BLACKWELL EQUILIBRIA IN REPEATED GAMES

COSTAS CAVOUNIDIS*, SAMBUDDHA GHOSH†, JOHANNES HÖRNER‡, EILON SOLAN§, AND SATORU TAKAHASHI¶

Abstract. We apply Blackwell optimality to repeated games. A Blackwell (subgame-perfect, perfect public, etc.) equilibrium is an equilibrium whose strategy profile is sequentially rational for all high enough discount factors simultaneously. The bite of this requirement depends on the monitoring structure. Under perfect monitoring, a "folk" theorem holds relative to an appropriate notion of minmax. Under imperfect public monitoring, absent a public randomization device, any perfect public equilibrium generically involves pure action profiles or stage-game Nash equilibria only. Under private conditionally independent monitoring, in a class of games that includes the prisoner's dilemma, the stage-game Nash equilibrium is played in every round. **Keywords.** Repeated games, Blackwell optimality.

* University of Warwick, c.cavounidis@warwick.ac.uk
† Chinese University of Hong Kong
‡ Yale University and CNRS (TSE)
§ Tel-Aviv University
¶ National University of Singapore
Date: December 5, 2022.
We wish to thank Gabriel Carroll, Aniruddha Dasgupta, Olivier Gossner, Bart Lipman, Elliot Lipnowski, Francesco Nava, Jawwad Noor, Juan Ortner, Phil Reny, Takuo Sugaya, and Balazs Szentes, as well as audiences at Boston University, the Econometric Society Summer Meetings, LSE, MIT, NYU, and the University of Warwick for useful comments and discussions. Solan thanks the support of the Israel Science Foundation, Grants #217/17 and #211/22.

1. Introduction

By and large, the economic literature on repeated games has adopted discounting as the payoff criterion. It is technically convenient, and captures the idea that the distant future does not matter much for current decisions, which is certainly "more realistic than its opposite," as a leading microeconomics textbook puts it.[1] Yet, at least in the context of repeated games, it has two consequences that are often viewed as undesirable. First, no action profile can typically be ruled out; in this sense little is said about behavior.[2] Second, predictions depend on common knowledge of the exact discount factor, undoubtedly a strong assumption.[3]

In this paper, we study *Blackwell equilibria*, that is, equilibria whose strategy profiles are optimal for all high discount factors simultaneously. Hence, they preserve the property that time isn't free, and that every round matters for the player's payoff, yet, by definition, they cannot depend on the exact value of the discount factor. This is a payoff criterion, not a solution concept. For the latter, we adopt what is commonly used depending on the environment: subgame-perfect Nash equilibrium, perfect public equilibrium, sequential equilibrium, etc. With some abuse, we refer to the relevant notion of equilibrium under the Blackwell criterion as Blackwell equilibrium. The name "Blackwell equilibrium" derives from "Blackwell optimality," the corresponding concept introduced for Markov decision processes by Blackwell (1962).

Robustness to the exact discount rate admits several interpretations. When the rate is thought of as arising from the random length of the actual interaction, there are many situations in which players are uncertain about exactly how long this interaction will take place, and this uncertainty might be sufficiently vague that modeling it explicitly seems futile.[4] The same applies when the discount rate pertains to the players' time preferences.[5] Uncertainty regarding future interest rates is both subjective and significant; and it has large, negative and persistent effects on the economy

[1] See Mas-Colell, Whinston, and Green (1995), p.734.
[2] See the discussion in Aumann and Maschler (1995), p.139.
[3] This second issue has led many game theorists (Aumann and Maschler, among many others) to favor undiscounted payoff criteria. This is throwing the baby out with the bathwater, as forsaking impatience reinstates the unrealistic "opposite" mentioned above.
[4] As Aumann and Maschler (1995, p.133) put it, "there is a limit to the amount of detail that can usefully be put into a model, or indeed that the players can absorb or take into account."
[5] Admittedly, in that case, by revealed preference, a player "knows" his own discount factor. The case in which each player knows his own discount factor, but not the others', is taken up in Section 5.

(see, *e.g.*, Istrefi and Mouabbi, 2018).[6] When players in the game are a convenient proxy for groups of agents (countries, political parties, firms, etc.), then Blackwell equilibria have the desirable feature that they are unanimously viewed as optimal by the constituents of each group, independent of exactly how patient each of them is, provided that they are all sufficiently patient. Finally, from the point of view of the analyst, they allow to predict, or explain, behavior that might apply to a variety of situations, which might differ in the details of the interaction length.

Our goal is to understand how this more restrictive payoff criterion affects the usual predictions about payoffs and action profiles in infinitely repeated games, under various monitoring structures. Our main result is that its (relative) bite is increased as monitoring "worsens," so to speak. Loosely speaking, this is because robustness to discounting makes it difficult to enforce mixed actions. Yet, the role of mixed strategies becomes progressively more important as the information structure shifts from perfect monitoring, to imperfect public monitoring, and finally to imperfect private monitoring.

In games of perfect monitoring, Blackwell (subgame-perfect) equilibria still span a large set of equilibrium payoffs, but not as large as under the (limit of) discounting criterion. Indeed, the standard folk theorem (see Fudenberg and Maskin, 1986, hereafter FM) requires punishments that, depending on the stage game, might involve mixing by the punishing players. To make the players indifferent over the support of their actions, these players must be compensated in the continuation game, as a function of the action they have chosen. To achieve indifference, this compensation must be finely tuned to the discount factor. We show that there is no way around this difficulty. As a result, a new notion of minmax payoff must be introduced, capturing the fact that punishing players must be myopically indifferent across all actions within the support of their mixed action (but not necessarily over all actions available to them).

We show that this is the only adjustment that must be made to the "standard" statement of the folk theorem – indeed, mixed actions play no other role in the usual proofs under perfect monitoring. The construction involves the same ingredients as the proof of FM, and can be achieved using a variation of simple strategies (Abreu,

[6]As discussed below, our analysis is readily adjusted to the case in which the discount factor isn't constant over time, or deterministic.

1988). Not too surprisingly, this folk theorem can be extended to imperfect public monitoring, in the special case in which monitoring satisfies product structure, individual full rank, and a public randomization device is available.[7]

In general, under imperfect public monitoring, it is known that unpredictable behavior serves another purpose. Mixed actions enlarge the set of detectable deviations, and hence affect the sufficient conditions usually made for the folk theorem to hold (Fudenberg, Levine, and Maskin, 1994, hereafter FLM). Yet, the impossibility of fine-tuning continuation payoffs in order to compensate players for mixing in a way that would be independent of the discount rate further restricts the action profiles that can be implemented. Absent a public randomization device, only stage-game Nash equilibria and pure action profiles can be played in a (perfect public) Blackwell equilibrium (generically, see Proposition 2). That is, the only mixed actions that can be played are stage-game Nash equilibria: it no longer suffices that players be myopically indifferent over the support of their mixed action. This is because, unless the action profile is a Nash equilibrium of the stage game, the continuation play must depend on the realized signal, which makes it impossible for players to be indifferent over multiple actions (as they generically induce distinct distributions over public signals), even if they are myopically indifferent over those.

A major difficulty in the analysis under imperfect monitoring is that such games are usually studied via recursive techniques involving the set of equilibrium payoffs (see Abreu, Pearce, and Stacchetti, 1990). Because payoffs of a given strategy profile, and the "self-generation" operator itself, depend on the discount rate, standard results are not as helpful here, since optimality must hold for an entire range of discount rates simultaneously.[8] Hence, our analysis must tackle directly the issue of the action profiles that can be enforced.

Finally, we show that behavior is further constrained once monitoring is private. In a class of games that includes the prisoner's dilemma, when monitoring satisfies conditional independence, the only Blackwell equilibrium outcome consists in the repetition of the stage-game Nash equilibrium. This is because indifference between actions is known to play a further role under private monitoring. Under conditional independence, with pure strategies, a player cannot tell, even statistically, whether his

[7]Admittedly, the product structure is very special, but it applies to important classes of games, such as games with one-sided imperfect monitoring, e.g. principal-agent games, and games with adverse selection and independent types.

[8]The alternative route in the literature involves review strategies, following Radner (1985). Unfortunately, review strategies don't specify behavior fully, making it difficult to ensure that behavior is independent of the discount rate.

opponent is supposed to "punish or reward" him; therefore, his opponent cannot be incentivized to select one or the other continuation strategy as a function of the signals he receives, unless he happens to be indifferent across those (see Matsushima, 1989). Hence, any non-trivial sequential equilibrium must involve indifferences (whether a player actually mixes or uses his private history to select one or the other continuation strategy). For the same reason as under public monitoring, such indifference is inconsistent with the robustness to the discount rate.

Related Literature. The Blackwell optimality criterion has been introduced by Blackwell (1962) for finite Markov decision processes, as a way of characterizing optimality in the undiscounted case. Blackwell shows that optimal policies exist, and provides (a pair of) optimality equations to solve for those. More recently, this criterion has been applied to stochastic games, both in discrete time (Singh, Hemachandra and Rao, 2013) and in continuous time (Singh, Hemachandra, 2016). The focus of these papers is to provide conditions under which (Nash) equilibria exist under this payoff criterion. They do this for games in which a single player controls the transitions and the payoff of the non-controller is additive in the players' actions. Indeed, existence is a non-trivial problem in the environments they consider. In repeated games, this is immediate, as the repetition of stage-game Nash equilibria is a Blackwell equilibrium. In contrast, we are interested in characterizing the set of such equilibria under different monitoring structures.

The motivation of our paper is related to Gossner (2020). Gossner's goal is also to define equilibria that are robust to slight perturbations in the repeated game. Gossner introduces incomplete penal codes as partial descriptions of equilibrium strategies, and studies to what extent such codes can be found, whose completion is allowed to depend on the fine details of the game. Because his class of perturbations does not only include the discount rate, but the payoff matrix itself, complete penal codes typically do not exist.

Slightly less related are some papers that focus on special classes of strategies. Kalai, Samet and Stanford (1988) study reactive equilibria, which are equilibria in which at least one player conditions his actions only on his own opponent's action and not on his own past actions. As they show, if a reactive strategy profile is robust to nearby discount factors, it must play the stage-game Nash equilibrium in the prisoner's dilemma.

2. Blackwell Equilibria Under Perfect Monitoring

This section studies Blackwell equilibria under perfect monitoring. First, we derive necessary conditions that such equilibria must satisfy. This leads to a modified notion of minmax payoff, and to a folk theorem relative to this notion.

2.1. **Notation and Definitions.** The set of players is $I = \{1, \ldots, n\}$. Player i's finite set of actions is A_i, and $A := \prod_{i \in I} A_i$ is the set of all pure action profiles. A mixed action of i is $\alpha_i \in \Delta A_i$, where ΔE is the set of all probability distributions on a set E; the set of (independent) mixed-action profiles is $\mathcal{A} := \prod_i \Delta A_i$. Player i's reward function is a map $g_i : A \to \mathbb{R}$, whose domain is extended to ΔA in the usual way, and $g := (g_i)_{i=1}^n$. The set of feasible payoffs is $F := \text{co}(g(A))$, where co denotes the convex hull. Denote this normal-form game by $G = \langle I; A, g \rangle$.

The stage game G is played at each $t \in \mathbb{Z}_+$. Denoting by $a^{(t)} \in A$ the action profile chosen in each round t, the history at the end of round $t \in \mathbb{Z}_+$ is $h^t = (a^{(1)}, \ldots, a^{(t)}) \in A^t =: H^t$, with h^0 the empty history and $H := \cup_{t=0}^{\infty} H^t$ the set of all histories. An outcome h^∞ is an infinite sequence $(a^{(t)})_{t=1}^{\infty}$. Given discount factor $\delta_i \in [0, 1)$, player i's (average discounted) payoff is defined as

$$U_i\left(h^\infty, \delta_i\right) := (1 - \delta_i) \sum_{t=1}^{\infty} {\delta_i}^{t-1} g_i(a^{(t)}). \tag{2.1}$$

This defines the repeated game $G^\infty(\boldsymbol{\delta})$, where the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$ is referred to as the discount factor vector; $G^\infty(\delta)$ is the special case with common discount factor δ.[9]

A pure strategy of player i is a function $s_i : H \to A_i$; a behavioral strategy is a function $\sigma_i : H \to \Delta A_i$. The set of (behavioral) strategies for i is denoted Σ_i, and the set of (behavioral) strategy profiles is denoted Σ. Player i's expected payoff (or payoff, for short) given σ, $U_i(\sigma, \delta_i)$, is defined the usual way. The payoff vector is denoted $U(\sigma, \boldsymbol{\delta})$.

Unless mentioned otherwise, no public randomization device (PRD) is assumed.

2.2. **Blackwell Equilibrium.** Given that monitoring is perfect, the natural solution concept is subgame-perfect Nash equilibrium (SPNE).

Definition 1. A strategy profile $\sigma \in \Sigma$ is a **Blackwell SPNE** (above $\underline{\delta}$) if there exists $\underline{\delta} \in [0, 1)$ such that σ is an SPNE of $G^\infty(\boldsymbol{\delta})$ at any $\boldsymbol{\delta} \geq \underline{\delta} \cdot (1, \ldots, 1)$.

[9] Bold symbols are used for only those vectors whose scalar counterparts are also used, *e.g.*, $\boldsymbol{\delta}$.

A vector $v \in \mathbb{R}^n$ is a **Blackwell SPNE payoff** at $\boldsymbol{\delta}$ if there exists a Blackwell SPNE σ above some $\underline{\delta}$, with $\boldsymbol{\delta} \geq \underline{\delta} \cdot (1, \ldots, 1)$, such that $v = U(\sigma; \boldsymbol{\delta})$.

A more general definition would allow discounting to vary with time (or even with the history). That is, one might consider an evaluation $(\delta_t)_{t=1}^{\infty}$ (a probability distribution over positive integers), where the weight of any round t is given by δ_t (see Renault, 2014). This might be particularly relevant in settings where discounting captures the uncertainty in the length of the interaction. The choice adopted here is made primarily for simplicity. Plainly, enlarging the set of discount sequences that the equilibrium must survive further restricts the set of equilibria; yet, our equilibrium constructions do not rely on the fact that the sequences we focus on are constant (or, for that matter, deterministic). Similarly, we could weaken the criterion without changing any result by requiring optimality to hold only for constant vectors $\boldsymbol{\delta} = (\delta, \ldots, \delta)$, if this is more appropriate in some context.

2.3. **A Necessary Condition.** Subgame-perfection puts little restriction on action profiles specified by an equilibrium, as long as the feasible and individually rational payoff set has non-empty interior. Matters are different under the Blackwell criterion.

The set $\mathcal{A}^{\mathrm{MI}}$ are those mixed action profiles α such that each player gets the same reward from each action in the support of α_i, given α_{-i}. This property is called **Myopic Indifference** (MI). Formally,

$$\mathcal{A}^{\mathrm{MI}} := \left\{ \alpha \in \mathcal{A} \,\middle|\, g_i(a_i, \alpha_{-i}) = g_i(\alpha) \ \forall i \in I, \forall a_i \in \mathrm{supp}(\alpha_i) \right\}. \tag{2.2}$$

To put it differently, α is in $\mathcal{A}^{\mathrm{MI}}$ if, and only if, it is a Nash equilibrium of the stage game $\langle I; (\mathrm{supp}(\alpha_i))_{i \in I}, (g_i)_{i \in I} \rangle$.

The motivation for the definition of $\mathcal{A}^{\mathrm{MI}}$ derives from the following result.

Proposition 1. *If σ is a Blackwell SPNE, then $\sigma(h) \in \mathcal{A}^{\mathrm{MI}}$ for any history $h \in H$.*

Proof. If σ is a Blackwell SPNE, it is an SPNE at all δ in an open interval $\mathcal{O} \subset (0, 1)$. Fix any history h^{t-1}, player $i \in I$, and actions $a_i, a_i' \in \mathrm{supp}(\sigma_i(h^{t-1}))$. Let player i's expected reward in round $\tau > t$ under the continuation strategy $\sigma|_{h^{t-1}}$ following the action a_i (resp., a_i') at t be $g_i^{(\tau)}$ (resp., $g_i'^{(\tau)}$). Since player i mixes over a_i and a_i', they yield the same payoff for any $\delta \in \mathcal{O}$:

$$g_i(a_i, \sigma_{-i}(h^{t-1})) + \sum_{\tau > t} \delta_i^{\tau - t} g_i^{(\tau)} = g_i(a_i', \sigma_{-i}(h^{t-1})) + \sum_{\tau > t} \delta_i^{\tau - t} g_i'^{(\tau)} \ \forall \delta_i \in \mathcal{O},$$

and hence

$$f(\delta_i) := g_i(a_i, \sigma_{-i}(h^{t-1})) - g_i(a_i', \sigma_{-i}(h^{t-1})) + \sum_{\tau>t} \delta_i^{\tau-t}(g_i^{(\tau)} - g_i'^{(\tau)}) = 0 \ \forall \delta_i \in \mathcal{O}. \tag{2.3}$$

The Identity/Uniqueness Theorem (see Ahlfors (1953), p.127) implies that if the set of zeros of an analytic function has an accumulation point in its domain, then it is identically zero; since (2.3) holds for an open interval of δ_i, it follows that f is identically zero in $(-1, 1)$; in particular, setting $\delta_i = 0$ gives:

$$g_i(a_i, \sigma_{-i}(h^{t-1})) = g_i(a_i', \sigma_{-i}(h^{t-1})).$$

Thus, both a_i and a_i' yield the same reward; hence, $\sigma(h^{t-1}) \in \mathcal{A}^{\text{MI}}$. This shows that myopically indifferent action profiles are played after any history. □

The strength of subgame-perfection is not needed for the conclusion: if attention is restricted to histories on path, the same holds for Nash equilibria.

