext}}, x)$ is feasible in $\mathcal{P}_1(\lambda')$ for all λ' close enough to λ. By (21), the inequality $\lambda \cdot v > \tilde{k}_1(\lambda) - \varepsilon_0$ implies

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

Since both sides of the inequality are continuous in λ and since Λ is compact, Lemma 8, and therefore Theorem 4, follows from Lemma 7. ■

Lemma 8 is the exact analog of Lemma 3 (in the proof of Theorem 2). Inspection of the proof of Theorem 2 then shows that Lemma 2 is still valid here and that all subsequent

arguments based on Lemmas 2 and 3 remain valid (that is, both the construction of strategies in Section 3 and the results from Section A.1.3 readily extend to the present, more general setup).[45] Thus, Theorem 4 follows.

B.2.2 Step 1: There is a strictly truthful pair $(\rho_{\text{ext},0}, x_0)$

We start to prove Lemma 8. In this first step, we construct a specific *ex post*, strictly truthful (pure) pair $(\rho_{\text{ext},0}, x_0)$. It will later be used as a perturbation, and will thus play a role analog to that of the pair (ρ_*, x_*) in Section 3.

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

For $s \in S$, define then $\mu(s) := \frac{1}{|I|}\sum_{i\in I}\mu^i(s^i)$ and $T^i(s) := \frac{1}{|I|}\tau^i(s^i)$ so that, for each $i \in I$ and $s \in S$, the map

$$\tilde{s}^i \mapsto T^i(\tilde{s}^i, s^{-i}) + \mathbf{E}_{a\sim\mu(\tilde{s}^i,s^{-i})}\left[\theta^i_{\vec{a},r}(s^i)\right]$$

has a *strict* maximum at $\tilde{s}^i = s^i$.

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

Formally, given the recommendation a_0 in the previous round, and reports m in the current round, the p.r.d. picks a recommended action profile $a_0' \in A$, which is equal to a_0 with probability $1-\eta_0$, and drawn according to $\mu(m)$ otherwise. We set $\rho_0(m, a_0) = a_0'$, $\phi_0(m, a_0, y) = a_0'$, and $x_0(m, a_0) = -\gamma_{\vec{a_0'}} + \eta_0 T(m)$.[46,47]

Thus, (ρ_0, K_0, ϕ_0) is irreducible. Denote by μ_{η_0} the invariant measure and by $\gamma_{\eta_0} \in \mathbf{R}^I$ and $\theta_{\eta_0} : S \times K_0 \to \mathbf{R}^I$ the long-run payoff and relative values respectively (including transfers).

[45] The only, quite minor modification is as follows. Elements of $\mathcal{S}$ are now triples $(\rho_{\text{ext}}, x, v)$, where $\rho_{\text{ext}} = (\rho, K, \phi)$ is an extended policy, and the auxiliary set K changes with ρ. At the beginning of the k-th block, once the extended policy $(\rho_{(k)}, K_{(k)}, \phi_{(k)})$ has been selected as a function of past public play, an initial state in $K_{(k)}$ still needs to be specified. This choice is irrelevant for the proofs.

[46] Consistent with our usage, the dependence of ρ_0, x_0 and ϕ_0 on the outcome of the p.r.d. does not appear explicitly.

[47] Recall that $\gamma_{\vec{a}}$ is the long-run payoff induced by the constant policy $\vec{a}$. That is, $\gamma_{\vec{a}} = \mathbf{E}_{s\sim\mu_{\vec{a}}}[r(s,a)] = r(\mu_{\vec{a}}, a)$.

Lemma 9 *Both* $\lim_{\eta_0 \to 0} \mu_{\eta_0}$ *and* $\lim_{\eta_0 \to 0} \theta_{\eta_0}$ *exist. In addition,* $\lim_{\eta_0 \to 0} \theta_{\eta_0}(s, k_0) - \theta_{\vec{k_0},r}(s)$ *only depends on* k_0*.*

Proof. The distribution μ_{η_0} is the unique solution to a linear system with coefficients affine in η_0. Therefore, $\eta_0 \mapsto \mu_{\eta_0}$ is a rational function and, being bounded, has a limit as $\eta_0 \to 0$. We refer to the online appendix for the proof relative to θ_{η_0}. The proof uses similar arguments, but the proof that $\eta_0 \mapsto \theta_{\eta_0}$ is bounded is more delicate. ■

Lemma 10 *For* η_0 *small enough, the pair* $(\rho_{\text{ext},0}, x_0)$ *is ex post strictly truthful.*

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

$$r^i(s^i, a') + x_0^i((\tilde{s}^i, s^{-i}), a') + \theta_{\eta_0}^i(t, a'),$$

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

Taking expectations over a' and t, and since $x_0(\tilde{s}^i, s^{-i}, a') = \eta_0 T(\tilde{s}^i, s^{-i}) - \gamma_{\vec{a'}}$, the expected payoff when reporting $\tilde{s}^i$ is

$$\begin{aligned} & (1 - \eta_0) \left\{ r^i(s^i, a) - \gamma_{\vec{a}} + \mathbf{E}_{t \sim p(\cdot \mid s, a)} \left[\theta_{\eta_0}^i(t, a) \right] \right\} \\ + \quad & \eta_0 \left\{ \mathbf{E}_{a' \sim \mu(\tilde{s}^i, s^{-i})} \left[r^i(s^i, a') + T^i(\tilde{s}^i, s^{-i}) - \gamma_{\vec{a'}} + \mathbf{E}_{t \sim p(\cdot \mid s, a')} \theta_{\eta_0}^i(t, a') \right] \right\}. \end{aligned}$$

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

$$r^i(s^i, a') + T^i(\tilde{s}^i, s^{-i}) - \gamma_{\vec{a'}} + \mathbf{E}_{t \sim p(\cdot \mid s, a')} \theta_{\vec{a'},r}^i(t^i),$$

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

B.2.3 Step 2: λ is not a coordinate direction

We here deal with the more difficult case where λ is not a coordinate direction. We rely on Proposition 5 below deals with the following setup. Let be given an irreducible MDP with state space Ω, action set B, transition function $q(\cdot \mid \omega, b)$, reward $r : \Omega \times B \to \mathbf{R}$ (all sets being finite). Assume that successive states are observed by a first agent, who makes a report

to a second one, who in turn chooses an action, the reward of both agents being r. Plainly, if the second agent follows a stationary optimal policy, it is weakly optimal for the first one to be truthful. According to Proposition 5, there are arbitrarily small report-contingent transfers and an optimal policy in the perturbed MDP, see **P1**, such that truth-telling is strictly optimal whenever the report affects the action (distribution) being played, see **P2**.

Proposition 5 *For each $\varepsilon > 0$, there exists $x : \Omega \times B \to \mathbf{R}$ and $\rho : \Omega \to \text{int}\Delta(B)$ such that the following holds, with $\theta := \theta_{\rho, r+x}$:*

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

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

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

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

Proposition 5 is immediate when transitions are action-independent. In that case indeed, and for $\omega \in \Omega$, set $B(\omega) := \text{argmax}_B r(\omega, \cdot)$, let $\rho(\omega)$ be the uniform distribution over $B(\omega)$ and set $x(\omega) := \eta|B(\omega)|$. For small $\eta > 0$, the pair (ρ, x) satisfies **P1** and **P2**. The proof is significantly more involved under action-dependent transitions. It is in the online appendix.

We now proceed in three (sub-)steps. We first rely on Proposition 5 to prove in Lemma 11 the existence of an extended policy $\rho_{\text{ext},1}$ and of transfers x_1 such that the long-run payoff under $\rho_{\text{ext},1}$ is close to $\bar{k}_1(\lambda)$ and such that truth-telling incentives are *ex post* strict unless reports do not affect the action being played. By perturbing the latter extended policy with the policy $\rho_{\text{ext},0}$ defined in Step 1, we next prove in Lemma 12 the existence of an *ex post* strictly truthful pair (ρ_{ext}, x) such that the long-run payoff under ρ_{ext} –excluding transfers– is close to $\bar{k}_1(\lambda)$. We conclude using AS and the action-identifiability assumption **A2**.

Lemma 11 *For all $\varepsilon > 0$, there exists an irreducible extended policy $\rho_{\text{ext},1} = (\rho_1, K_1, \phi_1)$ where $\rho_1 : S_+ \times K_1 \to \Delta(A)$ and transfers $x_1 : S_+ \times K_1 \to \mathbf{R}^I$, s.t.*

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

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

Proof. Proposition 5 holds for finite MDPs, hence we will have to rely on finite state approximations of the MDP $\mathcal{M}_\lambda$. We use the notations from Section B.1. Let $\varepsilon > 0$ be given. We let (K_1, ϕ_1) be an η-approximation of Δ_- such that the limit value v_ϕ of the MDP $\mathcal{M}_\phi$ induced by (K_1, ϕ_1) is close to $\bar{k}_1(\lambda)$: $|v_\phi - \bar{k}_1(\lambda)| < \frac{\varepsilon}{3}$. In addition, we assume that $\eta > 0$ is small enough[48] so that, for each irreducible $\rho : S_+ \times K_1 \to \Delta(A)$, one has

$$\left|\mathbf{E}_{(s_+,k,a)\sim\mu_\rho}\left[r_\lambda(s_+,k,a)\right] - \mathbf{E}_{(s,a)\sim\mu_\rho}\left[\lambda \cdot r(s,a)\right]\right| < \frac{\varepsilon}{3}, \tag{22}$$

where $\mu_\rho \in \Delta(S \times K_1 \times A)$ is the invariant distribution induced by ρ. Inequality (22) reads as follows: the two expectations are the long-run payoffs induced by ρ in the MDPs $\mathcal{M}_\phi$ and $\mathcal{M}_\lambda$ respectively. Consequently, the long-run λ-weighted payoff induced by any such policy ρ is close to the payoff induced in $\mathcal{M}_\phi$.

With this choice of (K_1, ϕ_1), we apply Proposition 5 to the MDP $\mathcal{M}_\phi$ with $\varepsilon/3$, and get $\rho_1 : S_+ \times K_1 \to \text{int } \Delta(A)$ and $\bar{x} : S_+ \times K_1 \times A \to \mathbf{R}$. Abusing notations, we will also view ρ_1 and $\bar{x}$ as maps defined on $S \times K_1$ and $S \times K_1 \times A$, independent of s^i for $i \in I_-$.

