BLACKWELL EQUILIBRIUM

COSTAS CAVOUNIDIS, SAMBUDDHA GHOSH, JOHANNES HÖRNER, EILON SOLAN, AND SATORU TAKAHASHI

ABSTRACT. We introduce 'Blackwell Equilibria'—applicable with slight modifications to various classes of discounted dynamic games, they require robustness to misspecification of discount factors. A Blackwell Equilibrium is a strategy profile that remains an equilibrium even if each player's discount factor is increased. We apply this to repeated games of perfect monitoring, and characterize the set of Blackwell SPNE payoffs via a novel minmax lying between the mixed and pure minmax values.

PRELIMINARY & INCOMPLETE VERSION: UNDER REVISION

This draft presents our results for games of perfect monitoring. We anticipate it will be enhanced by several results for the case of imperfect monitoring by June.

This paper merges the results obtained by Cavounidis and Ghosh with those by Hörner, Solan, and Takahashi. We wish to thank Kevin Cooke, Aniruddha Dasgupta, Kevin Lang, Bart Lipman, Francesco Nava, Jawwad Noor, Juan Ortner, Phil Reny, Takuo Sugaya and Balazs Szentes, as well as audiences at Boston University, the LSE, and the University of Warwick for useful comments and discussions. All errors remain our own.

1. Introduction

Repeated games constitute the simplest game-theoretic setting where concerns about the future of a repeated interaction can overcome myopic incentives to cheat. The canonical model with perfect monitoring is well studied. It involves a finite number of players choosing simultaneously from finite actions spaces; actions are revealed at the end of the period; total payoffs are defined as the discounted sum of the period payoffs. Best known among the repeated games results are the so-called folk theorems, which pin down the set of equilibrium payoff vectors as the common discount factor converges to unity. Friedman (1971) uses reversion to Nash equilibrium to produce 'trigger strategies' that can deliver any feasible average payoff better than some stage-game Nash equilibrium for each player. This result is expanded by Fudenberg and Maskin (1986), henceforth FM: Any payoff profile that is feasible and strictly individual rational (FSIR) in the stage-game is the payoff of a subgame-perfect Nash equilibrium (SPNE) of the infinitely repeated game if each player is sufficiently patient. From the best response property inherent in any Nash equilibrium it is clear that one cannot get strictly lower payoffs, making this a comprehensive theorem.

It is usual to assume in repeated games that the exact discount factors are known to all players and to the game theorist. In reality the game theorist who recommends a course of action, for example by designing mechanisms for repeated auctions, might not know the discount factor of the players exactly. If strategies are very sensitive to discount factors, small amounts of ignorance can have dramatic effects. Indeed, the strategies specified in FM cease to be equilibria if the analyst gets the discount factor even slightly wrong. We can also imagine that in real life players themselves cannot be certain that the others discount the future in exactly the same way as they do. Motivated by this we offer a solution concept that asks for equilibria to be robust to the discount factors. We identify when the game theorist can design strategies that form an equilibrium if all she can ensure is a lower bound on each player's discount factor. It is worth pointing out that our idea can be used to strengthen any solution concept that applies to discounted dynamic games; this paper does this for discounted infninitely repeated games, a natural starting point.

Kalai and Stanford (1988) provides a 'local' notion of discount robustness: a strategy profile is said to be a Discount-Robust Subgame Perfect (DRSP) equilibrium of a game if it remains an equilibrium for all sufficiently close discount factors. For games of perfect monitoring, we propose a stronger notion: Blackwell subgame perfect Nash equilibrium (Blackwell SPNE), where strategy profiles must be an SPNE if each player has a discount factor exceeding a cutoff. This draws inspiration from the work of Blackwell (1962), who shows that every finite Markov decision process has a strategy that is optimal at all high enough discount factors. This notion of robust optimality is now named 'Blackwell Optimality'. As our family of equilibrium concepts extends this idea to games, we name it in his honor as well.

Nash reversion strategies, introduced in Friedman (1971), can form Blackwell SPNE, but achieve only part of the FSIR set. On the other hand, folk theorems such as the one in FM that achieve the entire FSIR set as equilibrium payoffs are proven by constructing non-Blackwell SPNEs. Players who deviate are subjected to 'minmaxing', multiple periods where they are given their worst possible individually rational payoff, while players executing the punishment are rewarded. A player who is required to mix over actions with unequal stage-game payoffs should be rewarded more for using actions that give her lower stage-game payoffs, to make her indifferent over all pure actions while minmaxing the deviator. Such rewards are calculated using the exact discount factor—thus equilibria in FM need not be Blackwell SPNE.

One may wonder if, as the folk theorems were not designed with discount robustness in mind, an alternative technique could design robust SPNE to support any payoff in the FSIR set. This paper establishes that this is not so.

This is because we require robustness to an uncountable set of discount vectors[1]. The intuition for our results lies with the ability to calibrate punishments for deviations. If players' discount factors are fixed, players participating in the punishment of another can later be rewarded for mixing between actions that do not give the same stage-game payoff. However, if the discount factors may vary over an uncountable set, rewards cannot generate indifference under all possibilities. Therefore, certain useful kinds of punishments are impossible; as players will not accept payoffs that are not supported by worse punishments, they cannot get such payoffs in equilibrium. We thus show that the ability to punish, developed for (constructive) folk theorems, can therefore be thought of as a necessary condition, not merely a sufficient one, in thinking about equilibrium payoffs.

Blackwell equilibria are thus shown to require 'myopic indifference': if a player is required to mix at some history, then the support of that player's mixed action must offer the same current (not merely future!) expected payoffs, given what other

[1]This feature is shared by DRSP. Thus, even though it is a weaker notion of equilibrium, DRSP admits no larger a payoff set. Hence we focus on Blackwell equilibria.

players are doing. That is, mixing is *not* supportable by future rewards. This offers a strong behavioral implication for Blackwell equilibria: at any history, only action profiles in a finite 'myopically indifferent' set can be played. Using this, we show the existence of a critical value lying between the pure and mixed minmax such that feasible points that do not exceed this value cannot Blackwell SPNE payoffs.

Having precluded robust equilibria in a region of the FSIR, we set out to salvage the rest of the FSIR. We show that payoffs in (the interior of) the complement of the excluded region are always Blackwell SPNE payoffs. In particular, we prove a Blackwell folk theorem: for each such payoff, if players are sufficiently patient, there exist Blackwell equilibria that provide that payoff.

Curiously, the set of Blackwell SPNE payoffs coincides (up to a set of measure zero) with the set of payoffs that can be accomplished with Abreu (1988)'s Simple Strategy Profiles[2]. In fact, Simple Blackwell SPNEs are shown to be sufficient for the whole payoff set. This is despite the fact that the two notions are not nested.

2. Model

Fix a finite strategic-form stage-game $G = \langle I; (A_i)_i; (g_i)_i \rangle$, where I is the set of players $\{1, \ldots, n\}$, A_i is player i's finite set of actions, $A := \times_{i \in I} A_i$ is the set of all pure action profiles, and $g_i : A \to \mathbb{R}$ is player i's payoff function. A mixed action of i is $\alpha_i \in \triangle A_i$, where $\triangle E$ is the set of all probability distributions on a set E.