Standard constructions in the literature rely on action profiles that are not in $\mathcal{A}^{\text{MI}}$. More specifically, such action profiles enter in the definition of the minmax payoff, namely

$$\underline{v}_i := \min_{\alpha_{-i} \in \prod_{j \neq i}(\Delta A_j)} \max_{a_i \in A_i} g_i(a_i, \alpha_{-i}). \tag{2.4}$$

To keep a player to this level, the other players may have to randomize over actions over which they are not myopically indifferent. Given Proposition 1, we introduce the following notion of MI-minmax payoff:

$$\underline{v}_i^{\text{MI}} := \min_{\alpha \in \mathcal{A}^{\text{MI}}} \max_{a_i \in A_i} g_i(a_i, \alpha_{-i}). \tag{2.5}$$

Every pure action profile is in $\mathcal{A}^{\text{MI}}$. It follows that

$$\underline{v}_i^{\text{MI}} \leq \underline{v}_i^{\text{pure}} := \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} g_i(a_i, \alpha_{-i}).$$

Similarly, every Nash equilibrium of the stage game is included in $\mathcal{A}^{\text{MI}}$. Hence, it also holds that $\underline{v}_i^{\text{MI}} \leq \underline{v}_i^{\text{NE}}$, where $\underline{v}_i^{\text{NE}}$ is player i's lowest stage-game Nash equilibrium payoff.

The following example shows that the inequalities $\underline{v}_i \leq \underline{v}_i^{\text{MI}} \leq \underline{v}_i^{\text{pure}}$ can be strict.

Example 1. *Consider the payoff matrix given by Figure 1.*

To minmax player 1, player 2 must play $(\frac{1}{2}L + \frac{1}{2}R)$*. However, as* R *is a dominant action for player 2, he cannot be myopically indifferent between* L *and* R*. The MI-minmax of player 1 obtains when player 2 plays* $(\frac{3}{4}L + \frac{1}{4}M)$*, which is not as harsh a*

		Player 2		
		L	M	R
Player 1	T	$(1,0)$	$(0,0)$	$(0,3)$
	B	$(0,0)$	$(3,0)$	$(1,1)$

Figure 1: The game in Example 1.

punishment, but still worse for player 1 than pure minmaxing via (say) R, which is also player 2's action under Nash reversion. Hence, it holds that $\underline{v}_1 = \frac{1}{2} < \underline{v}_1^{\mathrm{MI}} = \frac{3}{4} < \underline{v}_1^{\mathrm{pure}} = \underline{v}_1^{\mathrm{NE}} = 1$.

Proposition 1 immediately implies the following.

Corollary 1. *Every Blackwell equilibrium payoff* v *satisfies* $v_i \geq \underline{v}_i^{\mathrm{MI}}$*, for all* $i \in I$.

2.4. **A "Folk" Theorem.** Under discounting, and subject to a mild dimensionality condition, FM establish a folk theorem for subgame-perfect Nash equilibrium: given any $v \in F$ such that for all i, $v_i > \underline{v}_i$, there exists $\underline{\delta} \in [0,1)$ such that, for any $\delta \in (\underline{\delta}, 1)$ there is a subgame-perfect Nash equilibrium σ of $G(\delta)$ with payoff $U(\sigma, \delta) = v$. Clearly, the same cannot hold under the Blackwell criterion, given Proposition 1.

Define

$$F^{\mathrm{MI}} := \{v \in F \mid v_i > \underline{v}_i^{\mathrm{MI}}, \quad \forall i \in I\}. \tag{2.6}$$

Whenever F^{MI} is full-dimensional, Proposition 1 implies that (the closure of) F^{MI} is an upper bound on the set of Blackwell SPNE payoff vectors. The following theorem shows that this upper bound is tight in general.[10]

Theorem 1. *Suppose that the dimension of* F^{MI} *is* n.[11] *For any* $v \in F^{\mathrm{MI}}$*, there exists* $\underline{\delta} < 1$ *such that for all* $\delta \in (\underline{\delta}, 1)$*,* v *is a Blackwell SPNE payoff at* δ.

The proof, which appears in Appendix A, follows FM (see also Abreu, 1988) in having stick-and-carrot punishment regimes, one for each player. Any unilateral

[10]Note that the statement of Theorem 1 refers to the equilibrium payoff vector evaluated at a common discount rate (yet, the strategy profile must be optimal for all possibly distinct discount factors high enough). This is because, as is well known (see Lehrer and Pauzner, 1999), the set of feasible payoffs evaluated at different discount factors can be larger than the convex hull of the stage-game payoffs, a topic that is orthogonal to our purpose.

[11]Unlike in FM, the full dimensionality assumption in Theorem 1 cannot be dropped for the case $n = 2$ in general. Presumably, the general case (without interiority assumption) can be dealt with by adapting the notion of effective minmax (Wen, 1994) to account for the constraint that action profiles must be in $\mathcal{A}^{\mathrm{MI}}$, along the lines of Fudenberg, Levine, and Takahashi (2007).

deviation from the prescribed strategies leads to a "stick-and-carrot regime." In FM, the stick phase involves minmaxing, during which the player who deviated is held to his minmax payoff. Play then moves to the carrot phase, in which all players earn strictly more than their minmax payoff.

In the case in which the target payoff v is achieved by a pure action profile, our construction is a straightforward adaptation of this construction. During the stick phase of player i, replace standard minmaxing with an action profile $\alpha^i \in \arg\min_{\alpha \in \mathcal{A}^{\text{MI}}} \max_{a_i \in A_i} g_i(a_i, \alpha_{-i})$. For each player $j \neq i$, actions within the support α^i_j yield j the same reward, and the selected one within it is subsequently ignored, whereas actions outside of the support are deterred as in FM.

What if v is not the payoff of a pure action profile? Lacking a PRD, we follow Dasgupta and Ghosh (2021) to construct action profile paths that deliver the target payoff while also keeping continuation payoffs near the target. We then show that if continuation payoffs given a pure action path and a certain discount remain bounded above and below, those same bounds apply at larger discount factors.[12] This ensures that continuation payoffs are similar enough to be enforced by the same punishments for all high enough discounts.

3. Imperfect Public Monitoring

3.1. **Generic Games.** This section turns to imperfect public monitoring, starting with the same finite set of players $I = \{1, 2, ..., n\}$ and finite sets of actions A_i, $i \in I$, with reward function $g_i : A \to \mathbb{R}$. A monitoring structure (Y, π) is a finite set of signals Y and a function $\pi \colon A \to \Delta Y$ mapping action profiles $a \in A$ into distributions over Y, indicating the probability that each signal $y \in Y$ is publicly observed. Let $G = \langle I; A, g; Y, \pi \rangle$. Given discount factor vector $\boldsymbol{\delta}$, we denote the infinitely repeated game by $G^\infty(\boldsymbol{\delta})$.

For each player i, a private history of length t, h_i^t, is a sequence $(a_i^{(1)}, y^{(1)}, \ldots, a_i^{(t)}, y^{(t)}) \in H_i^t := (A_i \times Y)^t$, and the set of i's private histories is H_i. A public history h^t is a sequence $(y^{(1)}, \ldots, y^{(t)}) \in H^t := Y^t$, with the set of all public histories denoted H. A behavior strategy $\sigma_i \in \Sigma_i$ maps private histories to i's mixed actions: $\sigma_i : H_i \to \Delta A_i$. It is public if, it is measurable with respect to H. We adopt perfect public equilibrium

[12] This is reminiscent of the Arrow-Levhari (1969) stopping theorem: if the value of a discardable security is weakly positive when evaluated at a certain discount (given optimal discarding), it is weakly positive at greater discounts. The same intuition applies here: any short-term setbacks and gains are smoothed out at higher discount factors. Therefore, continuation payoffs remain nearby at all higher discount factors, and hence deviations are deterred by the same punishments.

(PPE) as our solution concept: a strategy profile σ is a PPE if, for all i, σ_i is public, and for all public histories h^t, $\sigma|_{h^t}$ is a Nash equilibrium of the infinitely repeated game (under some payoff criterion).

Definition 1 is extended the obvious way. A strategy profile $\sigma \in \Sigma$ is a **Blackwell PPE** (above $\underline{\delta}$) if there exists $\underline{\delta} \in [0,1)$ such that σ is a PPE of $G^\infty(\boldsymbol{\delta})$ at any $\boldsymbol{\delta} \geq \underline{\delta} \cdot (1,\ldots,1)$. A vector $v \in \mathbb{R}^n$ is a **Blackwell PPE payoff** at $\boldsymbol{\delta}$ if there exists a Blackwell PPE σ above some $\underline{\delta}$, with $\boldsymbol{\delta} \geq \underline{\delta} \cdot (1,\ldots,1)$, such that $v = U(\sigma;\boldsymbol{\delta})$, where as before $U(\sigma;\boldsymbol{\delta})$ is the equilibrium payoff vector under σ given $\boldsymbol{\delta}$.

The next proposition shows that under imperfect public monitoring, the Blackwell criterion more severely restricts the set of action profiles that can be played in equilibrium. It assumes there is no PRD. Its proof, as well as the proofs of other results in this section, appear in Appendix B.

Proposition 2. *Fix I, A, and Y. For almost all (g,π), given any Blackwell PPE σ, $\sigma(h^t)$ is either a pure action profile or a stage-game Nash equilibrium, for all t, $h^t \in H^t$.*

That is, the set of reward functions and monitoring structures such that, in some Blackwell equilibrium, after some public history, players choose an action profile that is not pure or a stage-game Nash equilibrium, has measure zero. Other action profiles, even those satisfying myopic indifference, cannot arise. The selected action can no longer be simply ignored. To understand why, suppose that, after some history, some player, say i, is playing a (nondegenerate) mixed action α_i satisfying myopic indifference, yet some player j (perhaps i himself) is not playing a best-reply to α_{-j}. By definition, actions $a_i, a_i' \in \operatorname{supp}\alpha_i$ yield the same reward. However, they induce different distributions over public signals, in general. Since j isn't playing a best-reply in the stage game, the continuation strategy profile must depend on the public signal. This typically affects player i's continuation payoff, and hence, his preference between a_i and a_i' in the repeated game.

Proof. The proof is divided into three parts; we define a non-generic set of reward functions, then a non-generic set of signal distributions, and show that, for any (g,π) outside of this set, a Blackwell PPE specifies a pure action or Nash profile at every history.

Generically, a finite game possesses finitely many Nash equilibria (Harsanyi, 1973a). Because $\mathcal{A}^{\mathrm{MI}}$ is the union of sets of Nash equilibria over finitely many games (defined by the possible subsets of actions), there exists a subset $\mathcal{G} \subset \mathbb{R}^{I\times A}$ of measure zero

such that, for any $g \in \mathbb{R}^{I\times A} \setminus \mathcal{G}$, the set $\mathcal{A}_g^{\mathrm{MI}}$ (the subscript referring to the reward) is finite and has distinct payoffs, i.e. $\alpha, \alpha' \in \mathcal{A}_g^{\mathrm{MI}}$ implies that for all i, $g_i(\alpha) \neq g_i(\alpha')$.

Fix $g \in \mathbb{R}^{I\times A} \setminus \mathcal{G}$, a period $T > 1$, and a T-period strategy profile $\sigma^T \colon U_{t=0}^{T-1} H^t \to \mathcal{A}_g^{\mathrm{MI}}$ in which period-T play varies with the first-round signal, *i.e.*,

$$\exists y_1, y_1', y_2, \ldots, y_{T-1} \in Y \text{ s.t. } \sigma^T(y_1, y_2, \ldots, y_{T-1}) \neq \sigma^T(y_1', y_2, \ldots, y_{T-1}). \tag{3.1}$$

For each $i \in I$, $a_i', a_i'' \in A_i$ with $a_i' \neq a_i''$ let

$$\Pi_g^{i,T,a_i',a_i'',\sigma^T} := \left\{ \pi \in (\Delta Y)^A \;\middle|\; \mathbb{E}[v_i^T \mid a_i', \sigma^T] = \mathbb{E}[v_i^T \mid a_i'', \sigma^T] \right\}, \tag{3.2}$$

where $v_i^T = g_i \circ \sigma^T \mid_{Y^{T-1}}$ is player i's reward in round T. Both sides of the equality in (3.2) are polynomials of $\{\pi(y \mid a) \mid y \in Y \setminus \{y_0\}, a \in A\}$, where y_0 is an arbitrary signal (we omit $y_0 \in Y$ because the probability distribution $\{\pi(y \mid a) \mid y \in Y\}$ adds up to one). Since g features distinct rewards, player i's last-round reward varies with the first-round signal, and hence the polynomials are distinct. As the set of zeros of a non-zero polynomial, the set $\Pi_g^{i,T,a_i',a_i'',\sigma^T}$ is of measure zero (Caron and Traynor, 2005; Neeb, 2011). Define

$$\Pi_g := \bigcup_{i,T,a_i',a_i'',\sigma^T} \Pi_g^{i,T,a_i',a_i'',\sigma^T},$$

where $a_i', a_i'' \in A_i$, $a_i' \neq a_i''$ and σ^T varies with the first-round signal. T runs over a countably infinite set, for each element of which σ^T runs over a finite set because $g \notin \mathcal{G}$ ensures that the range $\mathcal{A}_g^{\mathrm{MI}}$ of σ^T is finite. Therefore, Π_g has measure zero.

Consider a game G with $g \notin \mathcal{G}$. Let σ be a Blackwell equilibrium that prescribes an action profile that is neither pure nor a stage-game Nash equilibrium after some public history; without loss of generality, after h^0. By hypothesis, there is a player $i \in I$ who mixes over (at least) two distinct actions a_i', a_i'', and a player j who does not use a stage-best-reply to $\sigma_{-i}(h^0)$. Then there exists an earliest round $T^* \in \mathbb{N}$ in which the action profile depends on the first-round signal; otherwise, j has a profitable deviation in the initial round. Since i mixes at the initial history, the payoff conditional on playing a_i' is equal to the payoff following a_i'', for all δ in an open interval $\mathcal{O} \subset (0,1)$; as in Proposition 1, it implies that the payoff at each round t is the same. In particular, applying this to T^*, and denoting by σ^{T^*} the T^*-period truncation of σ, this implies that $\pi \in \Pi_g^{i,T^*,a_i',a_i'',\sigma^{T^*}} \subset \Pi_g$. □

Again, an immediate implication of Proposition 2 is a lower bound on equilibrium payoffs. Recall that $\underline{v}_i^{\mathrm{pure}}$ is i's worst Nash payoff, and $\underline{v}_i^{\mathrm{NE}}$ is i's pure minmax.

Corollary 2. *Fix A, I and Y. For almost all (g, π), every Blackwell equilibrium payoff v satisfies $v_i \geq \min\{\underline{v}_i^{\text{pure}}, \underline{v}_i^{NE}\}$, for all $i \in I$.*

3.2. **A "Folk" Theorem.** Given Proposition 2 and Corollary 2, a folk theorem under the Blackwell criterion must involve a smaller payoff set, and stronger assumptions than those imposed by FLM. First, player i's equilibrium payoff is bounded below by his pure minmax, or his lowest Nash equilibrium payoff, whichever is lower.

Second, in general, mixed actions can help detect deviations, or discriminate among them. Hence, the identifiability assumptions must be strengthened.

Definition 2. The monitoring structure (Y, π) satisfies **pairwise full rank** for a profile α if for all $i, j \in I$ with $i \neq j$, the matrix $\Pi_{i,j}(a)$ whose rows are $\{\pi^\top(\cdot \mid a_i', \alpha_{-i}) \mid a_i' \in A_i\} \cup \{\pi^\top(\cdot \mid a_j', \alpha_{-j}) \mid a_j' \in A_j\}$ has rank $|A_i| + |A_j| - 1$.

Denote

$$F^* := \{v \in F \mid v_i \geq \min\{\underline{v}_i^{\text{pure}}, \underline{v}_i^{\text{NE}}\} \ \forall i \in I\}.$$

Theorem 2. *Fix I, A and Y. Suppose that π satisfies pairwise full rank for all pure action profiles. Suppose also that there exists $a^* \in A$, $Y^* \subset Y$ such that $\pi(Y^* \mid a_{-i}^*, a_i) < \pi(Y^* \mid a^*) < 1$, $\forall i \in I$, $a_i \neq a_i^*$. Then for any $v \in \operatorname{int} F^*$, there exists $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$, v is a Blackwell PPE payoff at δ.*

The assumption that a pair (a^*, Y^*) as stated in the theorem exists is technical; while it is relatively mild, it is needed in the proof, and we do not know whether some version of it is necessary for the result. It allows the players to emulate a PRD (and could be dispensed if a PRD was assumed).[13]

Proving the theorem involves several steps. It is more instructive to explain some of them in the special case in which monitoring takes a product structure, and assuming a PRD. Hence, the proof of Theorem 2 should be read after the proof of Theorem 3 and the discussion at the end of this section.

3.3. **Public Randomization.** If we now assume a PRD (a uniform draw from the unit interval), in the special case in which monitoring has a product structure, some non-Nash myopically indifferent mixed action profiles may be used. Hence, the Blackwell payoff set may exceed that of Theorem 2. We say that (Y, π) has a **product**

[13] That is, the event $\{y \mid y \in Y^*\}$ is a public, binary random variable, whose likelihood is maximized if players use a^*. Hence, it suffices to make its occurrence desirable to ensure that players are willing to generate that signal.

structure if

$$Y = \prod_i Y_i \text{ and } \pi(y \mid a) = \prod_i \pi_i(y_i \mid a_i),$$

where $\pi_i(\cdot \mid a)$ is the marginal distribution of $\pi(\cdot \mid a)$ on Y_i.

The relevant lower bound on player i's equilibrium payoff is given by the solution to the following program.

$$(P_i^{\mathrm{MI},\pi_i}): \qquad \min_{\alpha \in \mathcal{A}^{\mathrm{MI}}, x_i \colon Y_i \to \mathbb{R}} \left\{ g_i(\alpha) + \sum_{y_i \in Y_i} \pi_i(y_i \mid \alpha_i) x_i(y_i) \right\}$$

subject to

$$g_i(\alpha) + \sum_{y_i \in Y_i} \pi_i(y_i \mid \alpha_i) x_i(y_i) \geq g_i(a_i, \alpha_{-i}) + \sum_{y_i \in Y_i} \pi_i(y_i \mid a_i) x_i(y_i) \quad \forall a_i \in A_i, \tag{3.3}$$

$$x_i(y_i) \geq 0 \quad \forall y_i \in Y_i. \tag{3.4}$$

Let $\underline{v}_i^{\mathrm{MI},\pi_i}$ denote the minimum. We note that, since it is feasible to pick a stage-game Nash equilibrium for α, and to set $x_i(\cdot) = 0$, $\underline{v}_i^{\mathrm{MI},\pi_i} \leq \underline{v}_i^{\mathrm{NE}}$. Also, since $\mathcal{A}^{\mathrm{MI}}$ can be strictly larger than A, it is easy to find games such that $\underline{v}_i^{\mathrm{MI},\pi_i} < \min\{\underline{v}_i^{\mathrm{pure}}, \underline{v}_i^{\mathrm{NE}}\}$.