To repeat, the pair $(\rho_{\text{ext},1}, \bar{x})$ is such that for all $\omega, \tilde{\omega} \in S_+ \times K_1$, and denoting q the transition function in $\mathcal{M}_\phi$, one has

$$\begin{aligned} & r_\lambda(\omega, \rho_1(\omega)) + \bar{x}(\omega, \rho_1(\omega)) + \mathbf{E}_{q(\cdot|\omega,\rho_1(\omega))}\theta_{\rho_1, r_\lambda + \bar{x}}(\omega') \\ > \; & r_\lambda(\omega, \rho_1(\tilde{\omega})) + \bar{x}(\tilde{\omega}, \rho_1(\tilde{\omega})) + \mathbf{E}_{q(\cdot|\omega,\rho_1(\tilde{\omega}))}\theta_{\rho_1, r_\lambda + \bar{x}}(\omega') \end{aligned}$$

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

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

Together with (22), this proves **C1**.

Next, we follow Claim 9 and introduce transfers of the VCG type. For $i \in I_+$, we define $x_1^i : S_+ \times K_1 \to \mathbf{R}$ by

$$\lambda^i r^i(\omega, \rho_1(\omega)) + \lambda^i x_1^i(\omega) := r_\lambda(\omega, \rho_1(\omega)) + \bar{x}(\omega, \rho_1(\omega))$$

so that, as in Claim 9, one has $\lambda^i \theta^i_{\rho_1, r + x_1} = \theta_{\rho_1, r_\lambda + \bar{x}}$. Therefore, ρ_1 inherits the following truth-telling property in the one-shot game $\Gamma(\rho_{\text{ext},1}, x_1)$: at each state $(s_+, k) \in S_+ \times K_1$ and for each $i \in I_+$, player i strictly prefers reporting s^i over $\tilde{s}^i$ whenever $\rho_1(s_+, k) \neq \rho_1(\tilde{s}^i, s_+^{-i}, k)$.

For $i \in I_-$, set $x_1^i = 0$. Since ρ_1 is independent of $s^i \in S^i$, the latter property also holds for all $i \in I_-$. Thus, **C2** holds. ■

[48] It suffices to take $\eta < (1 - c)\varepsilon/3\bar{r}$, see Section B.1.

Lemma 12 *For all $\varepsilon > 0$, there exists an irreducible extended policy $\rho_{\text{ext}} = (\rho, K, \phi)$ such that*

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

C'2 *The pair (ρ_{ext}, x) is ex post strictly truthful.*

Proof. We construct (ρ, K, ϕ) as a perturbation of (ρ_1, K_1, ϕ_1) using (ρ_0, K_0, ϕ_0) so that the play alternates between long phases in $S \times K_1$ and long but much shorter, phases in $S \times K_0$.

Let $\eta_1 > 0$ be small, to be fixed later. We define the extended policy $\rho_{\text{ext}} = (\rho, K, \phi)$ as follows. We set $K = K_0 \cup K_1$. In each round, given the current public auxiliary state $k \in K$ and reports $s \in S$, the p.r.d. updates the public state to $k' \in K$ as follows. If $k \in K_1$, k' is set to k with probability $1 - \eta_1^2$, and $k' \sim \mu(s)$ otherwise. If $k \in K_0$, k' is set to a fixed $\bar{k}_1 \in K_1$ with probability η_1^2, and otherwise determined as under $\rho_{\text{ext},0}$ (*i.e.*, set to k with probability $1 - \eta_0$, and otherwise drawn according to $\mu(s)$).

We then define $\rho : S \times K \to \Delta(A)$ as $\rho(s,k) = \rho_0(s,k')$ if $k' \in K_0$ and $\rho(s,k) = \rho_1(s,k')$ if $k' \in K_1$. We also set $\phi(k,y) = \phi_1(k',y)$ if $k' \in K_1$ and $\phi(k,y) = k'$ if $k' \in K_0$.[49]

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

The irreducibility of (ρ, K, ϕ) follows from that of both (ρ_0, K_0, ϕ_0) and (ρ_1, K_1, ϕ_1). We denote by $\mu_{\eta_1} := \mu_{\rho_{\text{ext}}}$ the invariant measure as a function of η_1, and by $\theta_{\eta_1} := \theta_{\rho_{\text{ext}}, r+x} : S \times K \to \mathbf{R}^I$ the relative values. As in Section B.2.2, $\mu := \lim_{\eta_1 \to 0} \mu_{\eta_1}$ is well-defined. In addition, note that the limit transition function is the one induced by $\rho_{\text{ext},0}$ and $\rho_{\text{ext},1}$ on $S \times K_0$ and $S \times K_1$ respectively. Since transitions from K_0 to K_1 (resp., from K_1 to K_0) occur with probability η_1 (resp., η_1^2), one has $\mu_{\eta_1}(S \times K_1) = \dfrac{1}{1+\eta_1}$. As a consequence,

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

hence **C'1** hold for η_1 small enough.

We turn to **C'2**. As in Section B.2.2 (see supplementary material online), $\theta := \lim_{\eta_1 \to 0} \theta_{\eta_1}$ is also well-defined, and for $s \in S$, the differences $\theta(s,k_0) - \theta_0(s,k_0)$ and $\theta(s,k_1) - \theta_1(s,k_1)$ are independent of $k_0 \in K_0$ and $k_1 \in K_1$ respectively (where $\theta_n = \theta_{\rho_n, r+x_n}$ for $n = 0, 1$).

Fix $(s,k) \in S \times K$, $i \in I$ and $\tilde{s}^i \in S^i$. If $k \in K_0$ the strict incentive to report s^i follows from the strict truthfulness of $(\rho_{\text{ext},0}, x_0)$, for η_1 small enough.

[49] Note that ϕ therefore also depends here on the reports of the players.

Assume now that $k \in K_1$. If $\rho(s^i, s^{-i}, k) \neq \rho(\tilde{s}^i, s^{-i}, k)$ for *some* $s^{-i} \in S^{-i}$, player i strictly prefers reporting s^i over $\tilde{s}^i$ in $\Gamma(\rho_{\text{ext},1}, x_1)$.[50] And therefore in $\Gamma(\rho_{\text{ext}}, x)$ as well, for η_1 small enough. Assume finally that $\rho(s^i, s^{-i}, k) = \rho(\tilde{s}^i, s^{-i}, k)$ for *all* $s^{-i} \in S^{-i}$. Then the expected payoff of player i in $\Gamma(\rho_{\text{ext}}, x)$, conditional on the p.r.d. picking $k' = k$ is the same under both reports s^i and $\tilde{s}^i$. On the other hand, conditional on the p.r.d. picking some $a' \in K_0$, the expected payoff of player i converges as $\eta_1 \to 0$ to

$$\mathbf{E}_{a' \sim \mu(\cdot, s^{-i})} \left[r^i(s^i, a') + T^i(\cdot, s^{-i}) - \gamma_{\vec{a'}} + \mathbf{E}_{t \sim p(\cdot|s,a')} \theta_0^i(t, a') \right] = \mathbf{E}_{a' \sum \mu(\cdot, s^{-i})} \left[\theta_0^i(s, a') + T^i(\cdot, s^{-i}) \right]$$

which, by the choice of η_0, has a strict maximum for s^i. The strict truthfulness of (ρ_{ext}, x) follows, provided η_1 is small enough. ■

We now conclude the proof of Lemma 7. Inspection of the proofs of Claims 8, 9 and 10 shows that the successive modifications of the transfers preserve strict inequalities and do not rely on transitions being action-independent. That is, the same sequence of claims leads here to the existence of $x_4 : \Omega_{\text{pub}} \times S \times K \to \mathbf{R}^I$ (with $\Omega_{\text{pub}} = S \times K \times Y$) such that $\lambda \cdot x_4(\cdot) = 0$ and all truth-telling incentives in $\Gamma(\rho_{\text{ext}}, x_4)$ are strict.

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

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

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

B.2.4 Step 3: λ is a coordinate direction

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

[50] Whatever the choice of $\bar{\omega}_{\text{pub}}$ by nature in $\Gamma(\rho_{\text{ext},1}, x_1)$. Indeed, player i is *ex post* indifferent between s^i and $\tilde{s}^i$ when $\rho(s) = \rho(\tilde{s}^i, s^{-i})$ and strictly prefers s^i over $\tilde{s}^i$ if $\rho(s) = \rho(\tilde{s}^i, s^{-i})$. The claim thus follows since the belief of player i over S^{-i} has full support.

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

We view $\rho_1 : S \to \Delta(A)$ as a map defined over S (independent of s^{-i}). By **A2**, for $j \neq i$ and for each $a \in A$, there exists $x_a : Y \to \mathbf{R}^I$ that induce strict obedience to a:

$$r^j(s^j, \tilde{a}^j, a^{-j}) + \mathbf{E}_{p(\cdot|\tilde{a}^j, a^{-j})} x_a^j(y) + \mathbf{E}_{p(\cdot|s^j, \tilde{a}^j, a^{-j})} \theta^j_{\rho_1, r+\bar{x}}(t) < r^j(s^j, a) + \mathbf{E}_{p(\cdot|a)} x_a^j(y) + \mathbf{E}_{p(\cdot|s^j, a)} \theta^j_{\rho_1, r+\bar{x}}(t) \tag{23}$$

for each $s \in S$ and $\tilde{a}^j \neq a^j$.

For $j = i$, we ask for more. For any $s^i \in S^i$ and since ρ_1 is optimal in the MDP with payoff $r^i + \bar{x}$, any action $a \in A$ in the support of $\rho_1(s^i)$ maximizes

$$r^i(s^i, a) + \bar{x}(s^i, a) + \mathbf{E}_{t^i \sim p(\cdot|s^i, a)} \theta^i_{\rho_1, r+\bar{x}}(t^i).$$

Since $\|\bar{x}\| < \varepsilon$, the components x_a^i can be chosen so that the following holds:

B1 : (23) is modified and strengthened to

$$r^i(s^i, \tilde{a}^i, a^{-i}) + \bar{x}(\tilde{s}^i, a) + \mathbf{E}_{y \sim p(\cdot|\tilde{a}^i, a^{-i})} x_a^i(y) + \mathbf{E}_{t \sim p(\cdot|s, \tilde{a}^i, a^{-i})} \theta^i_{\rho_1, r+\bar{x}}(t)$$

$$< r^i(s^i, a) + \bar{x}(s^i, a) + \mathbf{E}_{y \sim p(\cdot|a)} x_a^i(y) + \mathbf{E}_{t^i \sim p(\cdot|s^j, a)} \theta^i_{\rho_1, r+\bar{x}}(t^i),$$

for every $\tilde{s}^i \in S^i$ and $\tilde{a}^i \neq a^i$.

B2 $\|x_a^i\| < k\varepsilon$ for some constant k that only depends on the primitives of the model and not on ε.