At each $t \in \mathbb{Z}_+$ the game G is played. Letting $\boldsymbol{a}^{(t)} \in A$ be the action profile actually chosen at time t,we denote the (public) history at the end of period t by $h^t = (\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(t)}) \in A^t =: H^t$ (starting with the empty history h^0). When player i discounts future payoffs using the discount factor δ_i, player i's average discounted utility defined over infinite sequences of pure actions in A is

$$u_i(\left\{ \boldsymbol{a}^{(t)} \right\}_{t=1}^{\infty}; \delta_i) := (1 - \delta_i) \sum_{t=1}^{\infty} \delta_i{}^{t-1} g_i \left(\boldsymbol{a}^{(t)} \right);$$

likewise the payoff $u_i(\sigma; \delta_i)$ of a strategy profile σ is defined as the expected utility of infinite action sequences, under the measure induced by the strategy profile. A pure strategy s_i of i comprises $s_i(t+1) : H^t \to A_i$ (for $t = 0, 1, \ldots$); for mixed strategies we replace A_i by $\triangle A_i$. This describes 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.

[2] Using a suitable extension of SSPs to mixed strategies.

Player i's minmax and pure-action minmax values, respectively, are

$$w_i := \min_{\boldsymbol{\alpha}_{-i} \in \times_j (\Delta A_j)} \max_{a_i \in A_i} g_i(a_i, \boldsymbol{\alpha}_{-i}), \quad w_i^p := \min_{\boldsymbol{a}_{-i} \in A_{-i}} \max_{a_i \in A_i} g_i(a_i, \boldsymbol{\alpha}_{-i}). \tag{2.1}$$

3. Blackwell SPNE

We are interested in identifying strategy and payoff profiles that can be supported under perfect monitoring using equilibrium notions that are robust to misspecification of discount factors in a sense made precise below. Blackwell (1956) showed that any finite[3] discounted Markov decision process has a stationary solution that is optimal above a certain cutoff discount factor. This is known as a 'Blackwell Optimal' solution. Motivated by this, our equilibrium concept, stated here for repeated games of perfect monitoring, asks for robustness to all discount factors above a cutoff.

Definition 1. A strategy profile σ is a **Blackwell SPNE** if there exists $\underline{\delta} \in [0,1)$ such that σ is an SPNE of $G^\infty(\boldsymbol{\delta})$ at any $\boldsymbol{\delta}$ with $\delta_i \geq \underline{\delta}$ for all $i \in I$.

We say that σ is a Blackwell SPNE above $\underline{\delta} \in [0,1)$ when $\underline{\delta}$ is such a sufficient discount factor cutoff. Our definition implies that if the game theorist knows that each player's discount factor exceeds a critical value, she can design strategy profiles from which no player would want to deviate unilaterally; Friedman (1971)'s folk theorem satisfies this criterion.

A few comments on the definition are in order.

- First, we could have chosen a more permissive defintion where the lower bound $\underline{\delta}_h$ above which $\sigma|_h$ is a Nash equilibrium can vary with h. However, this gives the same limiting results: all necessary conditions hold under the more permissive definition, while all sufficient conditions hold under the more stringent definition we adopt.
- Second, this is the natural multi-player version of the single-player notion of Blackwell Optimality. Indeed, if s is a Blackwell SPNE, then treating s_{-i} as fixed, player i faces a decision problem for which s_i is Blackwell Optimal.
- Third, there exists a literature studying Nash or SPNE arising from particular classes of strategies–reactive strategies in Kalai, Samet, and Stanford (1989), and finitely-complex ones in Kalai and Stanford (1988). Robustness with respect to nearby discount factors is invoked as an equilibrium refinement - which they term Discount Robust Subgame Perfect. Their primary concern is understanding not robustness to discount factors but equilibrium behavior

[3]Finite in terms of states and actions at each state

using relatively simple strategies. In contrast, we do not a priori rule out any strategy, and our focus is to construct strategies and payoffs robustly. A strategy profile that is a Blackwell SPNE must obviously be a DRSP equilibrium as in Kalai and Stanford (1988), but not vice versa.[4]

We now link payoffs and equilibria.

Definition 2. If σ is a Blackwell SPNE above $\underline{\delta}$ and $\boldsymbol{\delta} \geq \underline{\delta} \cdot \mathbf{1}$, we say $u(\sigma; \boldsymbol{\delta})$ is a **Blackwell SPNE payoff** at $\boldsymbol{\delta}$.

4. A NECESSARY CONDITION FOR BLACKWELL EQUILIBRIUM

The source of complication is that in order to punish a player the others may be required to mix over actions with different current payoffs, and the probabilities of mixing are unobservable. We define the set $\mathcal{X}$ of action profiles satisfying **myopic indifference**: Each player is indifferent over all pure actions he chooses with positive probability, given the others' actions. Let

$$\mathcal{X} = \left\{ \boldsymbol{\alpha} \in \prod_{i \in I} \Delta A_i \,\middle|\, \forall i \in I \forall a_i \in \operatorname{supp}(\alpha_i)[g_i(a_i, \boldsymbol{\alpha}_{-i}) = g_i(\boldsymbol{\alpha})] \right\}. \tag{4.1}$$

Note that myopic indifference does not imply that anyone is playing a best response. **A straightforward way to grasp $\mathcal{X}$ is that $\alpha \in \mathcal{X}$ iff α is a Nash Equilibrium of the stage-game restricted to α's support** $G_\alpha = \langle I; (supp(\alpha_i))_{i \in I}; (g_i)_{i \in I} \rangle$ **.** We now show that in Blackwell SPNE players must be myopically indifferent at each history.

Proposition 1. *If σ is a Blackwell SPNE, then $\sigma(h) \in \mathcal{X}$ for any public 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}, any player $i \in I$, and actions $a_i, a_i' \in \operatorname{supp}(\sigma_i(h^{t-1}))$. Let the expected stage-game payoff at time $\tau > t$ under the continuation strategy $\sigma|_{h^{t-1}}$ following the actions a_i and a_i' respectively (and no deviation at time τ by players other than i) be $g_j^{(\tau)}$ and $g_j'^{(\tau)}$. Since player i mixes over both these pure actions, they give the same total utility evaluated using any $\delta \in \mathcal{O}$:

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

[4]We choose not to focus on DRSP as our notion of discount robustness, as it is weaker than Blackwell SPNE but nevertheless produces the same payoff set. The reasons for this will become clear, shortly.

and hence

(4.2)
$$f(\delta_i) := g_i(a_i, \sigma_{-i}(h^{t-1})) - g_i(a_i', \sigma_{-i}(h^{t-1})) + \sum_{\tau > t} \delta_i^{\tau - t}(g_i^{(\tau)} - g_i'^{(\tau)}) = 0 \ \forall \delta_i \in \mathcal{O}.$$

The Identity Theorem implies that if the set of zeroes of an analytic function has an accumulation point in its domain, then it is identically zero; since (4.2) holds for an open interval, f must be identically zero in $(-1, 1)$, so that in particular we have by setting $\delta_i = 0$:

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

Thus, both a_i and a_i' give the same stage game payoff given opponents' actions; as this is true for all i, we have that $\sigma(h^{t-1}) \in \mathcal{X}$. Since h^{t-1} is arbitrary, this shows that myopically indifferent action profiles are played after *any* history. □

For generic g, we can show that $\mathcal{X}$ is a finite set, which pins down equilibrium actions quite sharply given this result. The (generic) finiteness of $\mathcal{X}$ depends crucially on mixing probabilities being unobservable; otherwise, we would have $\mathcal{X} = \times_i \triangle A_i$.

We note for later use that the proposition applies to all public histories and therefore to imperfect public monitoring.