This program is nothing but the "scoring algorithm" (in the direction that minimizes i's payoff) introduced by Fudenberg and Levine (1994), with the restriction that players $-i$ are constrained to choose from $\mathcal{A}^{\mathrm{MI}}$. Indeed, this constraint must be satisfied in a Blackwell equilibrium (given that this is already the case under perfect monitoring, see Proposition 1, this should come as no surprise). It immediately follows that $\underline{v}_i^{\mathrm{MI},\pi_i}$ is a lower bound on i's equilibrium payoff.

For this bound to be tight, a rank assumption is needed, for which we follow FLM.

Definition 3. A profile α satisfies **individual full rank** (IFR) if for all i the vectors $\{\pi(a_i, \alpha_{-i}) \mid a_i \in A_i\}$ are linearly independent.

Let

$$F^{\mathrm{MI},\pi} := \{v \in \mathrm{co}(u(A)) \mid v_i \geq \underline{v}_i^{\mathrm{MI},\pi_i} \, \forall i \in I\}.$$

The characterization is the following.

Theorem 3. *Assume a PRD. Suppose (Y, π) has a product structure. Every Blackwell PPE payoff vector v satisfies $v_i \geq \underline{v}_i^{\mathrm{MI},\pi_i}$ for all $i \in I$.*
Conversely, if π_i satisfies IFR for all i, then for any $v \in \operatorname{int} F^{\mathrm{MI},\pi}$, there exists $\underline{\delta} < 1$ such that for all $\delta \in (\underline{\delta}, 1)$, v is a Blackwell PPE payoff vector at δ.

The proof of the necessity part matches the proof of Proposition 1. The sufficiency part of the theorem has a two-step proof:

1. We show that at *some* $\delta_0 < 1$ there is a *robust equilibrium*, *i.e.*, a strategy profile that is a PPE in a neighborhood of δ_0 with payoff vector v at δ_0.
2. We use the PRD to periodically restart (or "reboot") the game, discarding the history up to that point, which allows us to lower the discount factor at which incentive compatibility must be checked.[14] That is, if σ is a PPE at some $\delta_0 \in (0,1)$, we can construct a related equilibrium at higher discount factors by rebooting appropriately to reduce the effective discount factor to δ_0.

3.3.1. *Rebooting.* We begin with the second step, *rebooting*. Fix σ and $p \in (0,1)$. Let σ^p denote a strategy profile that follows σ but reboots the game with probability p at the end of each round, independently across rounds; *i.e.*, if at the end of some round the value of the PRD is less than p, we discard the history and restart playing σ. More precisely, let $\omega_1, \omega_2, \ldots$ be the sequence of PRD draws; then, given $(y_1, \omega_1, \ldots, y_t, \omega_t)$, we let

$$\sigma^p(y_1, \omega_1, y_2, \omega_2, \ldots, y_t, \omega_t) := \sigma(y_{\tau+1}, \ldots, y_t), \text{ where } \tau := \max\{s \leq t \mid \omega_s \leq p\},$$

with the convention that $\max \emptyset = 0$.

If a player has a discount factor δ, the payoff stream from σ^p is evaluated at discount δ; however, a simple calculation shows that a player's incentive to deviate from σ^p at discount δ is the same as his incentive to deviate from σ at discount $\delta(1-p)$.

This definition and the preceding discussion imply the following lemma, which essentially reduces a global robustness problem to a local one.

Lemma 1 (Reboot Lemma)**.** *If σ is a PPE for all discount factors in some interval $(\delta_0 - \Delta, \delta_0 + \Delta) \subset [0,1)$, for some $\Delta > 0$, then σ^p is a Blackwell PPE above $\frac{\delta_0 - \Delta}{1-p}$ for $p \in (0,1)$ such that $\frac{\delta_0 - \Delta}{1-p} < 1 < \frac{\delta_0 + \Delta}{1-p}$, and*

$$U(\sigma^p, \delta_0/(1-p)) = U(\sigma, \delta_0).$$

3.3.2. *Constructing a Robust Equilibrium.* Given the Reboot Lemma, existence of the desired Blackwell equilibrium follows from the construction of a robust equilibrium. Existence of a robust equilibrium will be demonstrated by adapting arguments from Abreu, Pearce, and Stacchetti (1990) (henceforth APS 1990). Robustness requires incentives to hold for a range of discount factors; this motivates a stronger notion

[14] This is somewhat in the spirit of what is known as "Ellison's trick" (Ellison, 1994).

of self-generation than proposed in APS 1990. We provide some intuition for the two ways in which our definition needs to be stronger. The crux is that varying the discount factor, however slightly, could affect incentives.

If the current action profile is pure, changing the discount factor may reverse weak incentives to not deviate. This can be tackled by giving strict incentives (a slack of at least $(1-\delta)\eta$ in the definition below) not to deviate.

The problem is subtler when mixed actions are needed, such as when delivering $\underline{v}_i^{\mathrm{MI},\pi_i}$ to player i. Suppose α, x solve Program P_i^{MI,π_i}. Since $\alpha \in \mathcal{A}^{\mathrm{MI}}$, the current payoff $g_i(\alpha)$ is constant on $\mathrm{supp}(\alpha_i)$. Distinct actions in the support of α_i could generate different distributions over various continuation payoff vectors w, but with the same expected value for the i-th component w_i. But even if these actions induce the same distribution over w_i at δ, if they induce different distributions over continuation payoff vectors, they could induce different distributions over action paths. These could give different expected continuation payoffs for i at discount factors even slightly away from δ. To circumvent this, we need to ensure that any two such actions induce the same distribution over continuation payoff vectors, and hence the same probability distribution over action paths.

To this end we use the randomization device - a uniform draw (denoted ν) from $[0,1]^n$ - to "garble" the distribution of continuation payoff vectors.

Definition 4. For any $\eta > 0$, $W \subset \mathbb{R}^n$, and $\delta \in (0,1)$, the set $\mathcal{B}_\eta(W;\delta)$ comprises points $v \in \mathbb{R}^n$ such that $v = (1-\delta)g(\alpha) + \delta\,\mathbb{E}(w \mid \alpha)$ for some continuation payoff function $w : Y \times [0,1]^n \to \mathbb{R}^n$ taking finitely many values, and a current action profile α that is a NE of the normal-form game with payoffs $(1-\delta)g(a)+\delta\,\mathbb{E}(w \mid a)$, under the additional condition that for any $i \in I$, for any $a_i' \in A_i$, at least one of the following is true:

$$w|\alpha \overset{d}{=} w|a_i', \alpha_{-i} \tag{3.5}$$

$$v_i \geq (1-\delta)g_i(a_i', \alpha_{-i}) + \delta\mathbb{E}_\nu \sum_{y \in Y} \pi(y \mid a_i', \alpha_{-i}) w_i(y,\nu) + (1-\delta)\eta. \tag{3.6}$$

For any $\eta > 0$, the mapping $\mathcal{B}_\eta : 2^{\mathbb{R}^n} \times [0,1) \to 2^{\mathbb{R}^n}$ is called an '**η-strong APS mapping**'.

The slack in (3.6) does depend on the discount factor. Alternatively, we can view it as a slack of η in un-normalized or total payoffs. Our definition requests that all pure actions on the support of α_i output the same distribution of continuation payoffs as in (3.5). Moreover, a deviating pure action a_i off the support of α_i either

(a) satisfies (3.5) so that it leads to the same distribution over continuations, and is therefore unprofitable at any discount given that α is a Nash of the auxiliary game; or (b) satisfies (3.6) and therefore entails a loss of at least $(1-\delta)\eta$, which will allow us to show it is unprofitable at discount factors in some neighborhood of δ. Our strengthened notion of self-generation follows.

Definition 5. For $\eta > 0$, a set $W \subset \mathbb{R}^n$ is said to be **η-strong self-generating** at δ if $W \subset \mathcal{B}_\eta(W;\delta)$ for a strong APS mapping $\mathcal{B}_\eta$.

At this point the standard approach shows that any smooth set in the interior of the feasible and individually rational set is self-generating. Our proof differs in three ways from this. First, we use the notion of η-strong self-generation, to leave "wiggle room" for varying discounting and achieving robustness. Second, we show this property for closed balls, rather than directly for all smooth sets; this is analytically more tractable. Points in the interior of a ball are generated by playing a Nash equilibrium of the stage game; the required continuations are in the ball for high enough discount factors. Boundary points are harder –if the action profile chosen provides weak incentives, we "mix in" a small probability of continuation payoffs that provide strict incentives; these exist by individual full rank. Third, to satisfy condition (3.5) we modify each player's continuation payoffs to take only the two extreme values, the only variation being the probabilities with which these two values are chosen for various $y \in Y$; these stochastic continuation values require a public randomization device to carry out "garbling."

Following FLM 1994, we prove a local self-generation property, and then leverage this property to prove we can construct robust equilibria.

Lemma 2. *If π satisfies IFR, for every $c \in F^{\mathrm{MI},\pi}$ and $r > 0$, if the closed ball $B(c,r) \subset \mathrm{int}(F^{\mathrm{MI},\pi})$, then there is a $\underline{\delta} < 1$ and a $\eta > 0$, such that at each $\delta \in (\underline{\delta}, 1)$, $B(c,r)$ is η-strongly self-generating.*

The proof of the next lemma is immediate from Lemma 2, the continuity of payoffs in δ, and familiar self-generation arguments à la APS 1990 and FLM 1994.

Lemma 3 (Robust Equilibrium Exists)**.** *With a product monitoring structure and a PRD, if π satisfies IFR, for any compact $W \subset \mathrm{int}(F^{\mathrm{MI},\pi})$ there exists a $\underline{\delta} \in (0,1)$, a $\eta > 0$ and a $X \supset W$ such that for all $\delta \in (\underline{\delta}, 1)$ we have $X \subset \mathcal{B}_\eta(X;\delta)$. Moreover,for any $\delta \in (\underline{\delta}, 1)$ there is a strategy profile σ such that $v = U(\sigma;\delta)$ and σ is a PPE at all $\delta' \in (\delta - \Delta, \delta + \Delta)$ for some $\Delta > 0$.*

Theorem 3 follows from Lemmata 1 and 3. We also note that Lemma 3 is essentially a locally robust PPE folk theorem.

3.3.3. *From here to Theorem 2.* The full proof is contained in Appendix B, to which this section serves as a roadmap. Once again, we use a suitable version of self-generation to create robust equilibria at discount factors away from unity, and then use rebooting to create a Blackwell equilibrium out of these. Generating a robust equilibrium involves the same steps as before. Complications arise from having to emulate a PRD through the public signal. Our assumption on the existence of a suitable (a^*, Y^*) guarantees that, with suitable rewards, all players can be induced to maximize the probability of Y^*. "Test phases" during which players play a^* are therefore introduced to generate randomness for resets. Several difficulties arise; we tackle these in turn.

First, unlike the exogenous PRD of Theorem 3, our tests are open to manipulation. In that theorem, some players may desire a reset, others not; but they are not given a choice. Here, we need to provide incentives to players to choose the test action. To offer rewards and yet still wipe the history, we use two self-generating payoff sets, a "punishment" and a "reward" one; every payoff vector in the former is Pareto-dominated by every payoff vector in the latter. By maximizing the probability of Y^* during a test phase, players maximize the probability of getting to –or staying in– the 'reward' payoff set.

Second, to incentivize players as above, "normal" phase rounds –during which punishments and rewards are dished out– must be far more common than "test" phase rounds. The Reboot Lemma only applies when there is exactly one test between two normal-phase rounds linked by incentives –but we necessarily have more normal phase rounds between punishment phases. To fix this, we break up play into "cycles." Thus if we have N rounds in the normal phase and $T+1$ rounds of testing, rounds within a cycle are separated by $T+N+1$ rounds. Rounds that form part of the same normal phase are each linked to a different cycle, with continuation along that cycle coming $T+N+1$ rounds later. In other words, we play N copies of the strategies given by self-generation, punctuated by test phases of length $T+1$ each. The outcome (pass vs. fail) of each test phase affects every cycle identically.

Third, since the test is constructed to meet the need for a certain passing probability, it may give payoffs different than those we wish to deliver in equilibrium. As T is given by this need, we keep it fixed and raise both N and the permissible discount factor so that overall payoffs are largely due to normal phase play. Then, payoffs from

the normal phase are adjusted to take into account expected payoffs during the test phase so as to deliver an overall expected payoff equal to the target payoff.

Fourth, at the same time, the original robust equilibrium's incentives should apply; so $\delta^{T+N+1}(1-p)$, where p is the reset probability, must lie in the robust range.

4. Imperfect Private Monitoring

This section turns to imperfect private monitoring. Again, we are given a finite set of players $I = \{1, 2, ..., n\}$, for each $i \in I$ a finite set of actions A_i and a reward function $g_i : A \to \mathbb{R}$. A (private) monitoring structure is a pair (Y, π), with $Y = \prod_{i \in I} Y_i$, finite, and $\pi \colon A \to \Delta Y$ mapping $a \in A$ into the probability that the signal profile $(y_1, \ldots, y_n) \in Y$ obtains. Player i only observes y_i. Let $G = \langle I; A, g; Y, \pi \rangle$. Given discount factor vector $\boldsymbol{\delta}$, we denote the infinitely repeated game by $G^\infty(\boldsymbol{\delta})$.

A t-length private history h_i^t is a sequence $(a_i^{(1)}, y_i^{(1)}, \ldots, a_i^{(t)}, y_i^{(t)}) \in H_i^t$. The set of all private histories for i is denoted H_i. A behavior strategy $\sigma_i \in \Sigma_i$ maps private histories to mixed actions, $\sigma : H_i \to \Delta A_i$. We follow the literature by adopting sequential equilibrium as solution concept.

Definition 1 is extended the obvious way. A strategy profile $\sigma \in \Sigma$ is a Blackwell equilibrium (above $\underline{\delta}$) if there exists $\underline{\delta} \in [0, 1)$ such that σ is a (sequential) equilibrium of $G^\infty(\boldsymbol{\delta})$ at any $\boldsymbol{\delta} \geq \underline{\delta} \cdot (1, \ldots, 1)$. A vector $v \in \mathbb{R}^n$ is a Blackwell equilibrium payoff at $\boldsymbol{\delta}$ if there exists a Blackwell equilibrium σ above some $\underline{\delta}$, with $\boldsymbol{\delta} \geq \underline{\delta} \cdot (1, \ldots, 1)$, such that $v = U(\sigma; \boldsymbol{\delta})$, where as before $U(\sigma; \boldsymbol{\delta})$ is the equilibrium payoff vector under σ given $\boldsymbol{\delta}$.

Our focus is on games in which $\mathcal{A}^{\mathrm{MI}} = A$. An important property of such games is the following.

Lemma 4. *If $\mathcal{A}^{MI} = A$, the stage game has a unique and pure Nash equilibrium.*

Proof. Since $\mathcal{A}^{\mathrm{MI}} = A$, no player has a tie against any pure action profile of his opponents. If there were $a_i \neq a_i'$ and a_{-i} such that $g_i(a_i, a_{-i}) = g_i(a_i', a_{-i})$, then $(\frac{1}{2}a_i + \frac{1}{2}a_i', a_{-i})$ would belong to $\mathcal{A}^{\mathrm{MI}}$. Thus, every stage-game Nash equilibrium is strict, hence pure. By the index theorem, such an equilibrium is unique. □

The prisoner's dilemma satisfies $\mathcal{A}^{\mathrm{MI}} = A$. So does the product choice game, and 2x2 dominance solvable games. More generally, $\mathcal{A}^{\mathrm{MI}} = A$ if there is no tie against any pure action profile of the opponents, and for each player i, the ordinal ranking over i's actions is independent of $(a_{i+1}, \ldots, a_n)$. The converse of Lemma 4 does not

hold: there exist games with a unique and pure Nash equilibrium, yet $\mathcal{A}^{\mathrm{MI}} \neq A$. (For instance, dominance solvability is not enough in general.)

We focus on a special class of monitoring structures. Let $\pi_i(y_i \mid a) := \sum_{y_{-i}} \pi(y \mid a)$.

Definition 6. A monitoring structure (Y, π) is **conditionally independent** if $\pi(y \mid a) = \prod_i \pi_i(y_i \mid a)$ for all $a \in A$, $y \in Y$.

Conditional independence is a special and admittedly non-generic property. Yet, it plays an important role in the literature. In particular, in the case of the prisoner's dilemma, Matsushima (2004) establishes a folk theorem under conditional independence.

The monitoring structure (Y, π) has full support if $\pi_i(y_i \mid a) > 0$ for all $i \in I$, $y_i \in Y_i$, $a \in A$.

4.1. **An "Anti-Folk" Theorem.**

Theorem 4. *Suppose that $\mathcal{A}^{MI} = A$, and that (Y, π) is conditionally independent and has full support. The unique Blackwell equilibrium outcome is the repetition of the stage-game Nash equilibrium.*

Proof. Let σ be a Blackwell equilibrium. First, we show that σ is pure and history-independent on the equilibrium path. Let t be the first round at which σ prescribes either mixed or history-dependent actions, if it exists. Since players play pure and history-independent actions until round $t-1$, the conditional independence of the monitoring structure implies the independence of players' private histories at the beginning of round t, (H_t, p_t) with $p_t = p_{1t} \times \cdots \times p_{nt}$. Note that each player i is indifferent among all continuation strategies against the opponents' continuation strategies. By an argument similar to that in Proposition 1 (the Identity/Uniqueness Theorem), player i is indifferent among all actions played with positive probability at round t against the opponents' actions in the same round. Let $\overline{\alpha}_i(a_i) = \sum_{h_i^t} p_{it}(h_{it}) \sigma_i(h_i^t)(a_i)$ for each i and a_i. Then $(\overline{\alpha}_1, \ldots, \overline{\alpha}_n) \in \mathcal{A}^{\mathrm{MI}}$. This contradicts $\mathcal{A}^{\mathrm{MI}} = A$.