B3 $x_a^i(\cdot) \leq 0$ and $\mathbf{E}_{y \sim p(\cdot|a)} x_a^i(y)$ is independent of $a \in A$.

The substantive properties are **B1** and **B2**. Once they hold, **B3** follows by subtracting a small constant from x_a^i.

We view $\rho_1 : S \to \Delta(A)$ as being implemented by means of the p.r.d. picking a recommendation $a \sim \rho_1(s)$, and transfers being then given by $x_1^i(s, y) := \bar{x}(s^i, a) + x_a^i(y)$ and $x_1^j(s, y) = x_a^j(y)$ for $j \neq i$. The properties of x_a and of $\bar{x}$ ensure that the pair (ρ_1, x_1) is strictly obedient and satisfy the same truth-telling incentives as the pair $(\rho_1, \bar{x})$. Observe that $x_1^i(s, y) \leq (k+1)\varepsilon$. Since ρ_1 is optimal in the MDP with reward $r + \bar{x}$, one has

$$\mathbf{E}_{\mu_{\rho_{\text{ext}}, 1}} \left[r^i(s^i, a) + x_1^i(s, y) \right] \geq \bar{k}_1(e^i) - (k+1)\varepsilon.$$

We now recall the strictly truthful pair $(\rho_{\text{ext},0}, x_0)$ from Step 1 in Section B.2.2. Once again, we supplement x_0 with transfers inducing obedience. For $a \in A$, we let $x_a : Y \to \mathbf{R}^I$ be such that (i) $\mathbf{E}_{p(\cdot|a)} x_a(y) = 0$, and (ii) $\mathbf{E}_{p(\cdot|\tilde{a}^j, a^{-j})} x_a^j(y)$ is a large negative constant for each $j \in J$ and $\tilde{a}^j \neq a^j$. We next subtract the same constant to all maps $x_a^i(\cdot)$ $(a \in A)$ to get $x_a^i(\cdot) \leq 0$. (With an abuse of notation), transfers $x_0 = S \times K_0 \times Y \to \mathbf{R}^I$ are now defined by $x_0(s, k_0, y) = x_0(s, k_0) + x_a(y)$ where $a \in A$ is selected by the p.r.d. as specified in (K_0, ϕ_0).[51] With this updated definition of x_0, the pair $(\rho_{\text{ext},0}, x_0)$ is both strictly truthful and strictly obedient.

We now perturb. For $\eta_1 > 0$, we define the irreducible extended policy $\rho_{\text{ext}} = (\rho, K, \phi)$ from ρ_1 and $\rho_{\text{ext},0}$ and transfers $x : S \times K \times Y \to \mathbf{R}^I$ from x_0 and x_1, exactly as ρ_{ext} and x were obtained in Step 2 from $\rho_{\text{ext},1}$ and $\rho_{\text{ext},0}$. As in Step 2, it follows that for $\eta_1 > 0$ small, the pair (ρ_{ext}, x) is both strictly truthful and obedient –hence the triple $(\mathbf{E}_{\mu_{\rho_{\text{ext}}}} [r(s,a) + x(s,k,y)], \rho_{\text{ext}}, x)$ is feasible in $\mathcal{P}(e^i)$. Finally, since transitions from ρ_1 to $\rho_{\text{ext},0}$ (resp., from $\rho_{\text{ext},0}$ to ρ_1) occur with probability η_1^2 (resp., η_1) in each round, the expectation $\mathbf{E}_{\mu_{\rho_{\text{ext}}}} \left[r^i(s,a) + x^i(s,k,y) \right]$ is arbitrarily close to $\mathbf{E}_{\mu_{\rho_{\text{ext},1}}} \left[r^i(s,a) + x_1^i(s,a) \right]$ for $\eta_1 > 0$ small enough. The result follows.

The case $\lambda = -e^i$ is analogous. Let $\bar{a}^{-i} \in A^{-i}$ achieve the min in the definition of $\underline{v}^i$. Let next $\tilde{\mathcal{M}}^i$ be the MDP faced by player i when maximizing his own payoffs against the constant policy $\bar{a}^{-i}$. Hence $\tilde{\mathcal{M}}^i$ has S^i as state space, A^i as action set, and the rewards and transitions in $\tilde{\mathcal{M}}^i$ are deduced from r^i and p given $\bar{a}^{-i}$. We then repeat the proof of the case $\lambda = +e^i$.[52]

C Proof of Theorem 5

The overall pattern of the proof is that of the proofs of Theorems 2 and 4. We let a compact set Z be given, included in the interior of W. Given $z \in Z$, we construct a sequential equilibrium σ with payoff z. Under σ, the play is divided in a sequence of phases. In each phase, a direction $\lambda \in \Lambda$ is selected as a function of public past play. If λ is not close to some negative coordinate direction $-e^i$, the players follow as before some extended policy

[51] Recall from Section B.2.2 that the p.r.d. sets $a \in A$ to be equal to k_0 with probability $1 - \eta_0$ and otherwise draws $a \sim \mu(s)$.

[52] With obvious changes. Transfers x^i to player i are now required to be non-negative.

and the phase is of random duration. If instead λ is close to $-e^i$ for some $i \in I$, players play an equilibrium in an auxiliary "zero-sum" game (between i and $-i$) with fixed duration $T := \frac{1}{\sqrt{1-\delta}}$, final transfers and no communication (*i.e.*, reports are babbling).[53] These "zero-sum" games are defined and studied in Section F.1 below.

One new issue however arises. The equilibrium behavior in such a punishment phase depends on the continuation relative values at the end of the phase, which are themselves defined recursively from past public play –raising a potential circularity issue. To deal with it, we will insert a shorter transition phase at the end of each punishment phase, so as to ensure that the continuation values following punishment phases are predetermined. Given this change, we find it conceptually and technically more straightforward to insert such a transition phase between any two phases. Modulo this change, the proof will follow along earlier lines.

D Proof of Theorem 6

In Sections A.1 and B.2, the independence and private values assumptions are only used to obtain triples (v, ρ, x) and $(v, \rho_{\text{ext}}, x)$ with strict truth-telling incentives, see Lemmas 3 and 7. Since all truth-telling incentives are required to be strict in the optimization program $\mathcal{P}_2$, the analog of the latter two lemmas readily holds here, and the result follows as in Section A.1.

[53] The exponent $-\frac{1}{2}$ is somewhat arbitrary. What matters is that $T \ll \frac{1}{1-\delta}$ so that $\delta^T \to 1$ as $\delta \to 1$.

Supplementary Material: Truthful Equilibria in Dynamic Bayesian Games,

Johannes Hörner, Satoru Takahashi, Nicolas Vieille

This supplement contains additional material on Markov Decision Problems and details on the proof of Theorem 5.

E Markov Decision Problems

E.1 The ACOE

For the reader's convenience, we provide a statement and a self-contained proof of the Average Cost Optimality Equation for MDPs. The material in this section is standard.

We let an irreducible MDP $\mathcal{M}$ with finite primitives be given. The state space is S, the action set is A, the reward function is $r : S \times A \to \mathbf{R}$, and the transition function is $p(\cdot \mid s, a)$.[54] We let Σ denote the set of strategies in $\mathcal{M}$.

For $\delta < 1$ and $N \in \mathbf{N}$, we let

$$v_\delta(s) := \max_{\sigma \in \Sigma} \mathbf{E}_{s,\sigma} \left[(1-\delta) \sum_{n=1}^{\infty} \delta^{n-1} r(s_n, a_n) \right]$$

and

$$v_N(s) := \max_{\sigma \in \Sigma} \mathbf{E}_{s,\sigma} \left[\frac{1}{N} \sum_{n=1}^{N} r(s_n, a_n) \right]$$

denote the values of the discounted and finite horizon versions of $\mathcal{M}$, as a function of the initial state s.

Proposition 6 (ACOE) *There is a unique $v \in \mathbf{R}$ and a unique (up to an additive constant) map $\theta : S \to \mathbf{R}$ such that*

$$v + \theta(s) = \max_{a \in A} \left\{ r(s,a) + \mathbf{E}_{p(\cdot|s,a)} \theta(\cdot) \right\}, \text{ for all } s \in S. \tag{24}$$

In addition, $v = \lim_{\delta \to 1} v_\delta(s) = \lim_{N \to +\infty} v_N(s)$ for all $s \in S$.

[54] We are thus assuming that the sets S and A are finite and that for each policy $\rho : S \to \Delta(A)$, the induced Markov chain (s_n) is irreducible.

Proof. We first prove the existence of a solution to (24). For $\delta < 1$ the dynamic programming principle writes

$$v_\delta(s) = \max_{a \in A} \left\{ (1-\delta) r(s,a) + \delta \mathbf{E}_{p(\cdot|s,a)} v_\delta(\cdot) \right\}, \text{ for all } s \in S. \tag{25}$$

Let $a^*(s)$ achieves the maximum in (25), so that $v_\delta(s) = (1-\delta) r(s, a^*(s)) + \delta \mathbf{E}_{p(\cdot|s,a^*_s)} v_\delta(\cdot)$ for each s. This implies that $\delta \mapsto v_\delta(s)$ is a bounded and rational function on $[0,1)$. In particular, both $v(s) := \lim_{\delta \to 1} v_\delta(s)$ and $\theta(s) := \lim_{\delta \to 1} \dfrac{v_\delta(s) - v(s)}{1-\delta}$ exist. Irreducibility readily implies that $v(s)$ is independent of s.

Equation (25) then rewrites as

$$v + (v_\delta(s) - v) = \max_{a \in A} \left\{ (1-\delta) r(s,a) + \delta \mathbf{E}_{p(\cdot|s,a)} \left[v_\delta(t) - v \right] + \delta v \right\}.$$

Equation (24) follows when dividing by $1-\delta$ and letting $\delta \to 1$.

We next prove uniqueness, and start with v. Let (v, θ) be a solution to (24), so that

$$\theta(s) = \max_{a \in A} \left\{ r(s,a) + \mathbf{E}_{p(\cdot|s,a)} \theta(\cdot) \right\} - v. \tag{26}$$

Substituting (26) into the right-hand side of (24) yields first

$$2v + \theta(s) = \max_{\sigma} \mathbf{E}_{s,\sigma} \left[r(s_1, a_1) + r(s_2, a_2) + \theta(s_3) \right],$$

and, by induction,

$$v + \frac{\theta(s)}{N} = \max_{\sigma} \mathbf{E}_{s,\sigma} \left[\frac{1}{N} \sum_{n=1}^{N} r(s_n, a_n) + \frac{\theta(s_{N+1})}{N} \right]$$

for each N. This implies that $\lim_{N \to \infty} v_N(s)$ exists and is equal to v.