5. Folk theorem for Blackwell Equilibria

The benchmark result under perfect monitoring is Fudenberg and Maskin (1986): Given any $v \in F^* \backslash \underline{\partial} F$, there exists a $\underline{\delta} \in (0, 1)$ such that for any $\delta > \underline{\delta}$ there is a subgame perfect Nash equilibrium giving each $i \in I$ a discounted average payoff v_i.[5] This does not preclude the equilibrium from depending on the exact value of δ and therefore failing to be a Blackwell SPNE. Our goal is to see which of these payoffs survive when we strengthen the equilibrium notion from SPNE to Blackwell SPNE.

Given that only action profiles in $\mathcal{X}$ can be played in a Blackwell SPNE, a lower bound on any player i's payoff can be given by the best-response property:

$$w_i^{\mathcal{X}} = \min_{\boldsymbol{\alpha} \in \mathcal{X}} \max_{a_i \in A_i} g_i(a_i, \boldsymbol{\alpha}_{-i}). \tag{5.1}$$

Thus $w_i^{\mathcal{X}}$ is the highest payoff i can get if all other players are using the worst possible actions for her, subject to the constraint that these actions form a myopically

[5] Abreu, Dutta, and Smith (1991) can support points on the lower boundary where each player is strictly above his minmax.

indifferent action profile when coupled with some action α_i for i (not necessarily an action that is a best reponse to the punitive mixed actions).

Note that $w_i^{\mathcal{X}}$ is (weakly) higher than the lowest payoff i attains in $\mathcal{X}$ and (weakly) lower than the highest payoff i attains in $\mathcal{X}$ when he is best responding. Since pure actions profiles trivially satisfy myopic indifference, we have[6]:

$$\times_i A_i \subseteq \mathcal{X} \subseteq \times_i \Delta A_i, \tag{5.2}$$

and therefore the $\mathcal{X}$-minmax lies between the pure and mixed ones:

$$\min_{\boldsymbol{a}_{-i}\in A_{-i}} \max_{a_i\in A_i} g_i(a_i,\boldsymbol{\alpha}_{-i}) \geq \min_{\boldsymbol{\alpha}\in\mathcal{X}} \max_{a_i\in A_i} g_i(a_i,\boldsymbol{\alpha}_{-i}) \geq \min_{\boldsymbol{\alpha}_{-i}\in\times_j(\Delta A_j)} \max_{a_i\in A_i} g_i(a_i,\boldsymbol{\alpha}_{-i}) \tag{5.3}$$

$$\Rightarrow \qquad w_i^p \geq w_i^{\mathcal{X}} \geq w_i. \tag{5.4}$$

In Section **??** we present an example where both inequalities in (5.4) are strict.

Proposition 1 implies that no player i earns strictly below $w_i^{\mathcal{X}}$ in a Blackwell SPNE. We can see this by contradiction. If such a player exists, from (4.1) it immediately follows that there is a (public) history at which the players are choosing an action profile not in $\mathcal{X}$, which contradicts Proposition 1. If a player needs to mix among actions that give her different current payoffs, she must be compensated later for taking a worse action today; this compensation is easy if the discount factor is known,[7] but it is not obvious how–if at all–such incentives can be provided for a range of discount factors simultaneously. As we have seen, such incentives cannot hold simultaneously for an *uncountable* set of discount factors, as they must for both DRSP equilibrium and Blackwell SPNE. Thus myopic indifference is necessary for robustness.

Consider the set of feasible payoffs (strictly) consistent with this restriction:

$$F^{\mathcal{X}} := \{\boldsymbol{x} \in F \mid x_i > w_i^{\mathcal{X}} \forall i \in I\}. \tag{5.5}$$

In view of the negative result above the best we can hope for is that all payoffs in $F^{\mathcal{X}}$ are Blackwell SPNE payoffs; the following theorem shows this to be true. Note that as $\times_i A_i \subseteq \mathcal{X}$, the theorem presented below includes all payoffs $\boldsymbol{v}$ where each player gets strictly above his pure minmax payoff.

Theorem 1. *Fix any discounted repeated game of perfect monitoring with or without PRDs. Every Blackwell equilibrium payoff $\boldsymbol{v}$ satisfies $v_i \geq w_i^{\mathcal{X}}$ for all $i \in I$. Conversely, if F is full-dimensional, for any vector $\boldsymbol{v} \in F$ satisfying $v_i > w_i^{\mathcal{X}}$ there exists a $\underline{\delta}$ such that if $\delta > \underline{\delta}$ then $\boldsymbol{v}$ is a Blackwell SPNE payoff at δ.*

[6]We're abusing notation to identify pure actions with their degenerate mixed equivalents.

[7]See, for example, Fudenberg and Maskin (1986).

The sufficiency part is a folk theorem, whose proof follows Abreu (1988) and FM 1986 in having n stick-and-carrot punishment regimes, one for each player. In these papers, as well as in ours, a unilateral deviation from the prescribed strategies leads to a 'stick-and-carrot regime'. In FM 1986, the stick phase is the minmax phase, during which the player who deviated gets his minmax payof while playing a myopic best-response to the punitive strategies of the others. Thereafter FM 1986 moves to the carrot phase, where all players earn above their respective minmax values.

To explain how our strategies do (and must) differ from FM 1986 we focus on the straightforward case where the target payoff $\boldsymbol{v}$ is the payoff of some pure action profile. A unilateral deviation by any player i is punished by playing the action profile $\alpha^i \in \mathcal{X}$ that solves (5.1). Note that player i must cooperate in his own punishment in the sense that he does not play a myopic best response. If i is mixing, then all such pure actions give him the same payoff; however, there might be actions outside the support that are better responses. Player i does not deviate to one of these, as that would lead to the punishment phase being prolonged, and is instead willing to accept a lower stage-game payoff than $w_i^{\mathcal{X}}$ during his stick phase because the subsequent carrot phase gives more than $w_i^{\mathcal{X}}$. In a Blackwell SPNE player i's action during his own punishment is calibrated to make the others myopically indifferent.

The second point gets to the heart of the Blackwell approach. If punishing a player requires a pure action profile, this is not relevant. But how do we ensure that a player j mixes appropriately when punishing i? In FM 1986 this is achieved by adjusting j's continuation payoffs after the minmax phase such that j gets the same total payoff from the beginning of i's minmax phase no matter what sequences of pure actions j plays during i's stick phase, provided of course that there is no observable deviation. Calculating the adjustment above requires exact knowledge of palyer j's discount factor; even if the discount factor is ever so slightly different from what was supposed, the value of the adjustment would give player j a strict incentive to not mix. Our Blackwell requirement has bite because it rules out such fine-tuning of strategies to discount factors.

Instead, we make j myopically indifferent over all pure actions in the support of his mixed action. Note that j's utility can vary across various realisations of the mixing, once the others' mixed actions resolve into pure actions. However, given that j was ex-ante indifferent, we ignore this: all possible pure actions in the support of the minmaxing action lead to the same continuation strategies, and therefore the same expected payoff for j. Thus ex-ante *myopic indifference* during the stick phase replaces the fine-tuning of continuation payoffs to ensure indifference.