Since the monitoring structure has full support, there is no reason to play a sub-optimal action against the opponents' history-independent strategies. Thus, players play the unique stage-game Nash equilibrium on the equilibrium path. □

The intuition for this result relies on Matsushima (1991). Indeed, the first step of the proof of Theorem 4 follows his. He shows inductively that a pure strategy satisfying "independence of irrelevant information" must be history-independent. Hence, non-trivial equilibria must involve some indifference across some actions, for some

	H	D
H	$0,0$	$5,1$
D	$1,5$	$4,4$

Figure 2: The game in Example 2.

player, after some private history. This is inconsistent with $\mathcal{A}^{\mathrm{MI}} = A$, given that σ must be a Blackwell equilibrium.

If $\mathcal{A}^{\mathrm{MI}} \neq A$, *i.e.*, there exists $\alpha \in \mathcal{A}^{\mathrm{MI}} \setminus A$, then players may play α at, say, round 1, so that independence of private histories fails. Even if players have played pure actions so far, they can play possibly history-dependent actions at round t so long as their "averages" are equal to α_i. By using history dependence appropriately, one can engineer non-myopic equilibrium behavior at earlier rounds, see the following example.

Example 2. *Consider the repetition of the hawk-dove game given by the payoff matrix in Figure 2. The conditionally independent monitoring structure (Y, π) is given by $Y_1 = Y_2 = \{h, d\}$, and*

$$\pi_i(y_i = h \mid a_j = H) = \pi_i(y_i = d \mid a_j = D) = 0.9.$$

The stage game has three Nash equilibria: (H, D), (D, H), and $(\frac{1}{2}H + \frac{1}{2}D, \frac{1}{2}H + \frac{1}{2}D)$, hence $\mathcal{A}^{MI} \neq A$. Consider the following symmetric strategy profile: at every odd round, play D, and at every even round, play H if one's own previous action is D and the signal is h, play $\frac{5}{9}D + \frac{4}{9}H$ if one's own previous action is D and the signal is d, and play D if one's own previous action is H (off the equilibrium path). Note that at an even round on the equilibrium path, each player plays

$$0.1 \times H + 0.9 \times \left(\frac{5}{9}D + \frac{4}{9}H\right) = \frac{1}{2}H + \frac{1}{2}D$$

on average, and no player has an incentive to deviate. If a player deviates to H at an odd round, he receives a reward of 5 in that round, and at the next round, faces $0.9 \times H + 0.1 \times \left(\frac{5}{9}D + \frac{4}{9}H\right) = \frac{17}{18}H + \frac{1}{18}D$ on average, and so receives a payoff of $\frac{21}{18}$. Since

$$4 + \delta \times \frac{5}{2} \geq 5 + \delta \times \frac{21}{18} \Leftrightarrow \delta \geq \frac{18}{23},$$

this strategy profile is a Blackwell equilibrium.

5. Extensions

5.1. **Discount-free Equilibrium under Perfect Monitoring.** Provided a player's intertemporal choice satisfies classic axioms (Koopmans, 1960), discounted utility is a representation of his preferences, and "not knowing" them is meaningless. Hence, for applications in which discounting is part of such a representation, it is perhaps more natural to assume that player i "knows" his own discount factor δ_i, but is unsure about δ_{-i}.[15]

To account for this, we introduce the notion of discount-free equilibrium. To keep the discussion short, we restrict ourselves to the case of perfect monitoring. Definitions are as in Section 2.

Definition 7. A **discount-free equilibrium** above $\underline{\delta} \in (0,1)$ is a vector $(\sigma_i)_{i=1}^n$, with $\sigma_i : [\underline{\delta}, 1) \to \Sigma_i$, such that $(\sigma_i(\delta_i))_{i=1}^n$ is a SPNE of $G^\infty((\delta_1, \ldots, \delta_n))$, for all $(\delta_1, \ldots, \delta_n)$ such that $\delta_i \geq \underline{\delta}$, for all i. Its payoff at $\underline{\delta}$ is the payoff of $(\sigma_i(\underline{\delta}))_{i=1}^n$ in $G^\infty(\underline{\delta})$.

Theorem 5 states that knowing one's own discount factor is enough to restore the "standard" folk theorem, as stated in FM.[16] Specifically, we consider the repetition of the following extensive-form.

1. A PRD obtains (a uniform draw of the unit interval);[17]
2. The simultaneous-move game G is played;
3. Players publicly and simultaneously announce an element in $[0,1)$.

Theorem 5. *Fix $v \in \operatorname{int} F$ with $v_i > \underline{v}_i$, for all i. There exists $\underline{\delta} < 1$ such that, for all $\delta \in (\underline{\delta}, 1)$, there exists a discount-free equilibrium above δ with payoff v at $\delta \cdot (1, \ldots, 1)$.*

The proofs of results in this section appear in Online Appendix OB. The equilibrium that we construct follows FM, but involves repeated and truthful reporting of one's own discount factor, so that continuation payoffs can be adjusted to compensate players for their mixing during punishment phases. Repeated communication is

[15] Still, we suppose here that player i knows that the preferences of $-i$ can be represented by discounted utility, and he knows the functions g_{-i}.

[16] It is natural to wonder whether explicit communication cannot be replaced with phases in which communication occurs via actions. This is not obvious, as the natural candidate for a message space is the unit interval, the domain of the discount factor.

[17] Given that communication is allowed, adding a PRD is innocuous.

convenient to address issues arising after a history along which a player has misrepresented his preferences. Since the equilibrium must be subgame-perfect, it would no longer be possible to make such a player randomize appropriately otherwise.

The challenge is to provide incentives for truth-telling. This requires some care in defining continuation payoffs after punishment phases. Following the end of such a phase, let $W : (0,1) \to \mathbb{R}_+$ denote a player's reward, as a function of the discount factor, evaluated at the end of the punishment phase, that compensates a player for the way he has randomized on path. The function W can be chosen to be completely monotone, as we show. Further, it can be split over any two consecutive rounds t and $t+1$ in an incentive-compatible way, and further split over pairs of consecutive rounds so that each increment is small enough to be feasible and individually rational. More precisely, given $W(\delta)$ and $\epsilon > 0$, we pick T rounds, numbers $k_1, k_2, ..., k_T \in (0,1)$ adding up to one, such that a fraction k_t is dispensed in rounds $t, t+1$, with $|k_t W(\delta)| < \epsilon$ for all t.

5.2. **Limit Blackwell Payoffs.** Rather than fixing a discount factor, and characterizing the set of Blackwell equilibrium payoffs evaluated at this discount factor, one might wonder what payoffs can be achieved as limit payoffs of some Blackwell equilibrium.

Definition 8. A payoff vector v is a **limit Blackwell payoff** if there exists a Blackwell SPNE σ such that $U(\sigma; \boldsymbol{\delta}) \to v$ as $\boldsymbol{\delta} \to (1, \ldots, 1)$.

As one might surmise, the set of such payoffs matches the one that appears in Theorem 1, despite the fact that not all Blackwell equilibria have payoffs that converge as $\boldsymbol{\delta} \to (1, \ldots, 1)$.

Theorem 6. *Fix a repeated game under perfect monitoring, such that the dimension of F is n. A payoff vector is a limit Blackwell payoff if it is in F^{MI}, and only if it is in the closure of F^{MI}.*

The proof proceeds as that of Theorem 1. The main difference is that on-path play yields payoffs that converge to the target payoff. To this end, we construct sequences of pure actions that approximate the target payoff, yet preserve individual rationality for low discounting.

6. Conclusion

We apply the Blackwell optimality criterion to repeated games. This restricts equilibrium behavior by ruling out mixed (non-pure) strategies in general, except for

particular profiles that depend on the monitoring structure. This restriction on behavior implies bounds on equilibrium payoffs, which reflect and clarify the role that mixed strategies play under different monitoring structures. Under perfect monitoring, they are used during minmaxing. Under imperfect public monitoring, they also help detection. Under private, conditionally independent monitoring, they must be part of any equilibrium that is not the repetition of the stage-game Nash equilibrium.

As a result, the minmax levels must be adjusted under perfect and imperfect public monitoring, and the identifiability conditions must be strengthened under imperfect public monitoring. With these modifications, folk theorems apply. Finally, under private, conditionally independent monitoring, only the repetition of the stage-game Nash equilibrium survives in prisoner's dilemma-like games.

This paper is very much a first pass. Mixed strategies also play an important role when considering games with short-run vs. long-run players (see Fudenberg, Kreps, and Maskin, 1990).[18] Their importance under general private monitoring also remains to be seen.

Our results provide a somewhat nuanced justification for the skepticism with which mixed strategies are often viewed by empiricists when modeling long-run relationships, and their focus on pure strategies. At the same time, it may be that mixed strategies can be "purified" here as well (Harsanyi, 1973b). What equilibria survive under the Blackwell optimality criterion in a setting that includes random payoff shocks (see Bhaskar, Mailath and Morris, 2008; Pęski, 2012) is an open question.

[18]Here, under perfect monitoring, players are "hybrid:" they behave as if they were short-run as far as behavior strategies are concerned, but long-run for pure strategies.

Appendix A: Proofs for Section 2 (Perfect Monitoring)

Proof of Theorem 1. We start by stating a useful result, which allows construction of action sequences with desired payoffs and continuation payoffs within given bounds.

Theorem 7 (Dasgupta and Ghosh, 2021)**.** *For all $v \in F$ and $\varepsilon > 0$, there exists $\hat{\delta} > 0$ such that for any $\delta \geq \hat{\delta}$ there is a sequence of action profiles $(a^{(t)} : t \geq 0) =: a(v, \varepsilon, \delta)$ such that*

$$v = (1-\delta)\sum_{t\geq 0}\delta^t g(a^{(t)}); \quad \left\|(1-\delta)\sum_{t\geq\tau}\delta^{t-1}g(a^{(t)}) - v\right\| \leq \varepsilon \ \forall \tau \geq 1. \tag{6.1}$$

In words, given an ε and a high enough discount δ, the discounted payoff of the whole sequence is v, while the continuation payoff from any time τ onwards is ε-close to v. As our purposes require strategies designed without knowledge of the exact discount factors, we need to know how the continuation payoffs of a fixed sequence of actions change as individual discount factors increase. The next lemma answers this by showing that if *all* δ-discounted continuation payoffs of a sequence are bounded above and below, the same is true at higher discount factors. This helps us get discount robustness.

Lemma 5 (Patience Lemma)**.** *Given $\delta \in (0,1)$, if a sequence of real numbers $(x^{(t)})_{t\in\mathbb{Z}_+}$ satisfies*

$$\underline{x} \leq (1-\delta)\sum_{t=\tau}^{\infty}\delta^{t-\tau}x^{(t)} \leq \bar{x}, \forall \tau \geq 0 \tag{6.2}$$

for some $\overline{x}$ and $\underline{x}$ in $\mathbb{R}$ and some $\delta \in (0,1)$, then the same inequalities (6.2) hold for any $\delta' \in (\delta, 1)$.

Proof. Define $f : [\delta, 1) \times \mathbb{Z}_+ \to \mathbb{R}$ by

$$f(\delta', \tau) := (1-\delta')\sum_{t=\tau}^{\infty}\delta'^{t-\tau}(x^{(t)} - \underline{x}) \tag{6.3}$$

which is *ex hypothesi* non-negative when $\delta' = \delta$ for every $\tau \geq 0$. For any $\delta' > \delta$, we have after some standard substitutions and simplifications:

$$f(\delta', \tau) \ = \frac{1-\delta'}{1-\delta}f(\delta, \tau) + \frac{1-\delta'}{1-\delta}(\delta' - \delta)\sum_{t=\tau+1}^{\infty}\delta'^{t-\tau-1}f(\delta, t). \tag{6.4}$$

From $\delta' - \delta \geq 0$ and (6.3) it follows that the right side of (6.4) is non-negative. This leads to the inequality on the right side of (6.2) at δ'; the other follows similarly. □

We now give a constructive proof of the positive part of Theorem 1.

Proof. Fix $v \in F^{\mathrm{MI}}$. The overall structure follows folk theorems closely—unilateral deviations are followed by minmax punishments, followed by a post-minmax phase that rewards every player who carried out the minmax phase. Following Abreu, Dutta, and Smith (1994) we now construct n points that will serve as post-minmax payoffs at a given discount factor; we shall also show that at higher discount factors the actual post-minmax payoffs will be nearby.

First note that Full Dimensionality implies Abreu, Dutta, and Smith (1994)'s Non-Equivalent Utility; hence we are able to obtain points $\{x(i) \in F | i = 1, 2, \ldots, n\}$ satisfying payoff asymmetry (PA): $\forall i, j$ with $i \neq j$, $x_i(i) < x_i(j)$. Let $w(i) \in F$ be the point in F where i gets the lowest feasible payoff, ignoring considerations of individual rationality. Now for each pair $(\beta, \eta) \in (0,1)^2$ and each i, let

$$y(i) := \beta(1-\eta)w(i) + \beta\eta x(i) + (1-\beta)v.$$

For suitably small choices of β and η, by construction these points have the following properties for all i: (1) strict myopic indifference rationality (SMIR), *i.e.*, $y_j(i) > \underline{v}_j^{\mathrm{MI}}$ for all j; (2) PA; (3) target payoff dominance (TPD), *i.e.*, $y_i(i) < v_i$. Since all inequalities are strict, take any $\varepsilon > 0$ such that all the above inequalities hold with a slack of 3ε.

Each $y(i)$ is generated by a convex combination of the pure-action payoffs $\{g(a) | a \in A\}$, so $y(i) \in F$. We approximate each $y(i)$ within ε by a rational convex combination of the pure payoff points. Without loss of generality we can use these weights to construct sequences $(\widetilde{a}_t^i)_{t=0}^{t=T-1}$ of the same length T such that

$$\left\| \frac{1}{T} \sum_{t=0}^{T-1} g(\widetilde{a}_t^i) - y(i) \right\| < \varepsilon. \tag{6.5}$$

Defining

$$v(i) := \frac{1}{T} \sum_{t=0}^{T-1} g(\widetilde{a}_t^i) \in F^{\mathrm{MI}}, \tag{6.6}$$

we have obtained points $\{v(i) \in \mathbb{R}^n | i = 1, 2, \ldots, n\}$ and for each i a finite sequence of action profiles $\widetilde{a}^i = (\widetilde{a}_t^i)_{t=0}^{t=T-1}$ suitably reordered so that the payoff of i is increasing along the sequence for i (*i.e.*, $g_i(\widetilde{a}_t^i) < g_i(\widetilde{a}_{t'}^i)$ whenever $t < t'$). Since the SMIR, PA,

and TPD constraints for the $\{y(i)\}$ all held with a slack of 3ε, (6.5) implies that these constraints will hold with a slack of at least ε for $\{v(i)\}$:

$$\forall i,\ \underline{v}_i^{\mathrm{MI}} + \varepsilon < v_i(i), \tag{6.7}$$

$$\forall i,\ v_i(i) + \varepsilon < v_i, \tag{6.8}$$

$$\forall i \neq j,\ v_i(i) + \varepsilon < v_i(j). \tag{6.9}$$

As usual, we interpret each $v(i)$ as giving each player $j \neq i$ a 'reward' for punishing i. Extend each such finite sequence to a periodic sequence by defining $\widetilde{a}_t^i := \widetilde{a}_{t \bmod T}^i$ for $t \geq T$. For each i, the punishment profile for player i is a mixed action profile $\alpha^i \in \mathcal{A}^{\mathrm{MI}}$ such that

$$\alpha^i \in \arg\min_{\alpha \in \mathcal{A}^{\mathrm{MI}}} \max_{a_i} g_i(a_i, \alpha_{-i}). \tag{6.10}$$

Choose an $N \in \mathbb{N}$ such that for all i,

$$\max_{a \in A} g_i(a) + N\, g_i(\alpha^i) < (N+1)(v_i(i) - \varepsilon), \quad \forall i, \tag{6.11}$$

which is possible because of $g_i(\alpha^i) \leq \underline{v}_i^{\mathrm{MI}}$ and (6.7).

Choose a $\hat{\delta}$ high enough that for all $\delta > \hat{\delta}$, the implication of Theorem 7 holds. Then choose $\underline{\delta} \geq \hat{\delta}$ so that all of the following hold for each i and each $\delta \geq \underline{\delta}$:

$$\text{(Rewards)} \qquad \forall\, k \in \mathbb{Z}_+,\ \left\| \frac{1-\delta}{1-\delta^T} \sum_{t=0}^{T-1} \delta^t g(\widetilde{a}_{(t+k)}^i) - v(i) \right\| < \varepsilon,$$

$$\text{(IC-I)} \qquad (1-\delta) \max_{a \in A} g_i(a) + \delta[(1-\delta^N) g_i(\alpha^i) + \delta^N v_i(i)] < (1-\delta) \min_{a \in A} g_i(a) + \delta(v_i - \varepsilon),$$

$$\text{(IC-II(i))} \qquad \underline{v}_i^{\mathrm{MI}} < (1-\delta^N) g_i(\alpha^i) + \delta^N (v_i(i) - \varepsilon),$$

$$\text{(IC-II(j))} \qquad \forall t \leq N, \forall j \neq i,\ (1-\delta) \max g_i(a) + \delta[(1-\delta^N) g_i(\alpha^i) + \delta^N v_i(i))] < (1-\delta^t) g_i(\alpha^j) + \delta^t (v_i(j) - \varepsilon),$$

$$\text{(IC-III(i))} \qquad (1-\delta) \max_{a \in A} g_i(a) + \delta(1-\delta^N) g_i(\alpha^i) < (1-\delta^{N+1})(v_i(i) - \varepsilon),$$

$$\text{(IC-III(j))} \qquad \forall j \neq i,\ (1-\delta) \max g_i(a) + \delta[(1-\delta^N) g_i(\alpha^i) + \delta^N v_i(i)] < v_i(j) - \varepsilon.$$

For (Rewards) such a choice is possible by (6.6), and for (IC-I) (resp. (IC-II(i)), (IC-II(j)), (IC-III(i)), (IC-III(j))) because its limit as $\delta \uparrow 1$ reduces to $v_i(i) < v_i - \varepsilon$ (resp. $\underline{v}_i^{\mathrm{MI}} < v_i(i) - \varepsilon$, $v_i(i) < v_i(j) - \varepsilon$, (6.11), $v_i(i) < v_i(j) - \varepsilon$), which holds by (6.8) (resp. (6.7), (6.9), the choice of N, (6.9)).