We conclude with the uniqueness of θ. Let (v, θ) and (v, ψ) be two solutions to (24). This implies

$$\theta(s) - \psi(s) \leq \max_{a \in A} \mathbf{E}_{p(\cdot|s,a)} \left(\theta(\cdot) - \psi(\cdot) \right)$$

for each s. By irreducibility, it follows that $\theta(\cdot) - \psi(\cdot)$ is constant. ■

E.2 Perturbed Markov Chains and Relative Values

We discuss here two statements on the asymptotic properties of relative values of perturbed Markov chains, as the perturbation parameter converges to zero. These statements readily imply those used in the main body of the paper.

E.2.1 Result 1

The setup is as follows. Let be given (disjoint) sets S_l, with $1 \leq l \leq L$. Let also be given, for each l, an irreducible transition function p_l on S_l, with invariant measure ν_l, and a "payoff" $r_l : S_l \to \mathbf{R}$ with $\mathbf{E}_{\nu_l}[r_l(s)] = 0$. Let $\theta_l : S_l \to \mathbf{R}$ denote the associated relative value.

In addition, let p be an irreducible transition function on $\mathcal{S} := S_1 \cup \cdots \cup S_L$, and let $r : \mathcal{S} \to \mathbf{R}$ be the function that coincides with r_l on S_l. For $\varepsilon > 0$, we define a transition function p_ε on $\mathcal{S}$ as $p_\varepsilon(t \mid s) := (1-\varepsilon)p_l(t \mid s) + \varepsilon p(t \mid s)$ for $s \in S_l$ and $t \in \mathcal{S}$. Let $\mu_\varepsilon \in \Delta(\mathcal{S})$ be the invariant measure of p_ε, $\gamma_\varepsilon := \mathbf{E}_{\mu_\varepsilon}[r(s)]$ the long-run payoff, and $\theta_\varepsilon : \mathcal{S} \to \mathbf{R}$ the relative value. To fix ideas, we normalize θ_ε by imposing the condition $\mathbf{E}_{\mu_\varepsilon}[\theta_\varepsilon(\cdot)] = 0$.

Proposition 7 *The map $\varepsilon \mapsto \theta_\varepsilon$ is bounded. In addition,*

$$\lim_{\varepsilon \to 0} (\theta_\varepsilon(s') - \theta_\varepsilon(s)) = \theta_l(s') - \theta_l(s) \text{ for every } s, s' \in S_l.$$

Proof. We view each transition $p_\varepsilon(\cdot \mid s)$ as the succession of two random choices. First, it is randomly decided, with probability ε, whether to use p or p_l to draw the next state, which is next drawn accordingly. We denote by τ the random time of *first* "switch" (first round where p is used).

Given any two states $s, s' \in \mathcal{S}$ we denote by (s_n) and (s'_n) two Markov chains with transition functions p_ε starting from s and s' respectively, which are coupled in that (i) the successive switches occur in the same rounds for the two chains, and (ii) $s_n = s'_n$ after the first coincidence time $\omega := \inf\{n : s_n = s'_n\}$; yet all other random choices are independent.

Claim 13 *The following holds:*

- *There exists $c_1 > 0$ such that $\mathbf{E}\left[\sum_{n=1}^{\tau-1} (r(s_n) - r(s'_n))\right] \leq c_1$ for all $s, s' \in \mathcal{S}$ and $\varepsilon > 0$.*
- *There exists $c_2 > 0$ such that $\mathbf{P}(\omega \leq \tau) \geq c_2$ for every l, $s, s' \in S_l$ and $0 < \varepsilon \leq \frac{1}{2}$.*

Proof of the claim. Let $s \in \mathcal{S}$ be given, say $s \in S_l$. One has, with obvious notations

$$\mathbf{E}_\varepsilon \left(\sum_{n=1}^{\tau-1} r(s_n) \right) = \mathbf{E}_l \left(\sum_{n=1}^{\tau-1} r(s_n) \right).$$

By the ACOE, the latter is equal to $\theta_l(s) - \mathbf{E}_l[\theta_l(s_\tau)]$ and is therefore bounded as a function of ε.

The second statement follows from the irreducibility of p_l.[55] ■

Next, we denote by (τ_k) the successive switches – so that $\tau_1 = \tau$. Given $s, s' \in \mathcal{S}$, denote by ϕ the smallest index k such that s_{τ_k+1} and s'_{τ_k+1} belong to the same component S_l. Because p is irreducible, there exists $c_3 > 0$ such that $\mathbf{P}(\phi \leq L) \geq c_3$. Note that

$$\theta_\varepsilon(s) = \mathbf{E}\left[\sum_{n=1}^{\tau_{L+1}} (r(s_n) - \gamma_\varepsilon) + \theta_\varepsilon(s_{\tau_{L+1}+1})\right]$$

and a similar equality holds for $\theta_\varepsilon(s')$, hence

$$\begin{aligned}
\theta_\varepsilon(s') - \theta_\varepsilon(s) &= \mathbf{E}\left[\sum_{n=1}^{\tau_{L+1}} (r(s_n) - r(s'_n))\right] + \mathbf{E}\left[\theta_\varepsilon(s'_{\tau_{L+1}+1}) - \theta_\varepsilon(s_{\tau_{L+1}+1})\right] \\
&\leq Lc_1 + \max_{t,t' \in \mathcal{S}} (\theta_\varepsilon(t') - \theta_\varepsilon(t)) \times \mathbf{P}(\omega > \tau_{L+1}) \\
&\leq Lc_1 + (1 - c_2c_3) \max_{t,t' \in \mathcal{S}} (\theta_\varepsilon(t') - \theta_\varepsilon(t)),
\end{aligned}$$

using the previous claim. It follows that $\max_{s,s' \in \mathcal{S}} |\theta_\varepsilon(s') - \theta_\varepsilon(s)| \leq \dfrac{Lc_1}{c_2c_3}$. Together with the equality $\mathbf{E}_{\mu_\varepsilon}\theta_\varepsilon(\cdot) = 0$, this implies the first statement.

For $\varepsilon > 0$, θ_ε is the unique solution to the linear system ($s \in S_l$, $l \leq L$)

$$\gamma_\varepsilon + \theta_\varepsilon(s) = r(s) + (1 - \varepsilon)\mathbf{E}_{p_l(\cdot|s)}\theta_\varepsilon(t) + \varepsilon\mathbf{E}_{p(\cdot|s)}\theta_\varepsilon(t),$$

together with the normalization equation.[56] Therefore, $\theta_\varepsilon(s)$ is a bounded and rational function of s. Thus, $\theta(s) := \lim_{\varepsilon \to 0} \theta_\varepsilon(s)$ exists and satisfies the limit system obtained when setting $\varepsilon = 0$. That is, for fixed l and for each $s \in S_l$, one has

$$\theta(s) = r(s) + \mathbf{E}_{p_l(\cdot|s)}\theta(t).$$

All solutions of the latter system are equal to θ_l up to an additive constant, hence the result.

■

E.2.2 Result 2

The setup here is a variant of the previous one. We let be given two (disjoint) sets S_1 and S_2, an irreducible transition function p_l on S_l with invariant measure ν_l, a function $r_l : S_l \to \mathbf{R}$

[55] $\mathbf{P}(\omega \leq \tau)$ is continuous as a function of ε, converges to 1 as $\varepsilon \to 0$, and is less than one, except for $\varepsilon = 1$.

[56] Since μ_ε is the unique solution of a linear system with coefficients linear in ε, $\varepsilon \mapsto \mu_\varepsilon$ is a rational function, hence $\varepsilon \mapsto \gamma_\varepsilon$ is a rational function as well.

$(l = 1, 2)$ and θ_l the relative value. In addition, let $f : \mathcal{S} := S_1 \cup S_2 \to \mathcal{S}$ be such that $f(S_1) \subseteq S_2$ and $f(S_2) \subseteq S_1$ and let $r : \mathcal{S} \to \mathbf{R}$ be the map whose restriction to S_l is r_l.

For $\varepsilon = (\varepsilon_1, \varepsilon_2) \in (0, 1)^2$, we define a transition function p_ε over $\mathcal{S}$ by $p_\varepsilon(t \mid s) = (1 - \varepsilon_l)p_l(t \mid s) + \varepsilon_l f(s)$ for $s \in S_l$. Thus, transitions from S_1 to S_2 (resp., from S_2 to S_1) occur with probability ε_1 (resp., ε_2) in each round. Let $\theta_\varepsilon : \mathcal{S} \to \mathbf{R}$ denote the relative value.

Proposition 8 *One has* $\lim_{\varepsilon \to 0} (\theta_\varepsilon(s') - \theta_\varepsilon(s)) = \theta_l(s') - \theta_l(s)$ *whenever* $s, s' \in S_l$.

Note however that θ_ε is unbounded as a function of ε as soon as $\mathbf{E}_{\nu_1} r_1(\cdot) \neq \mathbf{E}_{\nu_2} r_2(\cdot)$.

Proof. We first prove that $\varepsilon \mapsto \theta_\varepsilon(s') - \theta_\varepsilon(s)$ is bounded whenever $s, s' \in S_l$. We use the same notations as in the proof of Proposition 7, and let (s_n) and (s'_n) be two Markov chains starting from s and s', with t.f. p_ε, and coupled as before. The constants c_1 and c_2 are as before. Whenever $s, s' \in S_l$ (and for ε_l bounded away from one), one has $\mathbf{P}(\omega \leq \tau) \geq c_2$, hence

$$|\theta_\varepsilon(s') - \theta_\varepsilon(s)| \leq c_1 + (1 - c_2) \max_{t, t' \in S_{3-l}} |\theta_\varepsilon(t') - \theta_\varepsilon(t)| .$$

It follows that $\max_{l=1,2} \max_{s,s' \in S_l} |\theta_\varepsilon(s') - \theta_\varepsilon(s)| \leq \dfrac{c_1}{c_2}$.

The limit claim follows as in the proof of Proposition 7. ■

E.3 Proof of Proposition 5

We let an irreducible MDP $\mathcal{M}_0$ be given, with primitives (Ω, B, q, r). We denote by $v \in \mathbf{R}$ and $\theta : \Omega \to \mathbf{R}$ the limit value and relative values of $\mathcal{M}_0$. For $\omega \in \Omega$, we let

$$B_0(\omega) := \operatorname{argmax}_{b \in B} \left\{ r(\omega, b) + \mathbf{E}_{\omega' \sim q(\cdot \mid \omega, b)} \theta(\omega') \right\}$$

be the set of actions that are optimal at $\omega \in \Omega$.