What if $\boldsymbol{v}$ is not the payoff of a pure action profile (or a Nash equilibrium payoff of the stage game)? If a PRD is available, the above argument goes through unchanged, as on-path actions resolve into pure actions given an observed value of the PRD. Lacking a PRD, the concept of *self-accessibility* proposed in Dasgupta and Ghosh (2017) is used to construct pure strategies that deliver our target payoff while also keeping continuation payoffs near the target, thus ensuring that continuation payoffs are similar enough to be enforced by using a uniform punishment. 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 higher discount rates.* This is reminiscent of, and in fact related to, the Arrow-Levhari (1969) stopping theorem. Therefore, continuation payoffs remain 'near' at higher discounts, and hence we can preclude deviations at any higher discount with the same punishment. The proofs of this and all subsequent results can be found in the appendix.

6. Folk theorem for Limit Blackwell Payoffs

Now, as the analyst is concerned with robustness to the discount factors, she may adopt a different standpoint on payoffs. It could well be that instead of focusing on payoffs at any particular $\delta \in (0, 1)$, she is instead interested in the limit of the payoffs of Blackwell SPNEs as $\delta \uparrow 1$. We therefore posit the following definition:

Definition 3. A payoff $\boldsymbol{v}$ is said to be a **limit Blackwell payoff** (LBP) if there exists a Blackwell SPNE σ such that for any $\varepsilon > 0$ we can find a cutoff discount factor $\underline{\delta} \in (0, 1)$ such that $\mid u(\sigma; \boldsymbol{\delta}) - \boldsymbol{v} \mid < \varepsilon$ for all $\boldsymbol{\delta} \geq \underline{\delta} \cdot \mathbf{1}$.

It is worth pointing out that the cutoff discount factor is allowed to vary with ε, but the strategy profile is not.

It's not the case, however, that Blackwell SPNEs generally have payoffs that converge in the limit as $\delta \uparrow 1$. We therefore have to deploy a different approach than before to ensure both that our equilibria have payoffs that converge under this limit, and that the payoffs converge to the right vector. As it turns out, the set of LBPs coincides with those payoffs characterized in Theorem 1 (up to a set of measure zero).

Theorem 2. *Consider a discounted repeated game of perfect monitoring with or without PRDs. If the feasible set F is full-dimensional, every vector $\boldsymbol{v} \in F^{\mathcal{X}}$ is a limit Blackwell payoff.*

The proof proceeds largely along the same lines, only now our on-path play is carefully calibrated to provide payoffs converging to $\boldsymbol{v}$. Instead of using self-accessibility, we create sequences of pure actions that rationally approximate $\boldsymbol{v}$, while ensuring individual rationality at all high enough discount factors. Off-path play remains largely the same, with similar 'stick' and 'carrot' phases. Of course, the required δ to attain $\boldsymbol{v}$ as a LBP and as a Blackwell SPNE payoff may differ.

7. Remarks

We remark on full dimensionality of the payoff sets first, followed by an alternative approach that uses 'approximate minmaxing'.

Unlike in Fudenberg and Maskin (1986), full-dimensional payoff sets are indispensable in Theorem 1 even for two-player games, because we cannot resort to mutual minmaxing after a deviation. If we dispense with the full-dimensionality assumption, we need to replace the notion of restricted minmax by *effective restricted minmax*:

$$w_i^{\mathcal{X},\mathrm{eff}} := \min_{\boldsymbol{\alpha}\in\mathcal{X}} \max_{j\in I_i^+, a_j\in A_j} g_i(a_j, \boldsymbol{\alpha}_{-j}),$$

where I_i^+ (resp. I_i^-) be the set of players who have positively (resp. negatively) equivalent utilities with player i, i.e., the set of players $j \in I$ such that there exist $c_{ij} > 0$ (resp. < 0) and d_{ij} satisfying $u_j(\boldsymbol{a}) = c_{ij}u_i(\boldsymbol{a}) + d_{ij}$ for all $\boldsymbol{a} \in A$.

Unlike FM and our approach, Gossner (1995) shows that deviations may be deterred by using 'codes of conduct' that specify on-path play exactly, but leave strategies during the minmax phase unspecified to allow them to vary with the discount factor; at the end of a minmax phase a test statistic is computed to determine the likelihood that players did indeed approximately minmax the deviator, and if not, which player needs to be punished subsequently. It should not come as a surprise that although robust equilibria for $\boldsymbol{v} \in F \setminus \overline{F^{\mathcal{X}}}$ do not exist, one can construct codes of conduct with payoffs $\boldsymbol{v}$ such that only off-path play need vary with the discount factor. This further reinforces the basic intuition that discount robustness has bite because it affects the ability to punish deviations. Indeed we develop this approach formally in Section 9 by introducing the possibility of messaging at the onset of a punishment phase; this will allow us to support the FSIR set as Blackwell equilibrium payoffs, clarifying that the Blackwell restriction has bite because communication is disallowed.

8. Simplicity

Abreu (1988) gives a definition of **simple strategy profiles**. A pure straegy profile is **simple** if it can be specified as $n+1$ paths of play $(a^{i(\cdot)})_{i\in\{0,1,\ldots,n\}}$. Play starts along path a^0, and if any player i deviates, 'i's punishment path' a^i is played starting from the first term. When i deviates from a punishment path, either another player's or their own, their punishment is started from the first term. Abreu is motivated by a desire for strategies that admit a simple description, and shows that (pure) simple SPNEs are able to attain any payoff attainable in pure SPNEs.

We extend Abreu's definition of simplicity to mixed strategies: paths are now potentially mixed $(\alpha^{i(\cdot)})_{i\in\{0,1,\dots,n\}}$ and only observable deviations (i.e. deviations outside the support of the mixed action specified by the path) are punished.

Given this definition, the choice of pure action in the support of a mixed action does not affect continuation play, and therefore future payoffs; thus, a simple SPNE can only involve myopically indifferent action profiles at any history. Therefore, simple SPNE payoffs are bounded below by $w^{\mathcal{X}}$.

Myopic indifference is only a necessary condition for simple SPNE and Blackwell SPNE, and a sufficient one for neither. It is *not* the case that simple SPNEs necessarily have the Blackwell property. Similarly, Blackwell SPNEs are not necessarily simple, and can have a rich structure. It is then perhaps surprising that we can show the following:

Theorem 3. *If F is full-dimensional, every vector $\boldsymbol{v} \in F$ satisfying $v_i > w_i^{\mathcal{X}}$ for all $i \in I$ is a simple Blackwell SPNE payoff at any δ above some $\underline{\delta} \in (0,1)$, and a Limit Blackwell Payoff of a simple Blackwell SPNE.*

That is, it is not only the case that the sets of Blackwell payoffs and simple payoffs coincide, but also that each such payoff is attainable by an SPNE that is both simple and Blackwell. In a sense, the analyst can get two properties for the price of one. If Blackwellness is required, it suffices to use simple strategies. If simplicity is desired, discount robustness can also be delivered.

9. Discount-free Equilibrium

Our definition of Blackwell equilibrium chooses the same strategy above a cutoff discount factor, independently of the actual profile of discount factors. In some applications, it seems natural for a player's strategy to vary with her own discount factor. On the other hand, we would want robustness to players' beliefs about others' discounts. Thus, we wish to design strategies which depend on one's own discount, while insisting that they still specify best-responses following any history to the way opponents play given any discount they may draw. In this sense, we require that player i's knowledge or belief of the other players' discount factors is irrelevant for the optimality of her strategy.[8]

Definition 4. A **discount-free equilibrium recipe** above $\underline{\delta} \in (0,1)$ comprises n sets of strategies $(\{\sigma_i(\delta_i) \mid \delta_i \in [\underline{\delta}, 1)\})_{i \in I}$ such that for every public history $h \in H$ and every $\delta \in [\underline{\delta}, 1)^I$, for each player i, if i evaluates payoffs using discount δ_i, then $\sigma_i(\delta_i)|_h$ is a best-reply to $\sigma_{-i}(\delta_{-i})|_h$.