Strategies.

For any $\delta > \underline{\delta}$, we define a strategy profile that is a Blackwell equilibrium above δ. Play is based on the following phases:
Phase I: Play $a(v, \varepsilon, \delta)$, a sequence of pure action profiles satisfying (6.1).
Phase II(i): Play α^i for N rounds.
Phase III(i): Play $\widetilde{a}^i$, starting at $\widetilde{a}^i_0$.

We construct a simple strategy profile à la Abreu (1988): We start in Phase I; unilateral deviations by a player j from any phase lead to Phase II(j) followed by Phase III(j).

As we used Theorem 7 to generate Phase I, the payoffs of the specified strategies evaluated at discount δ are v. It remains to show that for any $\delta' \geq \delta$, the strategies form an SPNE; *i.e.*, that the strategies are a Blackwell SPNE above δ.

For any $\delta' \in (0,1)$, let $v_i^t(\delta')$ and $v_i^t(j)(\delta')$ denote the δ'-discounted continuation payoff of the path in Phase I and Phase III(j) respectively, after $t-1$ rounds of the corresponding phase (not of the entire game) have elapsed. Note that, differently from FM and standard perfect-monitoring folk theorems, we do not ask a player to (myopically) best respond during her own punishment phase, as that would potentially not leave the others willing to mix.

From Theorem 7 and the Patience Lemma, we have

$$\forall i, \forall t, \forall \delta' \geq \delta,\ v_i^t(\delta') \geq v_i - \varepsilon. \tag{6.12}$$

From (Rewards) we have

$$\forall i, \forall j, \forall t, \forall \delta' \geq \delta,\ |v_i^t(j)(\delta') - v_i(j)| < \varepsilon. \tag{6.13}$$

From the fact that $g_i(\widetilde{a}_t^i) \leq g_i(\widetilde{a}_{t+1}^i)$ for $t \in \{0, 1, ...T-2\}$, we have

$$\forall i, \forall \delta' \geq \delta,\ v_i^1(i)(\delta') \leq v_i(i) \text{ and} \tag{6.14}$$

$$\forall i, \forall t, \forall \delta' \geq \delta,\ v_i^1(i)(\delta') \leq v_i^t(i)(\delta'). \tag{6.15}$$

Checking subgame perfection.
Step 1. Player i cannot profit by deviating from Phase I if for any $t \in \mathbb{N}$,

$$(1-\delta') \max_{a \in A} g_i(a) + \delta'[(1-\delta'^N) g_i(\alpha^i) + \delta'^N v_i^1(i)(\delta')] \leq (1-\delta') \min_{a \in A} g_i(a) + \delta' v_i^t(\delta').$$

Using (6.12), (6.14) and $g_i(\alpha^i) \leq \underline{v}_i^{\text{MI}}$ we need only show

$$(1-\delta') \max_{a \in A} g_i(a) + \delta'[(1-\delta'^N) \underline{v}_i^{\text{MI}} + \delta'^N v_i(i)] \leq (1-\delta') \min_{a \in A} g_i(a) + \delta'(v_i - \varepsilon),$$

which is identical to (IC-I), which applies as $\delta' \geq \delta \geq \underline{\delta}$.

Step 2. Player i cannot profit by deviating from Phase II(i) if for any $t = 1, 2, \ldots, N$,

$$(1-\delta')\underline{v}_i^{\mathrm{MI}} + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N v_i^1(i)(\delta')] < (1-\delta'^t)g_i(\alpha^i) + \delta'^t v_i^1(i)(\delta').$$

Using (6.13) and $g_i(\alpha^i) < v_i^{\mathrm{MI}} \leq v_i(i) - \varepsilon$ from (6.7), we can get the sufficient condition

$$(1-\delta')\underline{v}_i^{\mathrm{MI}} + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N(v_i(i)-\varepsilon)] < (1-\delta'^N)g_i(\alpha^i) + \delta'^N(v_i(i)-\varepsilon),$$

which reduces to (IC-II(i)), which applies as $\delta' \geq \delta \geq \underline{\delta}$.

Step 3. Player i cannot profit by deviating from Phase III(i) if for any $t \in \mathbb{N}$,

$$(1-\delta') \max_{a \in A} g_i(a) + \delta'[(1-\delta'^N)\, g_i(\alpha^i) + \delta'^N v_i^1(i)(\delta')] \leq v_i^t(i)(\delta').$$

Given (6.15), this inequality holds if

$$(1-\delta') \max_{a \in A} g_i(a) + \delta'(1-\delta'^N)\, g_i(\alpha^i) \leq (1-\delta'^{N+1})v_i^t(i)(\delta'),$$

so that we can now use (6.13) to get the sufficient condition

$$(1-\delta') \max_{a \in A} g_i(a) + \delta'(1-\delta'^N)\, g_i(\alpha^i) \leq (1-\delta'^{N+1})(v_i(i)-\varepsilon),$$

which is satisfied due to (IC-III(i)) given that $\delta' \geq \delta \geq \underline{\delta}$.

Step 4. Player i does not deviate (observably) from Phase II(j) if for all remaining punishment rounds $t \leq N$

$$(1-\delta') \max g_i(a) + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N v_i^1(i)(\delta')] < (1-\delta'^t)(g_i(\alpha^j) + \delta'^t v_i^1(j)(\delta').$$

It suffices to use (6.14) and (6.13) to obtain the sufficient condition

$$(1-\delta') \max g_i(a) + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N v_i(i)(\delta')] < (1-\delta'^t)(g_i(\alpha^j) + \delta'^t(v_i(j)-\varepsilon),$$

which is (IC-II(j)), which applies as $\delta' \geq \delta \geq \underline{\delta}$.

Step 5. Player i cannot profit by mixing differently in Phase II(j). Recall our definition of $\mathcal{A}^{\mathrm{MI}}$; since $\alpha^j \in \mathcal{A}^{\mathrm{MI}}$, mixing only occurs between myopically indifferent actions according to α^j; as future play does not vary over i's actions on $\mathrm{supp}(\alpha_i^j)$, he does not have a strict incentive to deviate.

Step 6. Player i cannot profit by deviating from Phase III(j) if for any $t \in \mathbb{N}$,

$$(1-\delta') \max g_i(a) + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N v_i^1(i)(\delta')] < v_i^t(j)(\delta').$$

so that using (6.14) and (6.13) we can use (IC-III(j)) as a sufficient condition.

Therefore for any $\delta' \geq \delta$, the specified strategies form an SPNE, and hence they are a Blackwell SPNE above δ.

If a PRD is available, the equilibrium strategies can be modified as follows:
Phase I: At each round, play the correlated action $p \in \Delta A$ such that $v = \sum_{a \in A} p(a)g(a)$.
Phase II(i): Play α^i for N rounds.
Phase III(i): Play $p^i \in \Delta A$ such that $v(i) = \sum_{a \in A} p^i(a)g(a)$ at each round of the phase.

It is easy to see that the resulting strategies constitute a Blackwell SPNE if δ satisfies the sufficient condition $\delta \geq \underline{\delta}$ in our PRD-free construction above. □

Appendix B: Proofs for Section 3 (Imperfect Monitoring)

Proof of Lemma 2. Following FLM 1994, we proceed by first showing a local version of the self-generation property we seek.

Definition 9. A set $W \subset \mathbb{R}^n$ is said to be **locally strong self-generating** if for any $v \in W$ we can find an open set $\mathcal{O}_v$, an $\eta_v > 0$ and a $\delta_v < 1$ such that

$$v \in \mathcal{O}_v \cap W \subset \mathcal{B}_{\eta_v}(W; \delta) \; \forall \delta \geq \delta_v.$$

We will show that a closed ball $B(c, r) \subset \mathbb{R}^n$ is locally strong self-generating if it lies in the interior of the set $F^{\mathrm{MI},\pi}$.

To this end it is useful to introduce the notion of MI-score. Following Matsushima (1989) and FL 1994, for any non-zero direction λ we can find a point $v^*(\lambda)$ that lies on the highest hyperplane in direction λ subject to the point itself being generated by a current action in $\mathcal{A}^{\mathrm{MI}}$, and continuation payoffs that lie below the said hyperplane.[19] Let H^- denote the lower half-space function, *i.e.*, $H^-(\lambda, k) := \{z \in \mathbb{R}^n : \lambda \cdot z \leq k\}$. Let $\mathcal{B}(\cdot, \alpha; \delta)$ denote the usual APS operator for a fixed current action profile α. That is, $v \in \mathcal{B}(W, \alpha; \delta)$ if there is a $w : Y \to W$ such that v is the payoff of the Nash equilibrium α of the normal-form game with payoffs $(1-\delta)g(a) + \delta \mathbb{E}[w|a]$.

Definition 10. The **MI-score** in direction λ is

$$k^{\mathrm{MI}}(\lambda) := \sup_{v \in \mathbb{R}^n} \left\{ \lambda \cdot v \middle| v \in \bigcup_{\alpha \in \mathcal{A}^{\mathrm{MI}}} \mathcal{B}(H^-(\lambda, \lambda \cdot v), \alpha; \delta) \right\}. \tag{6.16}$$

The region bounded by the MI-score in each direction serves as an upper bound on the equilibrium payoff set. The difference between the usual score and the MI-score in (6.16) is that we restrict the current action α to have the myopic indifference property.

[19] Matushima (1989) proposed an algorithm to characterize the upper boundary of the equilibrium payoff set, when first-order conditions suffice for a maximum; FL 1994 extended this to all directions to characterize the entire set, restricting attention to finite action spaces to enable sufficient conditions to be imposed explicitly.

The computation of the score is, loosely speaking, more flexible than the computation of self-generation because for any direction it allows us to generate payoffs using continuations in the lower half-space rather than in a smaller self-generating set. Note also that incentives aren't strict in calculating the score, since we use the standard APS operator $\mathcal{B}$ rather than our strong version $\mathcal{B}_\eta$ for some $\eta > 0$. While our equilibrium construction requires strict incentives, this is tackled perturbing it to create a point with almost the maximal score but with strict incentives.

Online Appendix OA shows how IFR implies that $F^{\mathrm{MI},\pi} = \{v \in \mathbb{R}^n \big| \lambda \cdot v \leq k^{\mathrm{MI}}(\lambda)\ \forall \lambda \neq 0\}$. So to prove Lemma 2 it suffices to show that a closed ball $B(c,r)$ is locally strong self-generating if the MI-score in any direction $\lambda \neq 0$ exceeds the value $\lambda \cdot v$ for any $v \in B(c,r)$.

Any $v \in B(c,r)$ falls in one of two cases.

Case 1: $v \in \mathrm{int}(B(c,r))$.

We can find $\mu > 0$ such that $B(v,\mu) \subset \mathrm{int}(B(c,r))$. Fixing a NE α^* of the stage game G, we pick δ_v large enough so that the implied continuation payoffs at each $v' \in B(v,\mu)$ (which are chosen to be constant in both the signal and the output of the randomization device) lie in $\mathrm{int}(B(c,r))$.

Case 2: $v \in \partial B(c,r)$.

The unit normal vector at v pointing away from $B(c,r)$ is $\lambda := (v-c)/\|v-c\|$. If there is an NE of G that lies above this hyperplane we follow the arguments as in Case 1. If not, pick a point v^* that gives the maximal MI-score in direction λ and find the associated mixed action $\alpha^* \in \mathcal{A}^{\mathrm{MI}}$ and $x^* : Y \to \mathbb{R}^n$, the associated normalized continuation payoff, [20] which satisfies that for all y, $\lambda \cdot x^*(y) \leq 0$.

For each i, order the actions of i as $A_i = \{a_{i,1}, \ldots, a_{i,K}\}$. Define $v_i' \in \mathbb{R}^{|A_i|}$ by introducing a 'penalty' of 1 for any action that is not in the support of the action that generates v^*:

$$v'_{i,k} = \begin{cases} v_i & \text{if } a_{i,k} \in \mathrm{supp}(\alpha_i^*) \\ v_i - 1 & \text{otherwise.} \end{cases}$$

By IFR, the matrix $\Pi_i(\alpha)$ whose rows are transposes of the column vectors $\pi(y|\alpha_{-i}, a_i)$, one for each $a_i \in A_i$, has full row rank; therefore the following linear equation has a

[20] For details, see Mailath and Samuelson (2006). For all $\delta \in (0,1)$, $(\alpha^*, \frac{1-\delta}{\delta}x^* + v^*)$ generates v^* on $H^-(\lambda, \lambda \cdot v^*)$ at δ.

solution $x_i' \in \mathbb{R}^{|Y|}$:

$$\begin{bmatrix} v'_{i,1} - g_i(a_{i,1}, \alpha^*_{-i}) \\ \vdots \\ v'_{i,K} - g_i(a_{i,K}, \alpha^*_{-i}) \end{bmatrix} = \Pi_i(\alpha^*) \begin{bmatrix} x_i'(y_1) \\ \vdots \\ x_i'(y_{|Y|}) \end{bmatrix}. \tag{6.17}$$

Therefore, we've converted our demand for strictness in payoffs into a signal-specific reward function. Having done this for each i, define the function $x' : Y \to \mathbb{R}^n$ by combining the x_i', *i.e.*, $x'(y) = (x_1'(y), \ldots, x_n'(y))$. Take $\beta, \gamma \in (0,1)$ small enough that for all y, we have $0 > \lambda \cdot (\beta\gamma x'(y) + \beta(1-\gamma)(v - v^*) + (1-\beta\gamma)x^*(y))$; this is possible from $\lambda \cdot x^*(y) \leq 0$ and $\lambda \cdot (v - v^*) < 0$. Let $x''(y) := \beta\gamma x'(y) + \beta(1-\gamma)(v-v^*) + (1-\beta\gamma)x^*(y)$, so that $\lambda \cdot x''(y) < 0$.

Let $v'' := \beta v + (1-\beta)v^*$. Since v^* maximises the MI-score and $\lambda \cdot v < k^{\mathrm{MI}}(\lambda)$, we have $\lambda \cdot v < \lambda \cdot v^*$, and therefore v'' satisfies $\lambda \cdot v < \lambda \cdot v''$.

Therefore, the following hold for all i:

$$v_i'' = g_i(a_i, \alpha^*_{-i}) + \mathbb{E}(x_i''|(a_i, \alpha^*_{-i})), \qquad \text{if } a_i \in \mathrm{supp}(\alpha_i^*), \tag{6.18}$$

$$v_i'' \geq g_i(a_i, \alpha^*_{-i}) + \mathbb{E}(x_i''|(a_i, \alpha^*_{-i})) + \beta\gamma, \qquad \text{otherwise.} \tag{6.19}$$

We define $\eta_v = \beta\gamma$, the degree of slack we've introduced.

If λ is not a negative coordinate direction, α^* can be chosen to be pure; otherwise, perform the following 'garbling' to ensure (3.5):

- Publicly draw ν from the uniform distribution on $[0,1]^n$.[21]
- Let $\underline{x}_j$ be the lowest value of the of x_j'' on Y, and let $\overline{x}_j$ be the highest.
- Let $\widetilde{x}(y, \nu) = (\widetilde{x}_j(y_j, \nu_j))_{j \in I}$ be given by

$$\widetilde{x}_j(y_j, \nu_j) = \begin{cases} \overline{x}_j, & \text{if } \nu_j \leq \frac{x_j''(y_j) - \underline{x}_j}{\overline{x}_j - \underline{x}_j}, \\ \underline{x}_j, & \text{otherwise.} \end{cases}$$

Note that $\widetilde{x}_j$ is a garbling of x_j'' that preserves the expectation, and that

$$\Pr(\widetilde{x}_j = \overline{x}_j \mid a_j) = \sum_{y_j \in Y_j} \pi_j(y_j \mid a_j) \frac{x_j''(y_j) - \underline{x}_j}{\overline{x}_j - \underline{x}_j},$$

[21]Combined with Lemma 1, an $(n+1)$-dimensional public randomization device suffices in each round, one dimension each to garble the continuation payoffs and an extra one for the reboot decision.

which is constant on $\text{supp}(\alpha_j)$.[22] This needs to be done only for negative-coordinate directions. For all other directions, set $\widetilde{x}_j(y_j, \nu_j) = x''_j(y_j)$ for all ν. Thus

$$v''_i = g_i(a_i, \alpha^*_{-i}) + \mathbb{E}_{y,\nu}(\widetilde{x}_i|(a_i, \alpha^*_{-i})), \qquad \text{if } a_i \in \text{supp}(\alpha^*_i), \tag{6.20}$$

$$v''_i \geq g_i(a_i, \alpha^*_{-i}) + \mathbb{E}_{y,\nu}(\widetilde{x}_i|(a_i, \alpha^*_{-i})) + \eta_v, \qquad \text{otherwise.} \tag{6.21}$$

Since $\lambda \cdot v < \lambda \cdot v''$, add $v - v''$ to both sides to translate the old normalised continuation function to a new one x:

$$v = g(\alpha^*) + \mathbb{E}_{y,\nu}(x|\alpha^*), \text{ where } x := \widetilde{x}_j + v - v''.$$

The continuation payoff point $w(y, \nu; \delta) = (w_j(y_j, \nu_j;\ \delta))_{j \in I}$ is given by

$$w_j(y_j, \nu_j; \delta) = v_j + \frac{1-\delta}{\delta} x_j(y_j, \nu_j), \tag{6.22}$$

which satisfies $\delta^2 \|w(y, \nu; \delta) - c\|^2 = (1-\delta)^2 \|x\|^2 + 2\delta(1-\delta)x \cdot (v-c) + \delta^2 r^2$. Now, using the fact that $\lambda = \frac{v-c}{\|v-c\|}$, we have $x \cdot (v-c) = \|v-c\|\lambda \cdot (\widetilde{x} + v - v'') = \|v-c\|\lambda \cdot \widetilde{x} + \|v-c\|(\lambda \cdot v - \lambda \cdot v'') < 0$, by $\lambda \cdot \widetilde{x} < 0$ and $\lambda \cdot v < \lambda \cdot v''$. Thus there is a $\delta_{y,\nu}$ such that for $\delta \geq \delta_{y,\nu}$, the continuation payoff $w(y, \nu; \delta)$ lies in the interior of $B(c, r)$. Take δ_v to be $\max_{y,\nu} \delta_{y,\nu}$, which is less than 1 as each $\delta_{y,\nu} < 1$. At each $\delta > \delta_v$, we have

$$v_i = (1-\delta)g(\alpha_i, \alpha^*_{-i}) + \delta \mathbb{E}_{y,\nu}(w_j(\cdot; \delta)|(\alpha_i, \alpha^*_{-i})), \qquad \text{if } a_i \in \text{supp}(\alpha^*_i), \tag{6.23}$$

$$v_i \geq (1-\delta)g(\alpha_i, \alpha^*_{-i}) + \delta \mathbb{E}_{y,\nu}(w_j(\cdot; \delta)|(\alpha_i, \alpha^*_{-i})) + (1-\delta)\eta_v, \qquad \text{otherwise.} \tag{6.24}$$

Thus

$$v \in \mathcal{B}_{\eta_v}(B(c, r); \delta)\ \forall \delta \geq \delta_v.$$

Since translating the continuation payoff function leaves incentives unaffected, there exists $\gamma_v > 0$ such that all points in $B(c, r) \cap B(v, \gamma_v)$ can be generated by using continuations in $B(c, r)$ at δ_v, preserving (6.23) and (6.24). Moreover, at $\delta > \delta_v$ translated continuation values are in $B(c, r)$, as they are convex combinations of translated original points and translated continuations at δ_v, both of which are in $B(c, r)$. Thus, for all $\delta > \delta_v$, we have $B(v, \gamma_v) \cap B(c, r) \subset \mathcal{B}_{\eta_v}(B(c, r); \delta)$.