Thus, for $\omega \in \Omega$ and $b \notin B_0(\omega)$, one has $r(\omega, b) + \mathbf{E}_{q(\cdot \mid \omega, b)} \theta(\omega') < v + \theta(\omega)$, and we let $c_0 > 0$ be strictly smaller than the difference between the two sides, for each ω and $b \notin B_0(\omega)$.

In the absence of transfers, assume that the second agent uses a distribution $\rho(\omega) \in \Delta(B_0(\omega))$ with full support, as a function of the report ω of the first agent. At state ω, it is strictly better to report truthfully ω rather than $\tilde{\omega}$ unless $B(\tilde{\omega}) \subseteq B(\omega)$. The main issue below will be to get rid of such indifference cases, and to prevent the first agent from reporting a state $\tilde{\omega}$ with $B(\tilde{\omega}) \subset B(\omega)$. The basic insight in the proof is to reward the first agent for reporting a state with many optimal actions.

We will construct a finite sequence $\mathcal{M}_1, \ldots, \mathcal{M}_n$ of perturbed MDPs. For all MDPs in the sequence, the state space is Ω and the action set is B.

We explain the construction of $\mathcal{M}_1$ before proceeding to the general case. Throughout, we fix an increasing function $\phi : \{1, \dots, |B|\} \to \mathbf{R}$ such that $\phi(|B|) < \frac{1}{|B|}$ (so that $\phi(m) < \frac{1}{m}$ for $m \leq |B|$). We then pick $\alpha > 0$ such that (i) $\alpha < \dfrac{1}{|B|} - \phi(|B|)$ and (ii) $\alpha < \phi(m+1) - \phi(m)$ for all $1 \leq m < |B|$.

Given $\varepsilon_1 > 0$, the reward r_1 and transition function q_1 of $\mathcal{M}_1$ are defined as

$$r_1(\omega, b) := (1 - \varepsilon_1) r(\omega, b) + \varepsilon_1 \left(r(\omega, \beta_0(\omega)) + c_0 \phi(|B_0(\omega)|) \right),$$

and

$$q_1(\cdot \mid \omega, b) := (1 - \varepsilon_1) q(\cdot \mid \omega, b) + \varepsilon_1 q(\cdot \mid \omega, \beta_0(\omega)),$$

where $\beta_0(\omega)$ is the uniform distribution over $B_0(\omega)$. We denote by v_{ε_1} and θ_{ε_1} the limit value and relative values of $\mathcal{M}_1$, and we let

$$B_1(\omega) := \operatorname{argmax}_{b \in B} \left\{ r_1(\omega, b) + \mathbf{E}_{\omega' \sim q_1(\cdot \mid \omega, b)} \theta(\omega') \right\}$$

be the set of optimal actions at ω in $\mathcal{M}_1$. Both v_{ε_1} and θ_{ε_1} are continuous w.r.t. ε_1, with $\lim_{\varepsilon_1 \to 0} v_{\varepsilon_1} = v$ and $\lim_{\varepsilon_1 \to 0} \theta_{\varepsilon_1} = \theta$. Hence $B_1(\omega)$ is upper hemi-continuous as a function of ε_1, so that $B_1(\omega) \subseteq B_0(\omega)$ for all $\varepsilon_1 > 0$ small enough, and $\omega \in \Omega$. We stop with the MDP $\mathcal{M}_1$ if there is a sequence $\varepsilon_1 \to 0$ such that $B_1(\cdot) = B_0(\cdot)$ along the sequence. We otherwise repeat the perturbation process with $\mathcal{M}_1$.

More generally, let $(\varepsilon_k)_{k \in \mathbf{N}}$ be a sequence of positive real numbers with $\sum_k \varepsilon_k < 1$. For $k \in \mathbf{N}$, we set $\vec{\varepsilon}_k := (\varepsilon_1, \dots, \varepsilon_k)$. For any such sequence (ε_k), we define inductively a sequence $\mathcal{M}_k(\vec{\varepsilon}_k)$ of MDPs with state space Ω and action set B, and with limit value denoted $v_{\vec{\varepsilon}_k}$ and $\theta_{\vec{\varepsilon}_k}$. The reward r_k and transition function q_k of $\mathcal{M}_k(\vec{\varepsilon}_k)$ are defined as

$$r_k(\omega, b) := \left(1 - \sum_{i=1}^{k} \varepsilon_i \right) r(\omega, b) + \sum_{i=1}^{k} \varepsilon_i \left(r(\omega, \beta_{i-1}(\omega)) + c_{i-1} \phi(|B_{i-1}(\omega)|) \right),$$

and

$$q_k(\cdot \mid \omega, b) = \left(1 - \sum_{i=1}^{k} \varepsilon_i \right) q(\cdot \mid \omega, b) + \sum_{i=1}^{k} \varepsilon_i q(\cdot \mid \omega, \beta_{i-1}(\omega)),$$

where $\beta_i(\omega)$, $B_i(\omega)$ and c_i are defined inductively as follows.

For each i,

$$B_i(\omega) := \operatorname{argmax}_b \{ r_i(\omega, b) + \mathbf{E}_{\omega' \sim q_i(\cdot \mid \omega, b)} \theta_{\vec{\varepsilon}_i}(\omega') \}$$

is the set of actions optimal at ω in $\mathcal{M}_i(\vec{\varepsilon}_i)$, $\beta_i(\omega) \in \Delta(B)$ is the uniform distribution over $B_i(\omega)$ and $c_i > 0$ is any number such that

$$c_i + r_i(\omega, b) + \mathbf{E}_{\omega' \sim q_i(\cdot|\omega,b)} \theta_{\vec{\varepsilon}_i}(\omega') < v_{\vec{\varepsilon}_i} + \theta_{\vec{\varepsilon}_i}(\omega)$$

for each $\omega \in \Omega$ and $b \notin B_i(\omega)$. This definition entails no circularity. Indeed, B_0, β_0 and c_0 are associated with $\mathcal{M}_0$ and, for $k \geq 1$, the definition of r_k and q_k, and therefore of $v_{\vec{\varepsilon}_k}$, $\theta_{\vec{\varepsilon}_k}$, B_k, β_k and c_k, only involves $v_{\vec{\varepsilon}_i}$ and $\theta_{\vec{\varepsilon}_i}$ for $i < k$.

Note also that, for given $\vec{\varepsilon}_{k-1}$, $v_{\vec{\varepsilon}_k}$ and $\theta_{\vec{\varepsilon}_k}$ are continuous as functions of ε_k, and $B_k(\omega)$ is therefore upper hemi continuous. It follows that, for every $\vec{\varepsilon}_{k-1}$, one has $B_k(\omega) \subseteq B_{k-1}(\omega)$ provided $\varepsilon_k > 0$ is small enough. In addition, $\lim_{\varepsilon_k \to 0} v_{\vec{\varepsilon}_k} = v_{\vec{\varepsilon}_{k-1}}$ and $\lim_{\varepsilon_k \to 0} \theta_{\vec{\varepsilon}_k} = \theta_{\vec{\varepsilon}_{k-1}}$.

In the sequel, we let a sequence (ε_k) be given such that for each k, ε_k is "very close to zero" given $\vec{\varepsilon}_{k-1}$. By this, we will mean that (i) $|v_{\vec{\varepsilon}_k} - v_{\vec{\varepsilon}_{k-1}}|$ and $\|\theta_{\vec{\varepsilon}_k} - \theta_{\vec{\varepsilon}_{k-1}}\|$ are smaller than some positive numbers which only involve $\vec{\varepsilon}_{k-1}$ (and which will appear in the computations below), and (ii) $B_k(\omega) \subset B_{k-1}(\omega)$ for every $\omega \in \Omega$.

We let $n \in \mathbf{N}$ be such that $B_n(\cdot) = B_{n-1}(\cdot)$, and we define $\rho : \Omega \to \Delta(B)$ as

$$\rho(\omega) := \left(1 - \sum_{k=1}^{n} \varepsilon_k\right) \beta_n(\omega) + \sum_{k=1}^{n} \varepsilon_k \beta_{k-1}(\omega).$$

Observe that $\operatorname{supp} \rho(\omega) = B_0(\omega)$ for each ω. We next define $x_{\text{eq}} : \Omega \times B \to \mathbf{R}$ as follows:

- for $b \in B_0(\omega)$, $x_{\text{eq}}(\omega, b)$ is defined by the equation

  $$x_{\text{eq}}(\omega, b) + r(\omega, b) + \mathbf{E}_{\omega' \sim q(\cdot|\omega,b)} \theta_{\vec{\varepsilon}_n}(\omega') = r(\omega, \rho(\omega) + \mathbf{E}_{\omega' \sim q(\cdot\omega,\rho(\omega))} \theta_{\vec{\varepsilon}_n}(\omega').$$

  Observe that $x_{\text{eq}}(\omega, \rho(\omega)) = \mathbf{E}_{b \sim \rho(\omega)} x_{\text{eq}}(\omega, b) = 0$ for each ω.

- For $b \notin B_0(\omega)$, we set $x_{\text{eq}}(\omega, b) = x_{\text{eq}}(\omega, \bar{b})$, where $\bar{b} \in B_n(\omega)$. Note that $x_{\text{eq}}(\omega, b)$ is independent of the choice of $\bar{b}$. Indeed, the actions of $B_n(\omega)$ are those that maximize $r_n(\omega, \cdot) + \mathbf{E}_{q_n(\cdot|\omega,\cdot)} \theta_{\vec{\varepsilon}_n}(\omega')$, or equivalently, that maximize $r(\omega, \cdot) + \mathbf{E}_{q(\cdot|\omega,\cdot)} \theta_{\vec{\varepsilon}_n}(\omega')$.

Finally, we define $x : \Omega \times B \to \mathbf{R}$ as

$$x(\omega, b) := x_{\text{eq}}(\omega, b) + \sum_{k=1}^{n} \varepsilon_k c_{k-1} \phi(|B_{k-1}(\omega)|).$$

We now prove that the pair (ρ, x) satisfies the desired properties.