Note that under the notion of a recipe for equilibrium, each possible profile of discount factors induces a profile of strategies that is an equilibrium.

The next theorem shows that allowing such discount-free strategies is enough for the minmax folk theorem to go through provided players can communicate their discount factor. Formally let $\bar{G}$ denote the stage-game with the following temporal structure:

(1) a random number ω is publicly drawn from the uniform distribution on $[0,1)$;
(2) the simultaneous-move game G is played;
(3) each player publicly and simultaneously announces a number in $[0,1)$, which is purportedly his discount factor.

Let $\bar{G}_\infty(\underline{\delta})$ denote the game where a discount factor profile $(\delta_i)_{i \in I}$ is drawn at the start of the game once and for all, each player is informed of his own δ_i only, and thereafter the game $\bar{G}$ is repeated infinitely many times with perfect monitoring (of actions and messages) throughout. We note that, given that we have communication, the use of an explicit PRD is only for convenience: we could use richer messages to construct jointly controlled lotteries.

Theorem 4. *Fix $v \in \text{int } F$ satisfying $\forall i, v_i > w_i$. There is $\underline{\delta} < 1$ and a discount-free equilibrium plan $\prod_{i \in I}\{\sigma_i(\delta_i) \mid \delta_i \in [\underline{\delta}, 1)\}$ above $\underline{\delta}$ of the game $\bar{G}_\infty(\underline{\delta})$ such that for every $\delta \in [\underline{\delta}, 1)^I$ the payoff of the equilibrium strategy profile $(\sigma_i(\delta_i))_{i \in I}$ is v.*

This theorem shows that the standard folk theorem goes through provided each player's strategy is responsive to his own discount factor, and players are able to send cheap-talk messages from the unit interval (which will be identified with a report about one's own discount factor). Minmaxing players i requires requires an adjustment to the post-punishment continuation payoff to leave each player with the same utility for all possible choices he could make with positive probability during the punishment phase. These adjustment must in turn be 'fine-tuned' based on the

[8] Thus this definition coincides with that of belief-free equilibrium in the game of incomplete information where each player's type is his discount factor $\delta_i \in [\underline{\delta}, 1)$.

exact discount factor of each player; we must elicit this information from the players in an incentive-compatible manner following the punishment. The challenge is two-fold: this payoff adjustment must be feasible, and fine-tuned to each player's announced discount factor to enforce truthful reporting.

Consider an off-path history where the most recent punishment phase lasted from period $-T$ to 0. To support non-myopically-indifferent mixing during the punishment phase, we need to offer rewards to players depending on how they played. We first show that, following the end of the punishment and a player's announcement of her discount, the required rewards (evaluated at time 0) to provide mixing incentives to different realizations of the discounts $W(\cdot) : \delta^I \to \mathbb{R}_+$ can be chosen to be 'completely monotone'; that is, $W \geq 0$, and the derivatives starting with W' alternate between non-positive and non-negative.

Lemma 1. *Without loss, the function W can be chosen to be completely monotone.*

Now, having chosen a completely monotone W, we show we can deliver the reward $W(\delta)$ to a player with discount factor $\delta \in [\delta_0, 1)$ over any two successive periods n and $n+1$ in an incentive-compatible way; in order words, if we base the reward on the player's annoucement of her discount factor, she will announce truthfully. In the statement of the lemma below $x_n(\delta)$ denotes the payoff that δ would get at time n, while $y_n(\delta)$ denotes the payoff that she would get at $n+1$; the total discounted adjustment must equal the target adjustment $W(\delta)$.

Lemma 2. *Fix $\delta_0 > 0$ and a completely monotone map $W : [\delta_0, 1] \to \mathbb{R}$, with $\max\{W, -W'\} < C_1$. For each $n \in \mathbb{N}$ there exist functions $x_n, y_n : [0,1] \to \mathbb{R}$, such that, for all pairs $\delta, \hat{\delta} \in [0,1]$,*

$$W(\delta) = \delta^n(x_n(\delta) + \delta y_n(\delta)) \geq \delta^n(x_n(\hat{\delta}) + \delta y_n(\hat{\delta})). \tag{9.1}$$

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.

The above lemma is not suitable for delivering adjustments in our repeated game because the required payoffs may be either infeasible or feasible but not individually rational. The next lemma shows that. we can spread the adjustment over sufficiently many periods so that each adjustment is small. Thus, given a total adjustment $W(\delta)$ we pick a large number N of periods; 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 periods n and $n+1$ using Lemma 2 (each W_n thus plays the role of h in Lemma 2). 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.

Lemma 3. *Let $\delta_0 > 0$ and let $W : [\delta_0, 1] \to \mathbb{R}_+$ be a completely monotonic function. For all $M > 0$, there exists a $N \in \mathbb{N}$, $\underline{\delta} \in (\delta_0, 1)$, and a collection of functions $\{z_n\}_{1 \le n \le N}$ with each $z_n : [\underline{\delta}, 1] \to [-M, M]$ such that*

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

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

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 period adjustments small enough that each period's target payoff is within the FSIR set.

Proof. The proof is standard except that we need to check that announcements are truthful (IC), which is ensured by adjusting according to Lemma 3. We pick M small enough (and, correspondingly, $\underline{\delta}$ high enough) that post-punishment payoffs are made feasible and individually rational. □

10. CONCLUSION

We have shown that a sizable part of the FSIR, the set of payoffs proven to be equilibrium payoffs in folk theorems, cannot be supported by discount-robust strategies. Even a local notion of robustness is unattainable for equilibria with payoffs in that region. However, we have shown in a robust folk theorem that the entire complementary payoff region can be supported by both locally and strongly discount-robust equilibria. We do this by construction, designing simple strategy profiles that deliver any point in the robust region.

Furthermore, in the course of addressing the robustness question, we discover a new point of interest, the restricted minmax. We have shown that it describes the ability to punish when punishers' incentives cannot be finely tuned via future rewards for mixing. This intuition explains why certain regions of the payoff space cannot be robustly attained, and additionally, how it is possible to construct equilibria in the robustly feasible region.

Therefore, the present research establishes that care must be taken when discount factors are not precisely known. While we present a tool to get around this problem, it - and all other tools - may not always be applicable.

APPENDIX: PROOFS

Proof of Theorem 1. The first part, positing that it is necessary to have $v_i \geq w_i^{\mathcal{X}}$ for all i for $\boldsymbol{v}$ to be attainable in a Blackwell SPNE, is dealt with in the text. It remains for us to show the positive part of the theorem.

We start by deploying the technique of self-accessibility, which will help us target payoffs without a PRD, while respecting bounds on continuation payoffs that are uniform over high enough δs.

Definition 5. Fix a finite set $C \subset \mathbb{R}^I$. A set $E \subset \mathbb{R}^I$ is self-accessible relative to C at the discount profile $\boldsymbol{\delta}$ if for each $\boldsymbol{x} \in E$ there exists $\boldsymbol{c} \in C$ and $\boldsymbol{y} \in E$ such that $\forall i, x_i = \delta_i c_i + (1 - \delta_i) y_i$.

For $\boldsymbol{x} \in \mathbb{R}^n$ and $\varepsilon \in \mathbb{R}_+$, the closed n-dimensional ball of radius ε centered on $\boldsymbol{x}$ is denoted by $B(\boldsymbol{x}, \varepsilon)$. The finite set C may be thought of as the set of pure-action payoffs (in $\mathbb{R}^n$) of the stage game.