Combining Cases 1 and 2 we see that $B(c, r)$ is locally strong self-generating. Now we can leverage our local self-generation property into a global one.

[22] The degenerate case of $\overline{x}_j = \underline{x}_j$ is trivial.

Let $Z \subset \operatorname{int} F^{\mathrm{MI},\pi}$ be a compact locally strong self-generating set. Since Z is locally strong self-generating, for any $z \in Z$ there exists $\gamma_z > 0$ and $\eta_z > 0$ such that $B(z,\gamma_z) \cap Z$ can be η_z-strongly generated by Z at all $\delta > \delta_z$. Then $\{(\operatorname{int} B(z,\gamma_z)) \cap Z\}_{z\in Z}$ forms an open cover of Z. Since Z is compact, extract a finite subcover $\{B(z_1,\gamma_{z_1}),\dots,B(z_L,\gamma_{z_L})\}$. Now let $\delta_Z := \max\{\delta_{z_1},\dots,\delta_{z_L}\}$, which is strictly less than 1 as a maximum of finitely many reals strictly less than 1; and $\eta^* := \min\{\eta_{z_1},...,\eta_{z_L}\}$, which is positive as the minimum of finitely many positive reals.

For all $\delta > \delta_Z$, since $B(z_\ell,\gamma_{z_\ell}) \cap Z \subset \mathcal{B}_{\eta^*}(B(c,r);\delta)$, we have that $\cup_{\ell=1}^{L}(B(z_\ell,\gamma_{z_\ell}) \cap Z) = Z$ is η^*-strongly self-generating. □

Proof of Theorem 2. The following doubly uniform version of robustness is important in this setting.

Definition 11. A pair $(W,\hat{\Delta}) \in F \times 2^{(0,1)}$ is a **doubly robust region** if for any $v' \in W$ and any $\delta \in \hat{\Delta}$, we can find a strategy profile that (i) is a PPE for any discount factor in $\hat{\Delta}$, and (ii) delivers the payoff v' at δ.

Lemma 6. *Given a closed ball $B \subset \operatorname{int} F^*$, there exists $\underline{\delta} \in (0,1)$ such that for any $\delta' \in (\underline{\delta},1)$ there is an open interval $(\delta_l,\delta_h) \ni \delta'$ such that $(B,(\delta_l,\delta_h))$ is a doubly robust region.*

Proof. First, we notice that Lemma 2 only used the PRD to get local self-generation in the negative coordinate direction. In our case, we are free to use either a pure or Nash action as in each negative coordinate direction $-e_i$ as we only need to attain a score of $\min\{\underline{v}_i^{\mathrm{pure}},\underline{v}_i^{\mathrm{NE}}\}$. Thus garbling is unnecessary, and strictness only has to be introduced when α^* is pure, which by assumption implies it satisfies IFR. Thus, a modified version of Lemma 2 holds with respect to closed balls in $\operatorname{int} F^*$, using a PRD-free version of the strong APS operator.

Thus, there is a $\underline{\delta} \in (0,1)$ and an $\eta > 0$ such that for all $\delta \in (\underline{\delta},1)$ we have $B \subset \mathcal{B}_\eta(B;\delta)$. Take any $\delta' > \underline{\delta}$; we will show there is an open interval satisfying the lemma.

Let $M := \max_{i,a} |g_i(a)|$. Take δ_l,δ_h such that

$$\underline{\delta} < \delta_l < \delta' < \delta_h \tag{6.25}$$

$$\delta_l > \delta_h - \frac{\eta(1-\delta_h)^2}{4M} \tag{6.26}$$

Given any $\delta \in (\delta_l,\delta_h)$ and any $v \in B$, using that $\underline{\delta} < \delta$ from (6.25), construct a PPE using Lemma 2 and η-strong self-generation at δ with payoff v. Then, take any

$\hat{\delta} \in (\delta_l, \delta_h)$. Take any public history and any action profile α played after that history. From the definition of strong self-generation, for any player i and action a_i one of the following is true:

- The distribution of continuation play does not vary between α and a_i, α_{-i}. Then, the fact α is a Nash equilibrium implies $g_i(\alpha) \geq g_i(a_i, \alpha_{-i})$ and hence a_i is never a better response than α_i regardless of the discount.
- Playing a_i entails a loss of at least $\eta(1-\delta)$ at discount δ. Notice the average discounted payoff function is differentiable in δ and for any outcome h^∞ we have the bound $|\frac{d}{d\delta_0} U_i(h^\infty, \delta_0)| \leq \frac{2M}{1-\delta_0}$. Thus, between any two $\delta_0, \delta_1 \in (\delta_l, \delta_h)$ each strategy can vary in payoffs by at most $\frac{2M}{1-\max\{\delta_0,\delta_1\}}$. Over the whole interval, then, which from (6.26) has length of at most $\frac{\eta(1-\delta_h)^2}{4M}$, this variation is bounded by $\eta(1-\delta_h)/2$. Thus, at $\hat{\delta}$, deviating to a_i entails a gain of less than $2\eta(1-\delta_h)/2 - \eta(1-\delta) < 0$; so the deviation is unprofitable.

Thus, the constructed strategies are a PPE at any $\hat{\delta}$ in (δ_l, δ_h), as requested. □

Now, we begin the proof of Theorem 2 in earnest. First, we construct statistical tests using (a^*, Y^*). These tests will be used to mimic a PRD and decide whether the game should or should not be reset. A success occurs when the signal is Y^*. Recall that any unilateral deviation from a^* strictly reduces the probability of a signal in the set Y^*. Let $\mathcal{T}(T^*, k^*)$ denote a test that is passed if and only if there are at least k^* successes in T^* Bernoulli trials, each with success probability $q^* = \pi(Y^*|a^*)$. The pass probability of the test is the probability of passing it if a^* is played for T^* rounds.

Play is divided into three phases — *Select*, *Normal*, and *Test*. Play begins in the *Select* phase. Intuitively, the *Select* is used to randomize whether the players next find themselves in a reward or punishment *Normal* phase. *Normal* phases are where players collect most of their payoffs. Each round n in a Normal phase is linked via incentives to the nth round in future Normal phases, until a reset is triggered. Following a *Normal* phase, play moves to the *Test* phase; if the test is failed, then the subsequent *Select* phase randomizes over punishment and reward once again; otherwise, the next *Normal* phase will continue where play left off.

Given a target payoff v, choose the following quantities in order, culminating in the choice of a cutoff discount factor $\delta_{\min}$.

Step 1. Choose $c^+, c^- \in \operatorname{int} F^*$ such that $c_i^+ > v_i > c_i^-$ for every i and the following holds:

$$v = q^* c^+ + (1 - q^*) c^-. \tag{6.27}$$

Step 2. Choose $r > 0$ such that $c_i^+ > v_i + 5r;\ v_i > c_i^- + 5r\ \forall i$; and

$$B(c^-, 5r), B(c^+, 5r) \subset \operatorname{int}\{v \in \operatorname{co} u(A) \mid \forall i \in I, v_i \geq \underline{v}_i^{\text{pure}} \wedge \underline{v}_i^{\text{NE}}\}.$$

As we will use balls of radius r rather than $5r$, the distance along each coordinate between any point in the balls we will use and the point v is at least $4r$.

Step 3. Choose a δ' high enough to satisfy the conditions of Lemma 6 for both $B(c^-, r)$ and $B(c^+, r)$. That is, for some $(\delta_l^-, \delta_h^-) \ni \delta'$ and $(\delta_l^+, \delta_h^+) \ni \delta'$ we have that $(B(c^-, r), (\delta_l^-, \delta_h^-))$ and $(B(c^+, r), (\delta_l^+, \delta_h^+))$ are both doubly robust regions. Take $(\underline{\delta}, \overline{\delta}) := (\delta_l^-, \delta_h^-) \cap (\delta_l^+, \delta_h^+)$.

Step 4. Find two binomial tests, both based on (a^*, Y^*), with pass probabilities $1 - p^+$ and p^-, such that

$$1 - p^+, 1 - p^- \in (\underline{\delta}, \overline{\delta}), \text{ and} \tag{6.28}$$

$$\left\| v - \frac{\frac{q^* c^+}{1-(1-p^+)\widehat{\delta}} + \frac{(1-q^*)c^-}{1-(1-p^-)\widehat{\delta}}}{\frac{q^*}{1-(1-p^+)\widehat{\delta}} + \frac{1-q^*}{1-(1-p^-)\widehat{\delta}}} \right\| < \frac{r}{2}, \quad \forall \widehat{\delta} \in \left(\max\{\frac{\underline{\delta}}{1-p^+}, \frac{\underline{\delta}}{1-p^-}\}, 1 \right). \tag{6.29}$$

Such tests can always be found. Since the binomial distribution approximates the normal distribution (which has a continuous CDF), we can use independent Bernoulli trials with success probability $q^* = \pi(Y^* \mid a^*)$ to design binomial tests of suitable lengths T^+ and T^- having pass probabilities arbitrarily close to $\frac{\underline{\delta}+\overline{\delta}}{2}$ and $1 - \frac{\underline{\delta}+\overline{\delta}}{2}$ respectively. If the approximations are close enough, (6.28) holds and so does inequality (6.29), because equation (6.27) implies that the LHS of inequality (6.29) goes to zero as p^+ and p^- get closer.

Step 5. As soon as the probability of the current test being passed hits zero or unity, we cannot give incentives to players to play the test action unless it is a Nash equilibrium. To ensure incentives throughout a test, we show how to modify it by truncating it appropriately. Given any test of length T^0, we can switch to playing a fixed Nash equilibrium α^{NE} of the stage game as soon as the test is conclusive (failed or passed for sure) and until T^0 rounds are up. Let $\mathcal{T}^-$ and $\mathcal{T}^+$ denote, respectively, the Nash-truncated versions of the minus and plus Binomial tests above extended to a common length $T := \max\{T^-, T^+\}$;

if $T^- < T^+ = T$, where Nash play is substituted following the conclusion of a test. If the continuation play after any test phase depends only on the test outcome (pass vs fail), playing α^{NE} is incentive compatible for any discount factor. Since the Nash equilibrium is played only when the test is conclusive, the Nash-truncated tests inherit the pass probabilities of the original tests.

Step 6. Tests not only impact subsequent play but also distort payoffs and incentives. We therefore must ensure they account for only a small proportion of all periods. This is both so that we can deliver the target payoff, and so that normal phases provide incentives to play the test action during test phases. Denoting by ρ the minimum change in the passing probability of a test when a player deviates from her test action,[23] choose N large enough that

$$N > 6\frac{\sqrt{n}M(T+1)}{r\rho}, \tag{6.30}$$

where $|u_i(a)| \leq M$ for all $a \in A, i \in I$.

Step 7. We now ensure that we can incentivize players to take the test action. We have already (in Step 6) found a N large enough to make most periods 'normal' - what remains is to find a bound on δ above which normal periods provide sufficient incentives during the test phases. Pick a $\widehat{\delta}$ such that for all $\delta \geq \widehat{\delta}$ and $i \in I$,

$$2MT + \frac{\delta^T}{1-\delta}\rho\left[\frac{(c_i^- + 3r/2)(1-\delta^N) - (1-\delta^{(N+T+1)})(v_i - 3r)}{1-(1-p^-)\delta^{(N+T+1)}}\right] < 0,$$

and

$$2MT - \frac{\delta^T}{1-\delta}\rho\left[\frac{(c^+ - 3r/2)(1-\delta^N) - (1-\delta^{(N+T+1)})(v_i + 3r)}{1-(1-p^+)\delta^{(N+T+1)}}\right] < 0.$$

In the limit as $\delta \to 1$, these become

$$\frac{2MT}{\rho}p^- + (T+1)(v_i - 3r) < N\left(v_i - c_i^- - \frac{9r}{2}\right) \tag{6.31}$$

$$\frac{2MT}{\rho}p^+ + (T+1)(v_i + 3r) < N\left(c_i^+ - v_i - \frac{9r}{2}\right). \tag{6.32}$$

Given that Step 2 guarantees that $c_i^+ - 5r > v_i > c_i^- + 5r$, both of these are implied by (6.30). Thus there is a $\hat{\delta}$ as we require.

[23] There are two tests, finitely many states in each test, finite action spaces and $n < \infty$ players, so this ρ is well-defined and, due to the assumed properties of the test action profile a^*, is strictly positive.

Step 8. We require that the payoff distortion due to tests is small so that we can achieve our target payoff. We have (6.30), but it does not take into account discounting. Thus, we find a $\delta_{\text{dist}} \in (0,1)$ such that for all $\delta > \delta_{\text{dist}}$ we have

$$\sum_{t=T+1}^{N+T} \delta^t r\rho/2 > \sum_{t=0}^{T} \delta^t 3\sqrt{n}M, \tag{6.33}$$

which is possible as in the limit as $\delta \to 1$ this inequality reduces to the undiscounted version in (6.30), which is strict.

Step 9. We want the equilibrium payoff to not vary much with the discount. To do this we find δ_s such that for all $i \in I$ and all $\delta > \delta_s$ we have

$$\left[1 - \delta \frac{1-\delta^N}{1-\delta^{(N+T+1)}}\right] |v_i| < r.$$

As δ increases to 1, the limiting condition is $(T+1)|v_i| < (N+T+1)r$, which is implied by Step 6; therefore, such a cutoff δ_s can be found.

Step 10. Finally, define

$$\delta_{\min} := \max\left\{ \left(\frac{\underline{\delta}}{1-\max\{p_-,p_+\}}\right)^{\frac{1}{N+T+1}}, \delta_{\text{dist}}, \widehat{\delta}, \delta_s \right\}. \tag{6.34}$$

Given a $\delta > \delta_{\min}$, proceed as follows.

Define $\delta^+ := \delta^{N+T+1}(1-p^+)$ and $\delta^- = \delta^{N+T+1}(1-p^-)$. By construction,

$$\delta^+, \delta^- \in (\underline{\delta}, \overline{\delta}) \tag{6.35}$$

First, we want to adjust the starting points of the self-generation algorithm in both balls to account for (slightly) different reset probabilities, but without taking into account the test rounds. Define

$$\lambda := v - \left[\frac{\frac{q^*(c^+)}{1-\delta^+} + \frac{(1-q^*)(c^-)}{1-\delta^-}}{\frac{q^*}{1-\delta^+} + \frac{1-q^*}{1-\delta^-}}\right]. \tag{6.36}$$

The expression in square brackets has the following interpretation. Suppose that we toss a coin with a probability q^* of coming up heads; if we toss heads we play a sequence of actions that gives us c^+ while if it comes up tails we play the sequence of actions that gives us c^-. However, as we reset back to the beginning with different probabilities, we do not get $q^*c^+ + (1-q^*)c^-$. Instead, we end up getting a combination of c^+ and $c^-$with weights in the ratio $q^*/(1-\delta^+)$ to $(1-q^*)/(1-\delta^-)$. The quantity λ can be interpreted as the adjustment to the starting point needed to get v; this follows as the above can be rearranged to write v as a convex combination of $c^+ + \lambda$

and $c^- + \lambda$:

$$v = \frac{\frac{q^*(c^+ + \lambda)}{1-\delta^+} + \frac{(1-q^*)(c^- + \lambda)}{1-\delta^-}}{\frac{q^*}{1-\delta^+} + \frac{1-q^*}{1-\delta^-}}.$$

By (6.29) we have that $\|\lambda\| < r/2$.