Claim 14 *ρ is an optimal policy in the MDP with stage payoff $r(\omega, b) + x(\omega, b)$.*

Proof. Recall that $v_{\vec{\varepsilon}_n}$ and $\theta_{\vec{\varepsilon}_n}$ are the limit value and relative values of the MDP $\mathcal{M}_n(\vec{\varepsilon}_n)$, and that $B_n(\omega)$ is the set of actions optimal at ω. Therefore, for each ω and by the ACOE, one has

$$v_{\vec{\varepsilon}_n} + \theta_{\vec{\varepsilon}_n}(\omega) = r_n(\omega, \beta_n(\omega)) + \mathbf{E}_{q_n(\cdot|\omega,\beta_n(\omega)}\theta_{\vec{\varepsilon}_n}(\omega').$$

Given the definition of r_n, q_n and $\rho(\omega)$, the right-hand side is also equal to

$$r(\omega, \rho(\omega)) + x(\omega, \rho(\omega)) + \mathbf{E}_{q(\cdot|\omega,\rho(\omega)}\theta_{\vec{\varepsilon}_n}(\omega').$$

Next, it follows from the definition of x_{eq} that

$$r(\omega, b) + x(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b}\theta_{\vec{\varepsilon}_n}(\omega')$$

is independent of $b \in \text{supp } \rho(\omega) = B_0(\omega)$.

On the other hand, for $b \notin B_0(\omega)$ and $\bar{b} \in B_n(\omega)$, one has $r_n(\omega, b) + \mathbf{E}_{q_n(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}(\omega') < r_n(\omega, \bar{b}) + \mathbf{E}_{q_n(\cdot|\omega,\bar{b})}\theta_{\vec{\varepsilon}_n}(\omega')$, which implies $r(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}(\omega') < r(\omega, \bar{b}) + \mathbf{E}_{q(\cdot|\omega,\bar{b})}\theta_{\vec{\varepsilon}_n}(\omega')$, which yields in turn

$$r(\omega, b) + x(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}(\omega') < r(\omega, \bar{b}) + x(\omega, \bar{b}) + \mathbf{E}_{q(\cdot|\omega,\bar{b})}\theta_{\vec{\varepsilon}_n}(\omega').$$

Together, these observations yield

$$v_{\vec{\varepsilon}_n} + \theta_{\vec{\varepsilon}_n} = \max_{b \in B} \left\{ r(\omega, b) + x(\omega, b) + \mathbf{E}_{q(\cdot|\omega,b)}\theta_{\vec{\varepsilon}_n}(\omega') \right\},$$

with the maximum being achieved by $\rho(\omega)$. This proves the claim. ■

Claim 15 *For every $\omega, \tilde{\omega} \in \Omega$, one has*

$$r(\omega, \rho(\omega)) + x(\omega, \rho(\omega)) + \mathbf{E}_{q(\cdot|\omega,\rho(\omega))}\theta_{\vec{\varepsilon}_n}(\omega') \geq r(\omega, \rho(\tilde{\omega})) + x(\tilde{\omega}, \rho(\tilde{\omega})) + \mathbf{E}_{q(\cdot|\omega,\rho(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}(\omega'), \quad (27)$$

with a strict inequality if $\rho(\omega) \neq \rho(\omega')$.

Proof. Fix $\omega, \tilde{\omega} \in \Omega$. Note that $\rho(\omega) = \rho(\tilde{\omega})$ if and only if $B_k(\omega) = B_k(\tilde{\omega})$ for $k = 0, \ldots, n$. Assume first that $\rho(\omega) = \rho(\tilde{\omega})$. Then, using $x_{\text{eq}}(\omega, \rho(\omega)) = 0$, one has

$$\begin{aligned} x(\omega, \rho(\omega) &= \sum_{k=1}^{n} \varepsilon_k c_{k-1} \phi(|B_{k-1}(\omega)|) \\ &= \sum_{k=1}^{n} \varepsilon_k c_{k-1} \phi(|B_{k-1}(\tilde{\omega})|) = x(\tilde{\omega}, \rho(\tilde{\omega})). \end{aligned}$$

Thus, (27) holds with equality.

Assume next that $\rho(\omega) \neq \rho(\tilde{\omega})$, and denote by $\bar{k}$ the smallest k such that $B_k(\omega) \neq B_k(\tilde{\omega})$. Since $B_n = B_{n-1}$, one has $\bar{k} < n$. We prove that (27) holds with a strict inequality by looking at the decomposition of ρ as a weighted sum of the uniform distributions β_k.

- For $k < \bar{k}$, one has $\beta_k(\omega) = \beta_k(\tilde{\omega})$, hence

$$r(\omega, \beta_k(\omega)) + c_k\phi(|B_k(\omega)|) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\omega))}\theta_{\vec{\varepsilon}_n}(\omega') = r(\omega, \beta_k(\tilde{\omega})) + c_k\phi(|B_k(\tilde{\omega})|) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}(\omega').$$

- For $\bar{k} < k < n$, we will rely on the assumption that ε_k is quite small compared to $\varepsilon_{\bar{k}}$. Plainly, one has, for some constant C which only depends on the primitives of the MDP

$$\begin{aligned} &r(\omega, \beta_k(\omega)) + c_k\phi(|B_k(\omega)|) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\omega))}\theta_{\vec{\varepsilon}_n}(\omega') \\ &\geq r(\omega, \beta_k(\tilde{\omega})) + c_k\phi(|B_k(\tilde{\omega})|) + \mathbf{E}_{q(\cdot|\omega,\beta_k(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}(\omega') - C. \end{aligned}$$

  Hence, when multiplied by $\varepsilon_{\bar{k}+1}$, the difference between the two sides of the latter inequality is very small compared to $\varepsilon_{\bar{k}+1}$, and in particular less that $\alpha\varepsilon_{\bar{k}+1}c_{\bar{k}}$.

- For $k = n$, and since $B_n(\omega)$ are the actions optimal at ω in $\mathcal{M}_n(\vec{\varepsilon}_n)$, one has as noted previously,

$$r(\omega, \beta_n(\omega)) + \mathbf{E}_{q(\cdot|\omega,\beta_n(\omega))}\theta_{\vec{\varepsilon}_n}(\omega') \geq r(\omega, \beta_n(\tilde{\omega})) + \mathbf{E}_{q(\cdot|\omega,\beta_n(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}(\omega').$$

We are left with $\bar{k} = k$, and distinguish two cases. Assume first that $b \notin B_{\bar{k}}(\omega)$ for some $b \in B_{\bar{k}}(\tilde{\omega})$. In that case,

$$r(\omega, \beta_{\bar{k}}(\omega)) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))}\theta_{\vec{\varepsilon}_{\bar{k}}}(\omega') > r(\omega, \beta_{\bar{k}}(\tilde{\omega})) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_{\bar{k}}}(\omega') + c_{\bar{k}} \times \frac{|B_{\bar{k}}(\tilde{\omega}) \setminus B_{\bar{k}}(\omega)|}{|B_{\bar{k}}(\tilde{\omega})|}$$

(because all actions in $B_{\bar{k}}(\tilde{\omega}) \setminus B_{\bar{k}}(\omega)$ are played with probability $\frac{1}{|B_{\bar{k}}(\tilde{\omega})|}$ and each leads to a loss of at least $c_{\bar{k}}$). Since $\varepsilon_{\bar{k}+1}, \ldots, \varepsilon_n$ are small (given $\varepsilon_{\bar{k}}$), the latter inequality still holds when $\theta_{\vec{\varepsilon}_n}$ is substituted to $\vec{\varepsilon}_{\bar{k}}$. This implies

$$\begin{aligned} & r(\omega, \beta_{\bar{k}}(\omega)) + c_{\bar{k}}\phi(|B_{\bar{k}}(\omega)|) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))}\theta_{\vec{\varepsilon}_n}(\omega') \\ > \;& r(\omega, \beta_{\bar{k}}(\tilde{\omega})) + c_{\bar{k}}\phi(|B_{\bar{k}}(\tilde{\omega})|) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}(\omega') + c_{\bar{k}}\left(\frac{1}{|B_{\bar{k}}(\tilde{\omega})|} + \phi(|B_{\bar{k}}(\omega)|) - \phi(|B_{\bar{k}}(\tilde{\omega})|)\right) \\ > \;& r(\omega, \beta_{\bar{k}}(\tilde{\omega})) + c_{\bar{k}}\phi(|B_{\bar{k}}(\tilde{\omega})|) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_n}(\omega') + c_{\bar{k}}\alpha, \end{aligned}$$

using property (i) of α.

Assume now that $B_{\bar{k}}(\tilde{\omega})$ is a strict subset of $B_{\bar{k}}(\omega)$, so that

$$r(\omega, \beta_{\bar{k}}(\omega)) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))}\theta_{\vec{\varepsilon}_{\bar{k}}}(\omega') = r(\omega, \beta_{\bar{k}}(\tilde{\omega})) + \mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_{\bar{k}}}(\omega'),$$

because $\beta_k(\omega)$ is optimal in $\mathcal{M}_k(\vec{\varepsilon}_k)$. This implies

$$r(\omega, \beta_{\bar{k}}(\omega))+c_{\bar{k}}\phi(|B_{\bar{k}}(\omega)|)+\mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\omega))}\theta_{\vec{\varepsilon}_{\bar{k}}}(\omega') > r(\omega, \beta_{\bar{k}}(\tilde{\omega}))+c_{\bar{k}}\phi(|B_{\bar{k}}(\tilde{\omega})|)+\mathbf{E}_{q(\cdot|\omega,\beta_{\bar{k}}(\tilde{\omega}))}\theta_{\vec{\varepsilon}_{\bar{k}}}(\omega')+\alpha c_{\bar{k}},$$

using property (ii) of α.

Since $\theta_{\vec{\varepsilon}_n}$ is very close to $\theta_{\vec{\varepsilon}_{\bar{k}}}$, the latter inequality still holds true when $\theta_{\vec{\varepsilon}_n}$ is substituted to $\theta_{\vec{\varepsilon}_{\bar{k}}}$. The desired inequality follows by summing over all $k = 1, \ldots, n$.

■

F Proof of Theorem 5

Most computations in Section F.2 will be omitted. Transition phases will rely on the strictly truthful pair $(\rho_{\text{ext},0}, x_0)$ constructed in Section B.2.2, with $K_0 = A$ and $\rho_0 : S \times K_0 \to A$. We supplement the transfers x_0 of Section B.2.2 with transfers $\bar{x}_0 : K_0 \times Y \to \mathbf{R}^I$ which induce obedience to ρ_0, and still denote by $x_0 : S \times K_0 \times Y \to \mathbf{R}^I$ the total transfers. We abbreviate the relative values $\theta_{\rho_0, r+x_0}$ to θ_0, and we let $\bar{r} \geq 1$ be a uniform bound on r and θ_0.

F.1 Auxiliary Zero-Sum Games

Throughout this section, we fix a player $i \in I$, and will introduce games between i and $-i$. W.l.o.g., all strategies of player i are here "babbling."