Theorem 5 (Self-Accessibility (Dasgupta and Ghosh 2017))**.** *For all $\boldsymbol{v} \in \mathbb{R}^N$ and $\varepsilon > 0$ such that $B(\boldsymbol{v}, \varepsilon) \subset int(F)$, there exists $\delta > 0$ such that $\forall \boldsymbol{\delta'} > \delta \cdot \mathbf{1}$ there exists a subset E of A that contains v and is self-accessible at $\boldsymbol{\delta'}$. Furthermore, under equal discounting closed balls in the interior of F or the relative interior of any of its k-faces, are self-accessible.*

This is useful because of the following result that follows by 'iteratively applying' self-acccessibility: if $\boldsymbol{v}$ lies in F and $\gamma > 0$, there exists $\delta_* > 0$ such that for any $\delta \geq \delta_*$ there is a sequence of action profiles $(\boldsymbol{a}^{(t)} : t \geq 0) =: a(\boldsymbol{v}, \gamma, \delta)$ such that

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

However, when strategies are designed without knowledge of the discount factors, we need to know how continuation payoffs of a given sequence of actions change with an increase in the discounts. The next lemma answers this by showing that if all δ-discounted continuation payoffs of a sequence remain above $y \in \mathbb{R}$, the same is true at higher discount factors. Therefore, using self-accessibility to create paths with continuation payoffs bounded below and continuation payoffs on deviation bounded above by a lower quantity, we can get discount robustness if the initial discount factor is high enough.

Lemma 4 (Patience Lemma). *Fix $\delta \in (0,1)$, a sequence of real numbers $(x^{(t)})_{t\in\mathbb{Z}_+}$ and $y \in \mathbb{R}$ such that*

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

Then the same inequalities hold for any $\delta' \in (\delta, 1)$.

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

$$f(\delta', t) := (1-\delta')\sum_{\tau=t}^{\infty}\delta'^{\tau-t}(x^{(\tau)} - \underline{x}) \tag{10.1}$$

which is ex hypothesi weakly positive when $\delta' = \delta$. For any $\delta' > \delta$, we have

$$f(\delta', t) = (1-\delta')\sum_{s=t}^{\infty}\left[\delta^{s-t} + \delta^{s-t}\left(\left[\frac{\delta'}{\delta}\right]^{s-t} - 1\right)\right](x^{(s)} - \underline{x})$$

$$f(\delta', t) = (1-\delta')\sum_{s=t}^{\infty}\left[\delta^{s-t} + (\delta' - \delta)\delta^{s}\delta'^{-t-1}\sum_{\tau=t+1}^{s}(\frac{\delta'}{\delta})^{\tau}\right](x^{(s)} - \underline{x})$$

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

Substituting in $f(\delta, t)$ and $f(\delta, \tau)$ we obtain

$$f(\delta', t)\ := \frac{1-\delta'}{1-\delta}f(\delta, t) + \frac{1-\delta'}{1-\delta}(\delta'-\delta)\sum_{\tau=t+1}^{\infty}\delta'^{\tau-t-1}f(\delta, \tau). \tag{10.2}$$

But this last formulation is weakly positive from $\delta' - \delta \geq 0$, and (10.1). This leads to one of the two desired inequalities; the other follows from a similar reasoning. $\square$

We can express $f(\delta', 0)$ as a weighted average of $f(\delta, \tau)$ at different τs with non-negative weights summing to 1. This is similar to the Arrow-Levhari (1969) stopping theorem which states that 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.

Finally, we can attack the theorem constructively.

Proof. Fix $\boldsymbol{v} \in F^{\mathcal{X}}$. Following Abreu, Dutta, and Smith (1991) pick player-specific punishments[9] $\{\boldsymbol{v}(i) \in F^{\mathcal{X}} \mid 1 \leq i \leq n\}$ and $\varepsilon > 0$ such that

$$\forall i,\ w_i^{\mathcal{X}} + 2\varepsilon < v_i(i) \tag{10.3}$$

$$v_i(i) + 3\varepsilon < v_i; \tag{10.4}$$

$$\forall i \neq j,\ v_i(i) + 3\varepsilon < v_i(j). \tag{10.5}$$

As usual, we interpret each $v(i)$ as giving each $j \neq i$ a 'reward' for punishing i. As before, let

$$\boldsymbol{\alpha}^{\boldsymbol{i}} \in \arg\min_{\boldsymbol{\alpha} \in \mathcal{X}} \max_{a_i} g_i(a_i, \boldsymbol{\alpha}_{-i}). \tag{10.6}$$

First, we want to preclude deviations from the reward phase following one's punishment. Choose $N \in \mathbb{N}$ s.t.

$$\max_{\boldsymbol{a} \in A} g_i(\boldsymbol{a}) + N\, g_i(\boldsymbol{\alpha}^{\boldsymbol{i}}) < (N+1)(v_i(i) - \varepsilon)\ \forall i, \tag{10.7}$$

which is possible because $g_i(\alpha^i) < w_i^{\mathcal{X}}$ and (10.3) holds.

Take $\underline{\delta} < 1$ such that for i and all $\delta \geq \underline{\delta}$ the following conditions hold:

(SA) The closed balls $B(\boldsymbol{v}, \varepsilon)$ and $\{B(\boldsymbol{v}(i), \varepsilon) : 1 \leq i \leq n\}$ are self-accessible at δ.

This is true for all high enough δ by Theorem 5.

(IC-I)

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

which in the limit as $\delta \uparrow 1$ reduces to $v_i(i) + \varepsilon < v_i - \varepsilon$ and is therefore true for high enough δ by (10.4).

(IC-II(i))
$$w_i^{\mathcal{X}} < (1-\delta^N) g_i(\alpha^i) + \delta^N (v_i(i) - \varepsilon)$$

which in the limit as $\delta \uparrow 1$ reduces to $w_i^{\mathcal{X}} < v_i(i) - \varepsilon$ and is therefore true for high enough δ by (10.3).

(IC-II(j))

$$\forall T \leq N, (1-\delta) \max g_i(a) + \delta[(1-\delta^N) g_i(\alpha^i) + \delta^N (v_i(i) + \varepsilon)] < (1-\delta^T)(g_i(\alpha^j) + \delta^T (v_i(j) - \varepsilon)$$

[9]The proof in FM excludes the lower boundary of the feasible strictly individually set. Abreu, Dutta, and Smith weaken the full-dimensionality (FD) condition to a non-equivalent utility condition (NEU). Both FD and NEU ensure the existence of the player-specific punishments we construct.

which in the limit as $\delta \uparrow 1$ reduces to $v_i(i) + \varepsilon < v_i(j) - \varepsilon$ and is therefore true for high enough δ by (10.5).

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

which in the limit $\delta \uparrow 1$ is equivalent to (10.7) and is therefore true for high enough δ.

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

which in the limit as $\delta \uparrow 1$ reduces to $v_i(i) + \varepsilon < v_i(j) - \varepsilon$ and is therefore true for high enough δ by (10.5).

Strategies.

We construct a simple strategy profile *à la* Abreu (1988): If j deviates unilaterally from any phase except Phase $III(j)$, impose *Phase II(j)* followed by *Phase III(j).* We differ from the standard folk theorems such as FM 1986 and follow Dasgupta and Ghosh (2017) in one respect: After player j deviates, we impose *Phase II(j)* and then enter *Phase III(j)*, continuing where we left off if *Phase III(j)* was active in the past, rather than restarting the final phase; this is needed because at the actual discount factor the payoff from the beginning of the phase might be higher than a subsequent continuation payoff.