We now further adjust the starting points in view of the non-normal phases. We compute the payoff v^+ such that one round of the average *Select* phase payoff z (detailed below), then N rounds of v^+, and finally T rounds of the test $\mathcal{T}^+$ (whose average δ-discounted payoffs we denote $x_{\mathcal{T}^+}$) gives the same expected utility as getting $c^+ + \lambda$ for $(1 + N + T)$ rounds, *i.e.*,

$$(1 - \delta^{N+T+1})(c^+ + \lambda) = (1 - \delta)z + \delta(1 - \delta^N)v^+ + \delta^{N+1}(1 - \delta^T)\, x_{\mathcal{T}^+}. \tag{6.37}$$

Subtracting v^+ from both sides and using the magnitude operator and the triangle inequality, we have

$$(1-\delta^{N+T+1})||(c^+ + \lambda - v^+)|| \leq (1-\delta)||(z-v^+)|| + \delta^{N+1}(1-\delta^T)\, ||(x_{\mathcal{T}^+} - v^+)||. \tag{6.38}$$

and replacing payoff differences by $2M$, the largest possible value, we have

$$(1 - \delta^{N+T+1})||(c^+ + \lambda - v^+)|| \leq (1 - \delta)2\sqrt{n}M + \delta^{N+1}(1 - \delta^T)\, 2\sqrt{n}M, \tag{6.39}$$

from which we deduce

$$||(c^+ + \lambda - v^+)|| < \frac{(1 - \delta^{T+1})}{(1 - \delta^{N+T+1})} 2\sqrt{n}M. \tag{6.40}$$

Finally, we use Step 8 to deduce

$$||(c^+ + \lambda - v^+)|| < \frac{\delta^{T+1} - \delta^{N+T+1}}{(1 - \delta^{N+T+1})} \frac{r}{2} < \frac{r}{2}. \tag{6.41}$$

Combining this with $\|\lambda\| < r/2$, the triangle inequality implies that $\|c^+ - v^+\| < r$; thus $v^+ \in B(c^+, r)$. Define v^- similarly, which guarantees $v^- \in B(c^-, r)$. By the above, Step 3, and Lemma 6, there exists a strategy profile σ^- that is a PPE for all $\delta'' \in (\underline{\delta}, \overline{\delta})$, while delivering v^- at δ^-, and similarly a σ^+ that is a PPE in that same set of discounts delivering v^+ at δ^+.

Strategy Profile. Now we describe the strategy $\sigma(\delta, v)$ as an automaton over 4 types of phases—*Select* (1 round), *Test* (T rounds), *Norm+*(N rounds), and *Norm−*(N rounds).

- Start in *Select*.
- *Select*:

  - In the first round of play or if the test immediately before triggered a reset, play a^* once; if the selection test succeeds (which it does with probability q^*), move to $Norm+$, else move to $Norm-$.
  - If the preceding test did not trigger a reset, play α^{NE} once, and then move to a Normal phase of the same type as the one leading up to the test.
- $Norm-$: If this is the first normal phase, or the last test concluded in favor of a reset, start playing σ^- using an empty history in each of the following N rounds; if not, play according to σ^- where each of the N cycles left off. That is, round t uses the action profile $\sigma^-(h_t^*)$ with the linked history h_t^* defined by $h_t^* := (y_{t-n(N+T+1)}, y_{t-(n-1)(N+T+1)}, ...y_{t-(N+T+1)})$ where n is the number of tests since the last reset. At the end of this phase move to $Test$.
- $Norm+$: As above, with plus instead of minus.
- $Test$: Play the Nash-truncated test $\mathcal{T}^-$, lasting T rounds, if the immediately preceding normal phase was $Norm-$; else play $\mathcal{T}^+$, also lasting T rounds. Then move to $Select$.

As the next result shows, the payoff under $\sigma(\delta, v)$ is close to v, for every high enough discount factor. For its proof, see the online appendix.

Lemma 7 (Bounding payoffs)**.** *For any $\delta^* \geq \delta$, $|U_i(\sigma(\delta, v), \delta^*) - v_i| < 3r$ for all $i \in I$.*

Incentives. We now show that the strategy $\sigma(\delta, v)$ is a PPE at any $\delta^* \geq \delta_{\min}$.

At a history h that leads to the start of a $Norm$- phase, the future payoffs due to play according to σ^- until a reset are

$$\sum_{j=1}^{N} \delta^{*j-1} \sum_{\tau=0}^{\infty} [(1-p^-)\delta^{*(N+T+1)}]^\tau g_i(\sigma^-(h^*)) \tag{6.42}$$

and reflect the N 'separate' cycles of play according to the self-generation algorithm. Play during the test and select phases gives at most M per period, so that the payoff due to test and reset phases until the next reset are bounded above by

$$\sum_{\tau=0}^{\infty} [(1-p^-)\delta^{*(N+T+1)}]^\tau [M(T+1)]. \tag{6.43}$$

Finally, a reset returns continuation payoffs to $U_i(\sigma(\delta, v), \delta^*)$, so that total future payoffs from play after resets are bounded above by

$$\delta^{*(N+T+1)} \sum_{\tau=0}^{\infty} [\delta^{*(N+T+1)}(1-p^-)]^\tau p^- \frac{U_i(\sigma(\delta, v), \delta^*)}{1-\delta^*} \tag{6.44}$$

Thus, we can bound payoffs at the outset of a *Norm-* phase, following a history h, by

$$\frac{U_i(\sigma(\delta,v)(h),\delta^*)}{1-\delta^*}$$

(6.45)

$$\leq \left[\sum_{j=1}^{N}\delta^{*j-1}\sum_{\tau=0}^{\infty}[(1-p^-)\delta^{*(N+T+1)}]^\tau g_i(\sigma^-(h^*)) + \sum_{\tau=0}^{\infty}[(1-p^-)\delta^{*(N+T+1)}]^\tau[M(T+1)]\right]$$
$$+\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p^-)]^\tau p^- \frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}.$$

Consider a deviation by some player during a *Test* phase. By the fact that a^* signal-maximizes Y^*, deviation from a^* therefore decreases the probability of passing the test. Suppose the test is following a *Norm−* phase; failing the test therefore decreases the probability of a reset. Take the smallest such change, ρ. An upper bound on the within-test-phase benefits of a deviation is $2MT$. So, a deviation is not profitable if

$$2MT + \delta^{*T}\rho\left[\frac{U_i(\sigma(\delta,v)(h),\delta^*) - U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\right] \leq 0, \tag{6.46}$$

where h is the relevant history.

We can substitute the expression in (6.45) - an upper bound to $U_i(\sigma(\delta,v)(h),\delta^*)$ - to bound the LHS of (6.46) above by

$$2MT + \delta^{*T}\rho\Bigg[\sum_{j=1}^{N}\delta^{*j-1}\sum_{\tau=0}^{\infty}\{(1-p^-)\delta^{*(N+T+1)}\}^\tau g_i(\sigma^-(h^*))$$
$$+\sum_{\tau=0}^{\infty}\{(1-p^-)\delta^{*(N+T+1)}\}^\tau M(T+1)$$
$$+\left\{\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}\{\delta^{*(N+T+1)}(1-p^-)\}^\tau p^- - 1\right\}\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\Bigg],$$

which is in turn bounded above (using Step 8) by

$$2MT + \delta^{*T}\rho\Bigg[\sum_{j=1}^{N}\delta^{*j-1}\sum_{\tau=0}^{\infty}[(1-p^-)\delta^{*(N+T+1)}]^\tau[g_i(\sigma^-(h^*))+r/2] - \frac{1-\delta^{*(N+T+1)}}{1-(1-p^-)\delta^{*(N+T+1)}}\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\Bigg].$$

Now we can apply the Patience Lemma to get a new bound:

$$2MT + \delta^{*T}\rho\left[\sum_{j=1}^{N}\delta^{*j-1}\frac{c_i^- + 3r/2}{1-(1-p^-)\delta^{*(N+T+1)}} - \frac{1-\delta^{*(N+T+1)}}{1-(1-p^-)\delta^{*(N+T+1)}}\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\right].$$

Using our bounds on $U_i(\sigma(\delta,v),\delta^*)$ from Lemma (7), we can get a further upper bound,

$$2MT + \frac{\delta^{*T}}{1-\delta^*}\rho\left[\frac{(c_i^- + 3r/2)(1-\delta^{*N}) - (1-\delta^{*(N+T+1)})(v_i - 3r)}{1-(1-p^-)\delta^{*(N+T+1)}}\right],$$

which we know from Step 7 is negative. Therefore, there is no incentive to deviate during a $Test$ phase following $Norm-$. The same argument applies mutatis mutandis to a $Test$ phase following $Norm+$, as well as the $Select$ phase.

Incentives during the $Norm+$ and $Norm-$ Phases: Since the strategy in the normal phase involves playing a PPE strategy with rebooting, it follows that there is no incentive for any player to unilaterally deviate from the prescribed actions in any round.

Equilibrium payoffs. Given the construction of v^+ and v^-, the payoff vector of this equilibrium at discount δ is $U(\sigma(\delta,v),\delta) = v$. This completes our proof. □

References

[1] Abreu, D. (1988). "On the theory of infinitely repeated games with discounting," *Econometrica*, **56**, 383-396.

[2] Abreu, D., Dutta, P.K., and L. Smith (1994). "The folk theorem for repeated games: a NEU condition," *Econometrica*, **62**, 939-948.

[3] Abreu, D., Pearce, D., and E. Stacchetti (1990). "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, **58**, 1041-1063.

[4] Ahlfors, L.V. (1953). *Complex analysis: an introduction to the theory of analytic functions of one complex variable*, McGraw-Hill, New York.

[5] Arrow, K., and D. Levhari (1969). "Uniqueness of the Internal Rate of Return with Variable Life of Investment," *Economic Journal*, **79**, 560-566.

[6] Aumann, R.J., and M. Maschler (1995). *Repeated Games with Incomplete Information*, MIT Press, Boston.

[7] Bhaskar, V., Mailath, G.J., and S. Morris (2008). "Purification in the infinitely-repeated prisoners' dilemma," *Review of Economic Dynamics*, **11**, 515-528.

[8] Blackwell, D. (1962). "Discrete dynamic programming," *The Annals of Mathematical Statistics*, **33**, 719-726.

[9] Caron, R.J., and T. Traynor (2005). "The zero set of a polynomial," technical report, University of Windsor.

[10] Dasgupta, A., and S. Ghosh (2021) "Self-accessibility," *Journal of Economic Theory*, **200**.

[11] Ellison, G. (1994). "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies*, **61**, 567-588.

[12] Fudenberg, D., Kreps, D., and E. Maskin (1990). "Repeated Games with Long-run and Short-run Players," *The Review of Economic Studies*, **57**, 555-573.

[13] Fudenberg, D., and Levine, D.K (1994), "Efficiency and Observability with Long-Run and Short-Run Players," *Journal of Economic Theory*, **61**,103-135.

[14] Fudenberg, D., Levine, D.K., and E. Maskin (1994). "The folk theorem with imperfect public information," *Econometrica*, **62**, 997-1039.

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

[16] Fudenberg, D., and E. Maskin (1986). "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, **54**, 533-554.

[17] Gossner, O. (2020). "The robustness of incomplete penal codes in repeated interactions," working paper, LSE.

[18] Harsanyi, J.C. (1973a). "Oddness of the number of equilibrium points: A new proof," *International Journal of Game Theory*, **2**, 235-250.

[19] —— (1973b). "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International Journal of Game Theory*, **2**, 1-23.

[20] Istrefi, K., and S. Mouabbi (2018). "Subjective interest rate uncertainty and the macroeconomy: A cross-country analysis," *Journal of International Money and Finance*, **88**, 296-313.

[21] Kalai, E., Samet, D., and W. Stanford (1988). "A note on reactive equilibria in the discounted prisoner's dilemma and associated games," *International Journal of Game Theory*, **17**, 177-186.

[22] Koopmans, T.C. (1960). "Stationary ordinal utility and impatience," *Econometrica*, **28**, 287-309.

[23] Lehrer, E. and A. Pauzner (1999). "Repeated Games with Differential Time Preferences," *Econometrica*, **67**, 393-412.

[24] Mailath, G., and Samuelson, L. (2005). *Repeated games and reputations: long-run relationships*. Oxford University Press, Oxford.

[25] Mas-Colell, A., Whinston, M.D., and J.R. Green (1995). *Microeconomic theory*, Oxford University Press, Oxford.

[26] Matsushima, H. (1989) "Efficiency in repeated games with imperfect monitoring," *Journal of Economic Theory*, **48**,428-442.

[27] Matsushima, H. (1991). "On the theory of repeated games with private information. Part I: anti-folk theorem without communication," *Economics Letters*, **35**, 253-256.

[28] Matsushima, H. (2004). "Repeated Games with Private Monitoring: Two Players," *Econometrica*, **72**, 823-852.

[29] Neeb, K.-H. (2011). *Holomorphy and convexity in Lie theory*, De Gruyter, Berlin.

[30] Pęski, M. (2012). "An anti-folk theorem for finite past equilibria in repeated games with private monitoring," *Theoretical Economics*, **7**, 25-55.

[31] Radner, R. (1985). "Repeated Principal-Agent Games with Discounting," *Econometrica*, **53**, 1173-1198.

[32] Renault, J. (2014). "General limit value in dynamic programming," *Journal of Dynamics & Games*, **1**, 471-484.

[33] Singh, V.V., Hemachandra, N., and K.S. Mallikarjuna Rao (2013). "Blackwell optimality in stochastic games," *International Game Theory Review*, **15**, 1-18.

[34] Singh, V.V., and N. Hemachandra (2016). "Blackwell-Nash Equilibrium for Discrete and Continuous Time Stochastic Games," arXiv preprint, arXiv:1605.00977.

[35] Wen, Q. (1994). "The "folk theorem" for repeated games with complete information," *Econometrica*, **62**, 27-49.

Online Appendix OA: Additional Proofs for Section 3

The goal of this online appendix is two-fold. First, we show that the intersection over all non-zero directions of the lower half-spaces defined by the corresponding MI-scores equals the set $F^{\mathrm{MI},\pi}$. Second, we prove Lemma 7.

The intersection of the maximal lower half-spaces is

$$H^* := \bigcap_{\lambda \neq 0} H^-(\lambda, k^{\mathrm{MI}}).$$

The MI-score differs from the usual score in that only action profiles with myopic indifference can be used to achieve it. Note also that incentives aren't strict in calculating the score, since we use the standard APS operator $\mathcal{B}$ rather than our strict version $\mathcal{B}_\eta$ for some $\eta > 0$. While our equilibrium construction requires strict incentives, this is tackled perturbing it to create a point with almost the maximal score but with strict incentives.

We consider the following mutually exclusive directions: (i) non-coordinate directions, *i.e.*, at least two coordinates are non-zero; (ii) positive coordinate directions, *i.e.*, exactly one coordinate is $+1$, while the rest are zero; (iii) negative coordinate directions, *i.e.*, exactly one coordinate is -1 while the rest are zero.

Lemma 8. *For any direction λ, the* MI*-score is attained by an action profile in $\mathcal{A}^{\mathrm{MI}}$; furthermore, a pure action profile achieves this score in directions other than negative coordinate ones. Additionally, $F^{\mathrm{MI},\pi} = \{v \in \mathbb{R}^n \big| \lambda \cdot v \leq k^{\mathrm{MI}}(\lambda)\ \forall \lambda\ [\|\lambda\| = 1]\}$.*

The proof of the lemma is standard. Intuitively, IFR allows us to generate incentives in all but negative-coordinate directions. These require us to use mixed actions in general, and indeed, the Blackwell restriction implies that any such action must also be in $\mathcal{A}^{\mathrm{MI}}$; this is just the MI-minmax defined earlier, rather than the standard minmax seen in the literature.

Proof. We divide the proof into the above cases.
Case 1: λ is a non-coordinate direction. One of the highest points in F in direction λ must be generated by a pure action profile, which is in $\mathcal{A}^{\mathrm{MI}}$, and using IFR we can satisfy the IC conditions with equality using continuation payoffs in the lower halfspace — this is the same argument as in Lemma 5.1 in FL 1994.
Case 2: λ is a positive coordinate direction e_i. The score maximization reduces to $\max v_i$ subject to the IC constraints. We pick the pure action profile that maximises the payoff of i in F. Thanks to IFR we can design continuation payoffs so that incentives hold with equality.