F.1.1 Preliminaries

For $k \in \mathbf{N}$ and $j \neq i$, we let $\mathcal{A}_k^j \subset \Delta(A^j)$ be a finite, $\frac{1}{k}$-dense subset of $\Delta(A^j)$. That is, for each $\alpha^j \in \Delta(A^j)$, there exists $\alpha_k^j \in \mathcal{A}_k^j$ such that $\|\alpha^j - \alpha_k^j\|_{L^1} < \frac{1}{k}$. We let Σ_k^j be the set of repeated game strategies of player j with the property that the mixed action of j in each round n belongs to $\mathcal{A}_k^j$ and only depends on the past public signals $y_1^i, \ldots, y_{n-1}^i$ relative to player i. We set $\Sigma_k^{-i} := \times_{j \neq i} \Sigma_k^i$, and let

$$w_k^i := \lim_{\delta \to 1} \min_{\sigma^{-i} \in \Sigma_k^{-i}} \max_{\sigma^i} \gamma_\delta^i(s^i, \sigma^i, \sigma^{-i})$$

be the long-run minmax payoff when players $-i$ are constrained to strategies in Σ_k^{-i}.[57] Thanks to the irreducibility assumption, there exists $c > 0$ such that the following holds: for each $k \in \mathbf{N}$, $j \neq i$ and each strategy σ^j, there exists $\sigma_k^j \in \Sigma_k^j$ such that

$$\gamma_\delta^j(s, \sigma^{-j}, \sigma_k^j) < \gamma_\delta^j(s, \sigma) + \frac{c}{k},$$

for every $\delta < 1$ and σ^{-j}.[58] Hence, $\lim_{k \to +\infty} w_k^i = w^i$.

Since strategies of player $-i$ ignore (y_n^j) $(j \neq i)$, we may restrict ourselves to strategies of player i which are independent as well of the public signals (y_n^j), $j \neq i$, relative to other players.

Let an arbitrary state $\bar{s}^i \in S^i$ be given, and $k \in \mathbf{N}$ be fixed. Given an horizon $T \in \mathbf{N}$, we let $G_k^i(\bar{s}^i, T)$ be the zero-sum game with T rounds between i and $-i$ with no communication, initial state $\bar{s}^i$, payoff $\frac{1}{T}\sum_{n=1}^{T} r^i(s_n^i, a_n)$ and where players $-i$ are restricted to Σ_k^{-i}. Denote by

$$w_k^i(T) := \min_{\sigma^{-i} \in \Sigma_k^{-i}} \max_{\sigma^i} \mathbf{E}_{\bar{s}^i, \sigma} \left[\frac{1}{T} \sum_{n=1}^{T} r^i(s_n^i, a_n) \right] \tag{28}$$

the minmax of $G^i(\bar{s}^i, T)$. Using irreducibility, one has $\lim_{T \to +\infty} w_k^i(T) = w_k^i$ for each $\bar{s}^i$.

Given k and T, we fix a strategy profile $\sigma_k^{-i} \in \Sigma_k^{-i}$ that achieves the minimum in (28). For $\alpha_k^{-i} \in \mathcal{A}_k^{-i}$, let $T(\alpha_k^{-i})$ be the (random) set of rounds in which σ_k^{-i} prescribes α_k^{-i}, and let $f_{\alpha_k^{-i}} \in \Delta(Y)$ denote the empirical distribution of the public signals received in $T(\alpha_k^{-i})$. Intuitively, if some player $j \neq i$ is playing according to σ_k^j, the signals (y_n^j) received in $T(\alpha_k^{-i})$ are *i.i.d.*, and drawn from $p^j(\cdot \mid \alpha_k^j)$. Hence, whenever $|T(\alpha_k^{-i})|$ is large and with high probability, $f_{\alpha_k^{-i}}$ should be close to the distribution $g_{\alpha_k^{-i}}^j \in \Delta(Y)$ defined as

$$g_{\alpha_k^{-i}}^j(y) = f_{\alpha_k^{-i}}(y^{-j}) p^j(y^j \mid \alpha_k^j).$$

This motivates the definition of

$$D^j := \sum_{\alpha_k^{-i} \in \mathcal{A}_k^{-i}} \frac{|T(\alpha_k^{-i})|}{T} \| f_{\alpha_k^{-i}} - g_{\alpha_k^{-i}}^j \|_{L^1}.$$

Claim 16 below formalizes this intuition. In words, and provided that T is large enough, player j can ensure that $D^j < \varepsilon$ with high probability by playing $\sigma_{k,T}^j$.

[57] It is independent of s^i.

[58] This assertion also relies on the product monitoring assumption. Under this assumption, public communication and public signals y_n^j relative to $j \neq i$ cannot be used by players $-i$ as a means to privately correlate their actions against $-i$.

Claim 16 *Given $\varepsilon > 0$, there exists $T_0 \geq 0$ such that*

$$\mathbf{P}_{\sigma^j_{k,T},\sigma^{-j}}(D^j > \varepsilon) < \varepsilon,$$

for all $T \geq T_0$, $j \neq i$ and σ^{-j}.

Claim 16 follows from Gossner (1995), who uses Blackwell's theory of approachability. It will be combined with the claim below, which asserts that player i is effectively punished when all players $j \neq i$ pass the test $D^j < \varepsilon$ with high probability.

Claim 17 *Let $\varepsilon > 0$ and T be given, and let σ be a strategy profile such that $\mathbf{P}_\sigma(D^j > \varepsilon) < \varepsilon$ for each $j \neq i$. Then*

$$\mathbf{E}_{\bar{s}^i,\sigma}\left[\frac{1}{T}\sum_{n=1}^{N} r^i(s^i_n, a_n)\right] < w^i_{k,T} + \bar{r}(I+2)\varepsilon.$$

Proof. On the event $\mathcal{D}^{-i} := \cap_{j\neq i}\{D^j \leq \varepsilon\}$ one has $\sum_{\mathcal{A}^{-i}_k} \frac{|T(\alpha^{-i}_k)|}{T}\|f_{\alpha^{-i}_k} - g^j_{\alpha^{-i}_k}\|_{L^1} \leq \varepsilon$, for each $j \neq i$, which implies, by repeated substitution,

$$\sum_{\mathcal{A}^{-i}_k} \frac{|T(\alpha^{-i}_k)|}{T}\left(\sum_y |f_{\alpha^{-i}_k}(y) - f_{\alpha^{-i}_k}(y^i) \times_{j\neq i} p^j(y^j \mid \alpha^j_k)|\right) < I\varepsilon.$$

We fix now an arbitrary private history (s^i_n, a^i_n, y_n) of player i, and compare the realized payoff $\frac{1}{T}\sum_{n=1}^{n} g^i(s^i_n, a^i_n, y_n)$ to its "expectation," assuming (y^j_n) are drawn using $\sigma^j_{k,T}$. Formally,

$$\frac{1}{T}\sum_{n=1}^{n} g^i(s^i_n, a^i_n, y_n) = \frac{1}{T}\sum_{\mathcal{A}^{-i}_k}\sum_{T(\alpha^{-i}_k)}\left(g^i(s^i_n, a^i_n, y_n) - \sum_{\tilde{y}^{-i}\in Y^{-i}} g^i(s^i_n, a^i_n, \tilde{y}^{-i}, y^i_n) \times p^{-i}(\tilde{y}^{-i} \mid \alpha^{-i}_k)\right)$$

$$+\frac{1}{T}\sum_{\mathcal{A}^{-i}_k}\sum_{T(\alpha^{-i}_k)}\sum_{\tilde{y}^{-i}} g^i(s^i_n, a^i_n, \tilde{y}^{-i}, y^i_n) \times p^{-i}(\tilde{y}^{-i} \mid \alpha^{-i}_k).$$

The expectation of the second term is independent of σ^{-i} and is equal to $\mathbf{E}_{\bar{s}^i,\sigma^i,\sigma^{-i}_{k,T}}\left[\frac{1}{T}\sum_{n=1}^{T} r^i(s^i_n, a_n)\right] \leq w^i_k(T)$. Since the first term is bounded by $2\bar{r}$, and by $\bar{r}I\varepsilon$ on the event $\mathcal{D}^{-i}$, the result follows. ■

F.1.2 Auxiliary Games

From now on and given $\delta < 1$, we set $T := \dfrac{1}{\sqrt{1-\delta}}$. Given $\delta < 1$, transfers $x : S \times Y^T \to \mathbf{R}^I$, and a state profile $s \in S$, we let $G(s, \delta, x)$ denote the game of T rounds (ending after the draw of s_{T+1}), with initial state profile s, no communication, and with payoff

$$\frac{1-\delta}{1-\delta^T}\left\{\sum_{n=1}^{T}\delta^{n-1}r(s_n,a_n) + \delta^T x(s,\vec{y}) + \delta^T\theta_0(s_{T+1},\bar{a}_0)\right\},$$

where $\vec{y} := (y_1, \ldots, y_T)$ is the sequence of public signals received along the play, and $\bar{a}_0 \in A$ is fixed.

The following result will serve as the building block of the equilibrium construction of punishment phases.

Lemma 13 *Given $\varepsilon > 0$, there exists $\kappa_* \in \mathbf{R}$ and $\delta_* < 1$ such that for all $\delta > \delta_*$, there exist $x : S \times Y^T \to \mathbf{R}^I$ and $\gamma \in \mathbf{R}^I$ with the following properties:*

(a) *For all $s \in S$, γ is a sequential equilibrium payoff of $G(s, \delta, x)$;*

(b) *$\gamma^i < w^i + \varepsilon$;*

(c) *$x^i \geq 0$ and $\|x\| \leq \kappa_* T$.*

Proof. Let $\varepsilon > 0$ be given and pick $\varepsilon' < \frac{\varepsilon}{2\bar{r}(I+5)}$. Choose $k \in \mathbf{N}$ such that $|w_k^i - w^i| < \varepsilon'$, choose $C > \dfrac{4\bar{r}}{\varepsilon'}$, and apply Claim 16 with ε' to get T_0. We will show that the result holds with $\kappa_* := 2C$ and $\delta_* < 1$ large enough so that (i) $T \geq T_0$, (ii) $|w_k^i - w_k^i(T)| < \varepsilon'$ (for each $\bar{s}^i$), (iii) $\dfrac{1-\delta}{1-\delta^T}C < 1$ and both inequalities displayed below hold for each $\delta > \delta_*$:

$$-\delta^T\bar{r} - \sum_{n=1}^{T}\delta^{n-1}\bar{r} + \delta^T CT(1-\varepsilon') > \sum_{n=1}^{T}\delta^{n-1}\bar{r} + \delta^T C(1-2\varepsilon') + \delta^T\bar{r}, \tag{29}$$

and for each sequence $(u_1, \ldots, u_T)$,

$$\left|\frac{1-\delta}{1-\delta^T}\sum_{n=1}^{T}\delta^{n-1}u_n - \frac{1}{T}\sum_{n=1}^{T}u_n\right| < \varepsilon'\max(u_1,\ldots,u_T).$$

Let $\delta > \delta_*$ be arbitrary, and define $x_* : Y^T \to \mathbf{R}^I$ by $x_*^i(\cdot) = 0$ and $x_*^j(\vec{y}) = -CT$ if $D^j > \varepsilon'$ and $x_*^j(\vec{y}) = 0$ otherwise, so that $\|x_*(\cdot)\| \leq \frac{1}{2}\kappa_* T$ and $x_*^i(\cdot) \geq 0$.