For a given $\delta > \underline{\delta}$, our strategies specify:

Phase I: Play $a(\boldsymbol{v}, \varepsilon, \delta)$.

Phase II(i): Play $\boldsymbol{\alpha}^{\boldsymbol{i}}$ for N periods.

Phase III(i): Play $a(\boldsymbol{v}(i), \varepsilon, \underline{\delta})$.

As we used Self-Accessibility 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 $t-1$ periods of the correspnding phase (not of the entire game) have elapsed. Note that the strategy, including parameters in it, is set independently of the actual discount factor(s) δ'.[10] Also note that we do not ask a player to (myopically) best respond during her own punishment phase, as that would not leave the others willing to mix.

[10] In contrast, the third phase in FM depends on the exact discount factor when mixing is involved.

Checking subgame perfection.

Step 1. Player i does not deviate from *Phase I* if for any progression in *Phase I* t,

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

By Self-Accessibility, for any t we have that $v_i^t(\delta) \geq v_i - \varepsilon$, whence the Patience Lemma implies that $v_i^t(\delta') \geq v_i - \varepsilon$ for any $\delta' \geq \delta$; similarly $v_i^1(i)(\delta') \leq v_i(i) + \varepsilon$ for any $\delta' \geq \delta$. Hence for the inequality to hold it is enough to show that

$$(1-\delta')\max_{\boldsymbol{a}\in A} g_i(\boldsymbol{a}) + \delta'[(1-\delta'^N)w_i^{\mathcal{X}} + \delta'^N(v_i(i)+\varepsilon)] \leq (1-\delta')\min_{\boldsymbol{a}\in A} g_i(\boldsymbol{a}) + \delta'(v_i - \varepsilon);$$

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

Step 2. Player i does not deviate from *Phase II(i)* if for any progression in *Phase III(i)* t,

$$(1-\delta')w_i^{\mathcal{X}} + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N v_i^t(i)(\delta')] < (1-\delta'^N)g_i(\alpha^i) + \delta'^N v_i^t(i)(\delta').$$

We can again use Self-Accessibility and the Patience Lemma to bound $v_i^t(i)(\delta')$ below by $v_i(i) - \varepsilon$ and get the sufficient condition

$$(1-\delta')w_i^{\mathcal{X}} + \delta'[(1-\delta'^N)g_i(\alpha^i) + \delta'^N(v_i(i)-\varepsilon)] < (1-\delta'^N)g_i(\alpha^i) + \delta'^N(v_i(i)-\varepsilon)$$

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

Step 3. Player i does not deviate from *Phase III(i)* if for any progression in it t,

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

We can rearrange the inequality as:

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

Using the same technique, we can bound $v_i^t(i)(\delta')$ below by $v_i(i) - \varepsilon$ and therefore use (IC-III(i)) as a sufficient condition, 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 periods $T \leq N$ and any progression in *Phase III(i)* t and *Phase III(j)* τ,

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

We can again use $v_i(i) + \varepsilon$ as an upper bound for $v_i^t(i)(\delta')$ and $v_i(j) - \varepsilon$ as a lower bound for $v_i^\tau(j)(\delta')$. Then, the inequality reduces to (IC-II(j)) which applies as $\delta' \geq \delta \geq \underline{\delta}$.

Step 5. Although player i does not deviate observably we need to show that she mixes as required in *Phase II(j)*. Recall our definition of $\mathcal{X}$; since $\boldsymbol{\alpha}^j \in \mathcal{X}$, mixing only occurs between myopically indifferent actions according to $\boldsymbol{\alpha}^j$; as future play does

not vary over i's actions on $\text{supp}(\alpha_i^j)$, she is indifferent.
Step 6. Player i does not deviate from *Phase III(j)* if for any progression in *Phase III(i)* t and *Phase III(j)* τ,

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

Using the same logic as in *Step 4.* 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. □

Note that with a PRD the equilibrium strategies can be modified as follows:
Phase I: At each time play the correlated action $p \in \triangle A$ such that $\boldsymbol{v} = \sum_{\boldsymbol{a}\in A} p(\boldsymbol{a})g(\boldsymbol{a})$.
Phase II(i): Play $\boldsymbol{\alpha^i}$ for N periods.
Phase III(i): Play $p^i \in \triangle A$ such that $\boldsymbol{v(i)} = \sum_{\boldsymbol{a}\in A} p^i(\boldsymbol{a})g(\boldsymbol{a})$ at each period.

It is easy to see that these strategies must be a Blackwell SPNE above $\underline{\delta}$ if the PRD-free strategies are. Furthermore, under this modified strategy profile we have $u(\sigma, \boldsymbol{\delta}) = \boldsymbol{v}$ for all $\boldsymbol{\delta} \geq \underline{\delta} \cdot \mathbf{1}$.

Proof of Theorem 2. First, we state a Tauberian theorem of immediate use:

Lemma 5. [Frobenius 1880] *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 $\lim_{\delta\uparrow 1} \sum_{t=0}^\infty \delta^t x_t = x^*$.

We use this to show that should 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 $\delta \uparrow 1$.

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

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

$$x^* = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} x_t = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} \sum_{k=0}^{t} (x_k - x_{k-1}) \tag{10.8}$$

as $\sum_{k=0}^t (x_k - x_{k-1}) = x_t$. Using Lemma 5, we obtain

$$x^* = \lim_{\delta\uparrow 1} \sum_{t=0}^\infty \delta^t (x_t - x_{t-1}) = \lim_{\delta\uparrow 1} \left[\sum_{t=0}^\infty (\delta^t - \delta^{t+1}) x_t \right] = \lim_{\delta\uparrow 1} (1-\delta) \sum_{t=0}^\infty \delta^t x_t. \tag{10.9}$$

□

In light of this, if we can construct a Blackwell SPNE whose on-path play yields an undiscounted average payoff of $\boldsymbol{v}$, then $\boldsymbol{v}$ is a LBP.

We are now ready to present the proof of the LBP folk theorem.

Proof. Fix $\boldsymbol{v} \in F^{\mathcal{X}}$. 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 $\boldsymbol{v}$.

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 $\mathcal{X}$-minmax value $w_i^{\mathcal{X}}$, not just the usual minmax w_i. In general, continuation payoffs of sequences in C need not even be individually rational.