Case 3: λ is a negative coordinate direction $-e_i$. Only here do might we need to achieve the MI-score using a mixed action which is in $\mathcal{A}^{\mathrm{MI}}$. In this case it follows from our definition of the program P_i^{MI,π_i} and FL 1994 that the score is the negative of the minmax $\underline{v}_i^{\mathrm{MI},\pi_i}$. □

Proof of Lemma 7. We first bound $U_i(\sigma(\delta,v),\delta^*)$ from above. Notice that terms are divided based on whether they come from the plus or the minus ball. Within each ball we can group terms into $1+N+T$ cycles, and finally sum over the N normal phase cycles, replacing the test phase cycles by the largest possible payoff. This gives

$$\begin{aligned}
\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*} \leq & q^*\sum_{j=1}^{N}\delta^{*j}\sum_{\tau=0}^{\infty}[(1-p^+)\delta^{*(N+T+1)}]^\tau g_i(\sigma^+(h^*)) \\
& +q^*\sum_{\tau=0}^{\infty}[(1-p^+)\delta^{*(N+T+1)}]^\tau[M(T+1)] \\
& +(1-q^*)\sum_{j=1}^{N}\delta^{*j}\sum_{\tau=0}^{\infty}[(1-p^-)\delta^{*(N+T+1)}]^\tau g_i(\sigma^-(h^*)) \\
& +(1-q^*)\sum_{\tau=0}^{\infty}[(1-p^-)\delta^{*(N+T+1)}]^\tau[M(T+1)] \\
& +q^*\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p^+)]^\tau p^+\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*} \\
& +(1-q^*)\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p^-)]^\tau p^-\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}
\end{aligned}$$

which, by Step 6, grants

$$\begin{aligned}
\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*} < & q^*\sum_{j=1}^{N}\delta^{*j}\sum_{\tau=0}^{\infty}[(1-p^+)\delta^{*(N+T+1)}]^\tau(g_i(\sigma^+(h^*))+r/2) \\
& +(1-q^*)\sum_{j=1}^{N}\delta^{*j}\sum_{\tau=0}^{\infty}[(1-p^-)\delta^{*(N+T+1)}]^\tau(g_i(\sigma^-(h^*))+r/2) \\
& +q^*\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p^+)]^\tau p_+\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*} \\
& +(1-q^*)\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p^-)]^\tau p^-\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}.
\end{aligned}$$

Since these actions are generated by the self-generation algorithm, which keeps all continuation payoffs in a ball of radius r around c^- or c^+, the Patience Lemma gives a bound on each cycle:

$$\begin{aligned}\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*} <& q^*\sum_{j=1}^{N}\delta^{*j}\frac{(c_i^+ + r + r/2)}{1-(1-p_+)\delta^{*(N+T+1)}} + (1-q^*)\sum_{j=1}^{N}\delta^{*j}\frac{(c_i^- + r + r/2)}{1-(1-p^-)\delta^{*(N+T+1)}}\\ &+q^*\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p_+)]^\tau p_+\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\\ &+(1-q^*)\delta^{*(N+T+1)}\sum_{\tau=0}^{\infty}[\delta^{*(N+T+1)}(1-p^-)]^\tau p^-\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}.\end{aligned}$$

Collecting all terms with $v(\delta^*)$ on the left, we have

$$\begin{aligned}&\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\left[1-\frac{q^*p^+\delta^{*(N+T+1)}}{1-(1-p^+)\delta^{*(N+T+1)}}-\frac{q^*p^-\delta^{*(N+T+1)}}{1-(1-p^-)\delta^{*(N+T+1)}}\right]\\ &\qquad<\sum_{j=1}^{N}\delta^{*j}\frac{q^*(c_i^+ + r + r/2)}{1-(1-p^+)\delta^{*(N+T+1)}}+\frac{(1-q^*)(c_i^- + r + r/2)}{1-(1-p^-)\delta^{*(N+T+1)}}.\end{aligned}$$

This in turn leads to

$$\begin{aligned}&\frac{U_i(\sigma(\delta,v),\delta^*)}{1-\delta^*}\left[\frac{q^*(1-\delta^{*(N+T+1)})}{1-(1-p^+)\delta^{*(N+T+1)})}+\frac{(1-q^*)(1-\delta^{*(N+T+1)})}{1-(1-p^-)\delta^{*(N+T+1)})}\right]\\ &\quad<\delta^*\frac{1-\delta^{*N}}{1-\delta^*}\left[\frac{q^*(c_i^+ + r + r/2)}{1-(1-p^+)\delta^{*(N+T+1)}}+\frac{(1-q^*)(c_i^- + r + r/2)}{1-(1-p^-)\delta^{*(N+T+1)}}\right],\end{aligned}$$

or simply

$$U_i(\sigma(\delta,v),\delta^*)<\delta^*\frac{1-\delta^{*N}}{1-\delta^{*(N+T+1)}}\left[3r/2+\frac{\frac{q^*c_i^+}{1-(1-p^+)\delta^{*(N+T+1)}}+\frac{(1-q^*)c_i^-}{1-(1-p^-)\delta^{*(N+T+1)}}}{\frac{q^*}{1-(1-p^+)\delta^{*(N+T+1)}}+\frac{(1-q^*)}{1-(1-p^-)\delta^{*(N+T+1)}}}\right].$$

This and Step 4 grants

$$U_i(\sigma(\delta,v),\delta^*)<\delta^*\frac{1-\delta^{*N}}{1-\delta^{*(N+T+1)}}[2r+v_i],$$

or,

$$U_i(\sigma(\delta,v),\delta^*)-v_i<\delta^*\frac{1-\delta^{*N}}{1-\delta^{*(N+T+1)}}2r-[1-\delta^*\frac{1-\delta^{*N}}{1-\delta^{*(N+T+1)}}]v_i,$$

so that Step 9 gives

$$U_i(\sigma(\delta,v),\delta^*)-v_i<3r.$$

Reasoning similarly with the lower bound implies

$$U_i(\sigma(\delta, v), \delta^*) - v_i > -3r.$$

The required bound follows. □

Online Appendix OB: Proofs for Section 5

Proof of Theorem 5. To prove Theorem 5, we'll first show a number of lemmata that will allow us to deliver rewards to players depending on their discount factor announcements, which make truthful reporting IC.

Lemma 9. *Without loss of generality, the function W can be chosen to be completely monotone, i.e., $W \geq 0$, and the derivatives starting with W' alternate between non-positive and non-negative.*

Proof of Lemma 9. Denote by 0 the round right after the punishment phase for player j ends. Let ξ_t denote the reward, in time-t payoff, we must pay $i \neq j$ for his action taken in round t while minmaxing j. We can choose a reward structure so that $\xi_t > 0$ whatever i's realized action at t was (in effect, choosing the lowest possible reward fully specifies the rest). Then, as $W(\delta)$ is just the sum of the rewards discounted forward to time 0, we have

$$W(\delta) = \sum_{t=-T}^{-1} \delta^t \xi_t = \sum_{t=1}^{T} \delta^{-t} \xi_{-t} > 0$$

Its derivatives are

$$W^{(n)}(\delta) = \sum_{t=1}^{T} \left[\delta^{-t-n} \xi_{-t} \prod_{m=0}^{n-1} (-t-m) \right],$$

where the product term alternates sign with n and the other terms are always positive, implying that W is completely monotone. □

Having chosen a completely monotone W, Lemma 10 delivers the reward $W(\delta)$ to a player with discount factor $\delta \in [\delta_0, 1)$ over any two successive rounds n and $n+1$ in an incentive-compatible way; in order words, if we base the reward on the player's announcement of his discount factor, he will announce truthfully. In the statement of the lemma below, $x_n(\delta)$ denotes the payoff that δ would get in round n, while $y_n(\delta)$ denotes the payoff that he would get at round $n+1$; the total discounted adjustment must equal the target adjustment $W(\delta)$.

Lemma 10. *Fix $\delta_0 > 0$ and a completely monotone map $W : [\delta_0, 1] \to \mathbb{R}$ such that both W and the absolute value of its derivative $-W'$ are bounded above by $C_1 > 0$. For each $n \in \mathbb{N}$ there exist functions $x_n, y_n : [0,1] \to \mathbb{R}$, such that, for all pairs $\delta, \widehat{\delta} \in [0,1]$,*

$$W(\delta) = \delta^n(x_n(\delta) + \delta y_n(\delta)) \geq \delta^n(x_n(\widehat{\delta}) + \delta y_n(\widehat{\delta})). \tag{6.47}$$

Furthermore, there exists C_2 such that $\max\{|x_n(\delta)|, |y_n(\delta)|\} < n\delta^{-n-1}C_2$ for all δ and n.

Proof of Lemma 10. Given W and n, define

$$x_n(\delta) = \delta^{-n}((n+1)\,W(\delta) - \delta W'(\delta));\ y_n(\delta) = \delta^{-n-1}(\delta W'(\delta) - nW(\delta)),$$

so that $W(\delta) = \delta^n(x_n(\delta)+\delta y_n(\delta))$. For any $\widehat{\delta} > \delta$ the fundamental theorem of calculus gives

$$\begin{aligned} & x_n(\widehat{\delta}) + \delta y_n(\widehat{\delta}) - (x_n(\delta) + \delta y_n(\delta)) \\ = & \int_\delta^{\widehat{\delta}} (x_n'(\mu) + \delta y_n'(\mu))\,\mathrm{d}\mu \\ = & \int_\delta^{\widehat{\delta}} \mu^{-n-2}(\delta - \mu)\left(n(n+1)\,W(\mu) - 2n\mu W'(\mu) + \mu^2 W''(\mu)\right)\mathrm{d}\mu \ \leq 0, \end{aligned}$$

where the inequality follows from the complete monotonicity of W. From here, equation (6.47) follows immediately for $\widehat{\delta} > \delta$. A symmetric argument establishes the same property for $\widehat{\delta} < \delta$. The last assertion of the lemma is immediate given the definitions of x and y, and the bound on W and $-W'$. □

Lemma 10 is not by itself suitable for delivering adjustments in our repeated game because the required payoffs may be either infeasible or feasible but not individually rational. Lemma 11 steps in and shows that adjustments can be spread over sufficiently many rounds so that each adjustment is small. Thus, given a total adjustment $W(\delta)$ and a δ we pick a large number N of rounds; we then pick fractions $k_1, k_2, ..., k_N$ that add up to one, and then deliver part n of reward, *i.e.*, $W_n(\delta) = W(\delta)\,k_n$, by splitting it over rounds n and $n+1$ using Lemma 10 (each W_n thus plays the role of h in Lemma 10).

While it seems natural to subdivide equally by letting each k_n equal $1/N$, it turns out that this is not enough to guarantee that rewards are sufficiently small, because later rewards increase too rapidly (the bounds on $x_n(\delta)$ and $y_n(\delta)$ increase too fast with n); to compensate for this we will choose a decreasing sequence of k_ns. This

shows that for high enough δ, with a PRD, we can compensate players for actions taken in the minmax phase in a way that makes it incentive compatible to reveal their discount factor δ, while keeping individual round adjustments small enough that each round's target payoff is feasible and above the standard minmax.

Lemma 11. *Let $\delta_0 > 0$ and let $W : [\delta_0, 1] \to \mathbb{R}_+$ be a completely monotonic function. For all $L > 0$, there exists a $N \in \mathbb{N}$, $\underline{\delta} \in (\delta_0, 1)$, and a collection of functions $\{z_n\}_{1 \leq n \leq N}$ with each $z_n : [\underline{\delta}, 1] \to [-L, L]$ such that*

$$W(\delta) = \sum_{t=1}^{N} \delta^n z_n(\delta); \tag{PK}$$

$$\forall \delta > \underline{\delta}, \delta \in \arg\max_{\widehat{\delta}} \sum_{n=1}^{N} \delta^n z_n(\widehat{\delta}). \tag{IC}$$

Proof of Lemma 11. Given C_2 as in Lemma 10, choose ε so that $2\varepsilon C_2 < L$. Let $N = \inf\{\widehat{N} | \sum_{n=1}^{\widehat{N}} \frac{\varepsilon}{n} > 1\}$, which exists since the harmonic series $\sum_{n \geq 1} \frac{1}{n}$ diverges. For $n = 1, ..., N-1$ set $k_n = \varepsilon/n$ and let $k_N = 1 - \sum_{n=1}^{N-1} k_n$. We will split split the total reward into N parts as $W(\delta) = \sum_n W_n(\delta)$, where $W_n(\delta) := k_n W(\delta)$; then using Lemma 10 we split each of these parts into a current and a delayed component: $W_n(\delta) = \delta^n(x_n(\delta) + \delta y_n(\delta))$. Putting $x_{N+1}(\delta) = 0$, the actual reward paid in round n for $1 \leq n \leq N+1$ is $z_n(\delta) = x_n(\delta)k_n + y_{n-1}(\delta)k_{n-1}$, where x_n is the current reward component of W_n, while y_{n-1} is the delayed reward component of W_{n-1}.

Lemma 10 gives $|x_n(\delta)| < n\delta^{-n-1}C_2$ for all δ, we have $|x_n(\delta)k_n| < \varepsilon\delta^{-n-1}C_2$. Similarly, as $|y_n(\delta)| < n\delta^{-n-1}C_2$, we have $|y_n(\delta)|k_n < \varepsilon\delta^{-n-1}C_2$; these two facts imply $|z_n(\delta)| < \delta^{-n-1}2\varepsilon C_2$. Therefore, there exists $\underline{\delta} > 0$, so that for all $\delta > \underline{\delta}$ and $n \leq N+1$ we have $|z_n(\delta)| < L$.

By construction we have for all $\delta \in [\underline{\delta}, 1]$,

(6.48)

$$\sum_{n=1}^{N+1} \delta^n z_n(\delta) = \sum_{n=1}^{N+1} \delta^n(k_n x_n + k_{n-1}y_{n-1}) = \sum_{n=1}^{N} \delta^n(k_n x_n + \delta k_n y_n) = \sum_{n=1}^{N} k_n W(\delta) = W(\delta).$$

Now, as we have $\delta \in \arg\max_{\widehat{\delta}} \delta^n(x_n(\widehat{\delta}) + \delta y_n(\widehat{\delta}))$ for each $n = 1, ..., N$ by Lemma 10, it follows that by the above equation that

$$\delta \in \arg\max_{\widehat{\delta}} \sum_{n=1}^{N} \delta^n(k_n x_n + \delta k_n y_n) = \arg\max_{\widehat{\delta}} \sum_{n=1}^{N+1} \delta^n z_n(\widehat{\delta}),$$

i.e., we can deliver the reward $W(\delta)$ to type δ in an incentive-compatible way while keeping each player's per-round adjustments in $[-L, L]$. □

Lemma 11 allows us to deliver rewards for punishment in an incentive-compatible way while keeping adjustments in an arbitrarily small $[-L, L]$. From there, the rest of the proof of Theorem 5 is standard.

Proof of Theorem 6. We start with a Tauberian result, which is used to show that if the average of the first T terms of a sequence of reals converge as $T \to \infty$, then the normalized discounted sum of the whole sequence converges to the same limit as $\delta \uparrow 1$. The following classic result, due to Frobenius, is our starting point.

Lemma 12 (Frobenius)**.** *For any sequence of reals $(x_t)_{t=0}^{\infty}$ satisfying*

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=0}^{t} x_t = x^* \in \mathbb{R},$$

we have $\lim_{\delta\uparrow 1} \sum_{t=0}^{\infty} \delta^t x_t = x^*$.

Corollary 1. If for a sequence of reals $(x_t)_{t=0}^{\infty}$ we have $\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} x_t = x^* \in \mathbb{R}$, then $\lim_{\delta\uparrow 1}(1-\delta) \sum_{t=0}^{\infty} \delta^t x_t = x^*$.

Proof. Defining $x_{-1} = 0$, we have

$$x^* = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} x_t = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=0}^{t} (x_k - x_{k-1}) \tag{6.49}$$

as $\sum_{k=0}^{t}(x_k - x_{k-1}) = x_t$. Using Lemma 12, we obtain

$$x^* = \lim_{\delta\uparrow 1} \sum_{t=0}^{\infty} \delta^t (x_t - x_{t-1}) = \lim_{\delta\uparrow 1} \left[\sum_{t=0}^{\infty} (\delta^t - \delta^{t+1}) x_t \right] = \lim_{\delta\uparrow 1} (1-\delta) \sum_{t=0}^{\infty} \delta^t x_t. \qquad \square$$

In light of this, if we can construct a Blackwell SPNE whose on-path play yields an undiscounted average payoff of v, then v is a limit Blackwell payoff. This leads to the proof of the folk theorem.

Fix $v \in F^{\mathrm{MI}}$. We wish to construct a strategy profile σ such that for any given $\varepsilon > 0$ there is some $\underline{\delta} < 1$ above which σ is a BE with discounted payoff within ε of v.

In view of Corollary 1, feasibility reduces to obtaining $\boldsymbol{v}$ as the limit of means of a sequence of pure action payoffs. While this is easy to do, some care is needed because in order for a sequence of actions to be an equilibrium path, we also need to make sure that the (discounted) continuation payoffs are individually rational; in fact our

insistence on Blackwell equilibria means that we should keep continuation payoffs of each $i \in I$ above the corresponding MI-minmax value $\underline{v}_i^{\text{MI}}$, not just the usual minmax $\underline{v}_i$.

The crux is that if the target payoff v is the discounted sum of pure-action payoffs, *i.e.*, of points in $C = g(A)$, all continuation payoffs are not necessarily individually rational even if v is. To circumvent this, we represent v as the discounted sum of individually rational payoffs. To this end, construct a full-dimensional set $D = \{d(1), ..., d(K)\}$, where $K \in \mathbb{N}$, such that (i) $v \in \text{int}\{\text{co}(D)\} \subset \text{co}(C)$; (ii) each $d(k) \in D$ is a rational convex combination of points in C; and (iii) there exists $\gamma > 0$ for which $d_i(k) > \underline{v}_i^{\text{MI}} + 3\gamma$ for all i and all $k = 1, 2, \ldots, K$. Since $v \in \text{co}(D)$, there exists a weight vector (non-negative components adding up to unity) $\lambda = (\lambda(k))_{k=1}^K$ such that $v = \sum_{k=1}^K \lambda(k) d(k)$. Let $(\lambda^m)_m$ be a sequence of weight vectors with rational components such that $\lambda^m \to \lambda$ as $m \to \infty$. Let $v^m := \sum_{k=1}^K \lambda^m(k) d(k)$. Since each $d(k)$ is a rational convex combination of points in C, a finite sequence ('subcycle k') from C averages to $d(k)$. Without loss of generality, these k subcycles have the same length.[24] Similarly we write v^m as a finite sequence ('cycle m') of points in D. Concatenate the cycles for $m = 0, 1, 2, ...$; then replace each occurance of $d(k)$ in each cycle by subcycle k to create the sequence $(x_t : t \geq 0)$ of payoff profiles in C, called the payoff path; the corresponding sequence of actions is the action path. Since there are finitely many distinct subcycles, choosing $\widetilde{\delta} < 1$ high enough ensures that for $\delta \geq \widetilde{\delta}$ the following two conditions hold—(i) the δ-discounted sum of any subcycle differs from the simple mean by at most γ; (ii) $(1 - \delta^L)M < \gamma$, where L is the maximum length of a subcycle and all individual payoffs of the stage game are in $[-M, M]$. Property (i) implies that any δ-discounted continuation payoff of the path from the start of any subcycle is at least $\underline{v}_i^{\text{MI}} + 2\gamma$; properties (i) and (ii) together imply that the continuation payoff of the path from any time (even when it is not the start of a subcycle) is at least $\underline{v}_i^{\text{MI}} + \gamma$.

The means of the cycles are v^m and since $\|v^m - v\| \to 0$, we have $\frac{1}{T+1} \sum_{t=0}^T x_t \to v$; Corollary 1 then implies that for any some $\widehat{\delta} \in (\widetilde{\delta}, 1)$ we have

$$\left\| v - (1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t \right\| \leq \varepsilon \; \forall \, \delta \geq \widehat{\delta}. \tag{6.50}$$

[24] If not, find the least common multiple L of the lengths of the subcycles, and repeat each of the subcycles the appropriate number of times to create K new subcycles each of length L.

The rest of the construction, as well as the proof that the resulting strategies constitute a Blackwell Equilibrium, follows the same path as the proof of Theorem 1.