Given $s \in S$, let σ_s be any sequential equilibrium of $G(s, \delta, x_*)$, with payoff $\gamma_s(\sigma_s) \in \mathbf{R}^I$. By the choice of C and δ_*, one has $\mathbf{P}_{s,\sigma}(D^j > \varepsilon') < 2\varepsilon'$ for every $j \neq i$.[59] Therefore, by Claim 17, one has

$$\mathbf{E}_{s,\sigma}\left[\frac{1}{T}\sum_{n=1}^{T} r^i(s_n^i, a_n)\right] < w_k^i + \varepsilon' + 2\bar{r}(I+2)\varepsilon',$$

which implies

$$\gamma_s^i(\sigma_s) < w^i + 2\bar{r}(I+4)\varepsilon'.$$

Since $\mathbf{P}_s(D^j > \varepsilon') < 2\varepsilon'$ for each $j \neq i$, it follows from the specification of x_* and δ_* that $\|\gamma_s(\sigma_s)\| \leq 14\bar{r}$. Set then $\bar{x}^j(s) := \max_{s' \in S} \gamma_{s'}^j(\sigma_{s'}) - \gamma_s^j(\sigma_s)$ for each $s \in S$ and $j \in I$, and

$$x(s, \vec{y}) := x_*(\vec{y}) + \bar{x}(s).$$

Plainly, σ_s is still a sequential equilibrium of $G(s, \delta, x)$ for each s, and the payoff vector induced by σ_s is now independent of s. Moreover, since $0 \leq \bar{x}(\cdot) \leq 14\bar{r}$, and by the choice of ε', both (b) and (c) hold as well. ■

We denote by $\mathcal{G}_\varepsilon^i$ the compact set of all accumulation points of such equilibrium payoffs $\gamma \in \mathbf{R}^I$, as $\delta \to 1$. Before we move on to the equilibrium construction, two remarks are in order. Note first that property (c) can be strengthened to $x^i(\cdot) \geq \varepsilon''$, where $\varepsilon'' < \varepsilon$ is arbitrary. (Indeed, for given $0 < \varepsilon'' < \varepsilon$, it suffices to first apply the current version of Lemma 13 with $\varepsilon - \varepsilon''$, and then add ε'' to x^i).

Because of irreducibility, there is a constant c (which only depends on the primitives of the game) such that, for $j \in I$, $s \in S$ and $t^j \in S^j$, the highest payoff achievable by j against σ_s^{-j} in the two games $G(t^j, s^{-j}, \delta, x(s, \cdot))$ and $G(s, \delta, x(s, \cdot))$ differ by at most $(1-\delta)c$. Since the latter payoff is equal to γ^j, the former does not exceed $\gamma^j + (1-\delta)c$. Since γ^j is also the payoff induced by $\sigma_{t^j, s^{-j}}$ in the game $G(t^j, s^{-j}, \delta, x(t^j, s^{-j}, \cdot))$, this implies that the benefit to player j of pretending that his initial state is s^j when it is t^j is bounded by $(1-\delta)c$.

F.2 Equilibrium Construction

We only provide a sketch. We start as in Section B.2. To unify notations, we set $\hat{k}_1(\lambda) = \bar{k}_1(\lambda)$ for $\lambda \neq -e^i$ and $\hat{k}_1(-e^i) = -w^i$ for $i \in I$. Since $\hat{k}_1(\cdot)$ is lower semi-continuous on Λ, there exists $\varepsilon_0 > 0$ such that

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

[59] Indeed, by (29), any strategy $\tilde{\sigma}^j$ such that $\mathbf{P}_{\tilde{\sigma}^j, \sigma^{-j}}(D^j > \varepsilon') < \varepsilon'$ is strictly preferred to any strategy $\tilde{\sigma}^j$ such that $\mathbf{P}_{\tilde{\sigma}^j, \sigma^{-j}}(D^j > \varepsilon') > 2\varepsilon'$. And $\sigma_{k,T}^j$ satisfies the former condition by Claim 16.

For each player i, we apply Lemma 13 with $0 < \varepsilon_0' < \varepsilon_0$ (so that $x^i(\cdot) \geq \varepsilon_0'$) and get κ_* and δ_*. We next pick $\varepsilon_0'' < \dfrac{\varepsilon_0'}{\kappa_*}$. With these choices, for fixed i and $\delta > \delta_*$, the payoff vector γ and the transfers x satisfy $\gamma^i < w^i + \varepsilon_0$, $\|x\| < \kappa_* T$, and $\lambda \cdot x_*(\cdot) \leq 0$ whenever $\|\lambda - (-e^i)\| < \varepsilon_0''$.

Parameters are chosen as follows. We first pick the parameter $0 < \beta < \frac{1}{2}$ of the length of transition phases, next choose κ to be large enough. Next, as before, pick $\varepsilon > 0$ small enough. Finally, we choose $\bar{\delta} < 1$ high enough. Computations are highly similar to those in Sections A.1.2 and A.1.3. They are therefore omitted, and we do not list conditions to be satisfied by κ, ε and $\bar{\delta}$.

We let $z \in Z$ be given, and let $\pi_1 \in \times_i \Delta(S^i)$ be the distribution of the initial state. The play is divided in a sequence of phases, with odd phases being transition phases. Slight adjustments in the strategies are needed (as compared with Section A.1.2), and we detail the updating from one transition phase to the following transition phase. The transition phase k starts with with a target payoff $z_{(k)}$ which is deduced from past public play. We set $(\rho_{(k)}, x_{(k)}) = (\rho_{\text{ext},0}, x_0)$, $v_{(k)} := \mathbf{E}_{\mu_{\rho_{\text{ext},0}}}[r(s,a) + x_0(s, k_0, y)]$, and $\theta_{(k)} := \theta_0$. In each round, the p.r.d. chooses with probability $\xi_* := (1-\delta)^\beta$ whether to start a new phase. In the first round $n = \tau_{(k+1)}$ of the following phase $k+1$, we first define the auxiliary target $w_{(k+1)}$ according to

$$\xi_* w_{(k+1)} + (1 - \xi_*) z_{(k)} = \frac{1}{\delta} z_{(k)} - \frac{1-\delta}{\delta} v_{(k)} + \frac{1-\delta}{\delta} x_{(k)}(\omega_{\text{pub},n-1}),$$

next apply Lemma 1 with $z := w_{(k)}$ to get $\lambda_{(k+1)}$.

If $\|\lambda_{(k+1)} - (-e^i)\| \geq \varepsilon_0''$ for all i, we apply Lemma 2 to get $(v_{(k+1)}, \rho_{(k+1)}, x_{(k+1)}) \in \mathcal{S}$, and finally update $z_{(k+1)}$ as

$$z_{(k+1)} = w_{(k+1)} + (1-\delta)\left(\left(1 + \frac{1-\delta}{\delta\xi}\right)\theta_{(k)}(m_{n-1}, m_n) - \theta_{(k+1)}(\omega_{\text{pub},n-1}, m_n)\right).$$

Then in each round, the p.r.d. chooses with probability ξ whether to start a new phase. In round $\tau_{(k+2)}$ the auxiliary target will be updated to $w_{(k+2)}$ according to (4) and $z_{(k+2)}$ in the following transition phase is defined by (5).

If instead $\|\lambda_{(k+1)} - (-e^i)\| < \varepsilon_0''$ for some i, we apply Lemma 13 with player i, and get $x : S \times Y^T \to \mathbf{R}^I$ and γ. We set $v_{(k+1)} = \gamma$, and $x_{(k+1)} = x$. In that case the duration of phase $k+1$ is T. In round $\tau_{(k+2)} := \tau_{(k+1)} + T$, we set

$$z_{(k+2)} = \frac{1}{\delta^T} z_{(k+1)} - \frac{1-\delta^T}{\delta^T} v_{(k+1)} + (1-\delta) x_{(k+1)}(m_{\tau_{(k+1)}}, y_{\tau_{(k+1)}}, \ldots, y_{\tau_{(k+1)}+T-1}).$$

That this recursive construction is well-defined follows as in Lemma 5.

Under σ, players report truthfully and play $\rho_{(k)}$ in any phase k that is not a punishment phase. If $\|\lambda_{(k)} - (-e^i)\| < \varepsilon_0''$, we let $\sigma_{(k)}$ be a sequential equilibrium in $G(m_{\tau_{(k)}}, \delta, x_{(k)})$ with payoff $v_{(k)}$. Under σ, player j plays $\sigma^j_{(k)}$ if his report in round $\tau_{(k)}$ is truthful, and otherwise plays a (sequentially) best reply to $\sigma^{-j}_{(k)}$ in the game $G(s^j_{\tau_{(k)}}, m^{-j}_{\tau_{(k)}}, \delta, x_{(k)})$.

As in Section A.1.3, one can establish that the continuation payoff under σ is equal to $z_{(k)}$ at the beginning of a punishment phase, and $z_{(k)} + (1-\delta)\theta_{(k)}$ in any round that does not belong to a punishment phase.

That a player cannot profitably deviate at the action step follows from the definition of σ in a punishment phase, and as in Theorem 2 otherwise. That a player cannot profitably deviate at the reporting step of a non-transition phase is clear during punishment phases since reports are ignored, and otherwise follows as before.

Consider finally the reporting step in a round n belonging to a transition phase. In the specific case where n is the first round following a punishment phase, reports are ignored, and the action being played is $\bar{a}_0$, hence truthful reporting is trivially optimal. Otherwise, the belief of player j over S^{-j} has full support, and the optimality of truth-telling follows along earlier lines, using that (i) $(\rho_{\text{ext},0}, x_0)$ is strictly truthful, and that (ii) the (*ex post*) marginal benefit of having misreported, conditional on the p.r.d. choosing to start a new phase, is at most of the order of $(1-\delta)$ –see the remark at the end of Section F.1.2.