Starting from the set $C = g(A)$ of pure-action payoffs, construct a full-dimensional set $D = \{\boldsymbol{d}(1), ..., \boldsymbol{d}(K)\}$ such that $\boldsymbol{v} \in \text{int}\{\text{co}(D)\} \subset \text{co}(C)$, each $\boldsymbol{d}(k) \in D$ is a rational convex combination of the points in C, and there is some $\gamma > 0$ for which $d_i(k) > w_i^{\mathcal{X}} + 3\gamma$. Clearly we find convex weights λ so that $\boldsymbol{v} = \sum_{k=1}^{K} \lambda(k)\boldsymbol{d}(k)$. For each $k = 1, ..., K-1$ define a sequence of convex rational weights $\lambda_m(k) \to \lambda(k)$, and define

$$\lambda^m(K) := 1 - \sum_{k=1}^{K-1} \lambda^m(k),\ \boldsymbol{v}^m := \sum_{k=1}^{K} \lambda^m(k)\boldsymbol{d}(k).$$

Since $\boldsymbol{d}(k)$ is a rational convex combination of points in C, a finite sequence ('subcycle k') from C averages to $\boldsymbol{d}(k)$. Without loss of generality, these k subcycles have the same length[11]. Similarly we write $\boldsymbol{v}^m$ as a finite sequence ('cycle m') of points in D. Concatenate the cycles for $m = 0, 1, 2, ...$; then replace each occurance of $\boldsymbol{d}(k)$ in any cycle by subcycle k to create the sequence $(\boldsymbol{x}_t : t \geq 0)$ of payoff profiles in C, called the payoff path; the corresponding sequence of actions is simply the 'path'. Since the are finitely many distinct subcycles, choosing $\tilde{\delta} < 1$ high enough ensures that for $\delta \geq \tilde{\delta}$ 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

[11]This can be achieved by finding the least common multiple of subcycles' lengths and creating new subcycles that repeat the corresponding old subcycles the appropriate number of times so that their lengths equate.

in $[-M, M]$. Property (i) implies that any δ-discounted continuation payoff of the path from the start of any subcycle is at least $w_i^{\mathcal{X}} + 2\gamma$; 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 $w_i^{\mathcal{X}} + \gamma$.

The means of the cycles is $\boldsymbol{v}^m$ and since $\| \boldsymbol{v}^m - \boldsymbol{v} \| \to 0$, we have $\frac{1}{T+1} \sum_{t=0}^{T} \boldsymbol{x}_t \to \boldsymbol{v}$; Corollary 1 then implies that for some $\hat{\delta} \in (\tilde{\delta}, 1)$ we have

$$\left\| \boldsymbol{v} - (1-\delta) \sum_{t=0}^{\infty} \delta^t \boldsymbol{x}_t \right\| \leq \varepsilon, \ \delta \geq \hat{\delta}. \tag{10.10}$$

Pick player-specific punishments $\{\boldsymbol{v}(i) : 0 \leq i \leq n\}$ in $F^{\mathcal{X}}$ as before and define the strategy profile σ by the following 'automaton description', which mirrors our earlier theorem except for the construction of the path:

Phase I: Play the action path above.

Phase II(i): Play $\boldsymbol{\alpha}^{\boldsymbol{i}} \in \mathcal{X}$ for N periods, where N is long enough to wipe out gains from a single deviation.

Phase III(i): Play $a(\boldsymbol{v}(i), \gamma, \hat{\delta})$.

Unimprovability: The same arguments as in the earlier proof show that σ must be a Blackwell SPNE above some $\underline{\delta} \geq \hat{\delta}$.

□

Proof of Theorem 3.

Proof. To achieve a particular $\boldsymbol{v} >> \boldsymbol{w}^{\mathcal{X}}$ in simple strategies, we edit the proofs of Theorems 1 and 2 in the following ways:

(1) Deviations during the 'carrot' phase were treated differently according to when they occur in the original proof - so we'll need to edit the 'carrot' phases to get a simple strategy profile. We begin by constructing, for each i, a finite sequence of pure action profiles $(a_t'^i)_{t \leq T_i}$ whose average payoff (undiscounted) is within ε of $v(i)$. The carrot phase during i's punishment will now consist of repeating the sequence a'^i - regardless of the history prior to the last deviation.

(2) We lengthen the 'stick' phases enough so that from any point in any 'carrot' phase, deviations are not profitable in undiscounted terms.

(3) We increase the bound on the discount vector enough so that going along with these longer 'stick' phases and their subsequent 'carrot' phases becomes individually rational.

On-path play is calculated in exactly the same ways as before, using the Self-Accessible construction for the simple Blackwell folk theorem and the cycles construction to achieve v as a simple Limit Blackwell payoff. We can now describe the punishments (both the 'carrot' and 'stick' parts) as sequences of action profiles, so that given the pure nature of on-path play, we can write our equilibria as Simple Strategy Profiles. □

Proof of Lemma 1.

Proof. Denote by 0 the period 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 period t while minmaxing j. We can choose a reward structure so that $\xi_t > 0$ whatever$i$'s realized action at time 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. □

Proof of Lemma 2. .

Proof. 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 $W(\delta) = \delta^n(x_n(\delta) + \delta y_n(\delta))$. Now, for any $\hat{\delta} > \delta$ the fundamental theorem of calculus gives

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

where the inequality follows from the complete monotonicity of W. From here, equation (9.1) follows immediately for $\hat{\delta} > \delta$. A symmetric argument establishes the same property for $\hat{\delta} < \delta$. The last assertion of the lemma is immediate given the definitions of x and y, and the bound on $W, -W'$. □

Proof of Lemma 3.

Proof. Given C_2 as in Lemma 2, choose ε so that $2\varepsilon C_2 < M$. Let $N = \inf\{\hat{N} | \sum_{n=1}^{\hat{N}} \frac{\varepsilon}{n} > 1\}$, 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 $k_N = 1 - \sum_{n=1}^{N-1} k_n$. Our proof strategy is to 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 2 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 period 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 2 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 $|z_n(\delta)| < M$.

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

(10.11)
$$\sum_{n=1}^{N+1} \delta^n z_n(\delta) = \sum_{n=1}^{N+1} \delta^n(k_n x_n + k_{n-1}y_{n-1}) = \sum_{n=1}^{N} \delta^n(k_n x_n + \delta k_n y_n) = \sum_{n=1}^{N} k_n W(\delta) = W(\delta).$$

Now, as we have $\delta \in \arg\max_{\hat{\delta}} \delta^n(x_n(\hat{\delta}) + \delta y_n(\hat{\delta}))$ for each $n = 1, ..., N$ by Lemma 2, it follows that by the above equation that

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

i.e. we can deliver the reward $W(\delta)$ to type δ in an incentive-compatible way while keeping per-period adjustments in $[-M, M]$. □

References

[1] Abreu, Dilip. "On the theory of infinitely repeated games with discounting." Econometrica: Journal of the Econometric Society (1988): 383-396.

[2] Blackwell, David. "Discrete Dynamic Programming." The Annals of Mathematical Statistics 33 (1962): 719-726

[3] Arrow, Kenneth and David Levhari. "Uniqueness of the Internal Rate of Return with Variable Life of Investment." Economic Journal 79 (1969): 560-566

[4] Dasgupta, Aniruddha, and Sambuddha Ghosh. Folk Theorems without Public Randomization: A New Approach.

[5] Friedman, James W. "A non-cooperative equilibrium for supergames." The Review of Economic Studies 38.1 (1971): 1-12.

[6] Frobenius, Georg.: "Über die Leibnizsche Reihe." Journal für die reine und angewandte Mathematik 89, 262-264 (1880).

[7] Fudenberg, Drew, and Eric Maskin. "The folk theorem in repeated games with discounting or with incomplete information." Econometrica: Journal of the Econometric Society (1986): 533-554.

[8] Gossner, Olivier. "The folk theorem for finitely repeated games with mixed strategies." International Journal of Game Theory 24, no. 1 (1995): 95-107.

[9] Kalai, Ehud, and William Stanford. "Finite rationality and interpersonal complexity in repeated games." Econometrica: Journal of the Econometric Society (1988): 397-410.

Current address: Cavounidis: Department of Economics, University of Warwick, ccavouni@bu.edu; Ghosh: SOE, Shanghai University of Finance and Economics, sghosh@mail.shufe.edu.cn; Hörner: Yale University; Solan: Tel-Aviv University; Takahashi: National University of Singapore.