Uniform Folk Theorems in Repeated Anonymous Random Matching Games[*]

Joyee Deb

Stern School of Business

New York University

Julio González Díaz

Department of Statistics and Operations Research

University of Santiago de Compostela

Jérôme Renault

Toulouse School of Economics

Université Toulouse 1

Abstract

We study infinitely repeated anonymous random matching games played by communities of players, who only observe the outcomes of their own matches. It is well known that cooperation can be sustained in equilibrium for the prisoner's dilemma (PD) through grim trigger strategies. Little is known about games beyond the PD. We study a new equilibrium concept, *strongly uniform equilibrium (SUE*, which refines the notion of uniform equilibrium (UE) and has additional properties such as a strong version of (approximate) sequential rationality. We establish folk theorems for general games and arbitrary number of communities. Interestingly, the equilibrium strategies we construct are easy to play. We extend the results to a setting with imperfect private monitoring, for the case of two communities. We also show that it is possible for some players to get equilibrium payoffs that are outside the set of individually rational and feasible payoffs of the stage game. In particular, for the PD we derive a bound on the number of "free-riders" that can be sustained in society. A by-product of our analysis is an important result relating uniform equilibrium and strongly uniform equilibria: we show that, in general repeated games with finite players, actions, and signals, the set of UE and SUE payoffs coincide.

Keywords. ANONYMOUS RANDOM MATCHING, UNIFORM EQUILIBRIA, REPEATED GAMES

1

This version: November 5, 2013

1 Introduction

We study the feasibility of cooperation in large communities, in which members of society interact repeatedly over time. In large communities, players may not recognize each other, and often do not observe everyone's actions. For example, consider a large market in which people trade (bilaterally) with strangers. The central question of this paper is whether people will cooperate or cheat in such anonymous transactions. We model these interactions as an infinitely repeated anonymous random matching game (RARMG) in which, in every period, players from different communities are anonymously and randomly matched to each other to play a stage game. Each player observes only the actions played in his own match: so he does not receive any information about the identity of his opponents or about how the other players have been matched and what actions they have played. We ask what payoffs can be achieved in equilibrium in RARMG.

Seminal papers by Kandori (1992) and Ellison (1994) show that cooperation can be sustained in the case of the prisoner's dilemma (PD). This is done by way of community enforcement, which means that desirable behavior or cooperation is sustained because deviating against one agent triggers sanctions by others. Such an equilibrium, prescription of desirable behavior and appropriate sanctions, is often interpreted as a social norm. How is such a social norm sustained by rational agents? In the PD, if a player ever faces a defection, he punishes all future rivals by switching to defection forever. After a defection, more and more people realize that there has been a deviation and start defecting as well. This information spreads until, eventually, the whole population is defecting. The credible threat of such a breakdown of cooperation can deter players from defecting in the first place. Yet, the arguments in the aforementioned papers exploit the fact that, in the PD, the Nash equilibrium of the stage-game is in strictly dominant actions. Thus, although punishing may lower continuation payoffs, it always gives a short-term gain. In general games, it is harder to provide incentives to "sanction" or punish, since doing so can both lower future continuation payoffs and entail a short-term loss. An open question is whether cooperation can be sustained beyond the PD.

We establish several strong possibility results about cooperation in the RARMG environment. Our main departure from existing literature lies in the solution concept, *strongly uniform equilibrium*, which we discuss in more detail below. We establish that given *any* stage game played by *any* number of communities, it is possible to sustain *any* individually rational and feasible payoff in a strongly uniform equilibrium. This is the first folk theorem in a RARMG environment, with no non-trivial assumptions on stage-game payoffs and without adding any informational assumptions.[1]

---

[1] Deb and González-Díaz (2013) establish a possibility result that applies to a relatively large class of two-player games beyond the PD. Other papers that go beyond the PD introduce verifiable information about past play to sustain cooperation. For instance, Kandori (1992) assumes the existence of a mechanism that assigns labels to players based on their history of play: players who have deviated or have seen a deviation can be distinguished from those who have not, by their labels. For related approaches, see Dal Bó (2007), Hasker (2007), Okuno-Fujiwara and Postlewaite (1995), and Takahashi (2010). More recently, Deb (2012) obtains a general folk theorem for any game by allowing cheap talk.

We also obtain a folk theorem for the setting with two communities and imperfect private monitoring within a match. Specifically, we consider a setting in which players do not perfectly observe their rival's action, but receive a noisy signal about it. We identify a mild sufficient condition on the signal structure under which the folk theorem obtains.

Our analysis also leads to some qualitative observations about the payoffs that can be sustained in equilibrium. First, we show that there can be players in the community who get equilibrium payoffs that lie outside the set of individually rational and feasible payoffs of the stage game.[2] We do so by establishing a bound on the number of members of one community that can be "allowed" to defect whilethe rest of society cooperates. We can interpret such an equilibrium as a social norm that sustains a proportion of free riders in society. Second, we also investigate whether there is room for using correlated punishments in our setting, presenting examples of equilibria in which some members of society are held strictly below their independent minmax payoff.

A substantial part of the contribution of this paper lies in the equilibrium concept, strongly uniform equilibrium, a refinement of the notion of uniform equilibrium (Sorin, 1986; Fudenberg and Levine, 1991) in which players' strategies are uniformly approximate best responses to their rivals' strategies and, moreover, have several extra features. One of them is a strong version of sequential rationality after every history in which all the payoffs (including those after deviations) are obtained almost surely and not just in expectation. Maybe more importantly, the equilibrium strategies we construct are easy to play and, therefore, it is plausible to think that they can be followed by economic agents.[3] In a nutshell, the strategies consist of (1) cooperative blocks, where the players play the actions that lead to the chosen target payoff, (2) testing phase, where, if needed, the players run simple statistical tests to decide whether or not someone has deviated, and (3) punishment blocks, to be played after a test has detected a deviation.[4]

Interestingly, we also make an important technical contribution to the literature on uniform equilibrium in standard repeated games. We show that while strongly uniform equilibrium is a strict refinement of uniform equilibrium in terms of strategies, the two solution concepts are identical in terms of payoff sets. This result applies not just to our particular random matching environment, but to any infinitely repeated game with general monitoring structure. An important consequence is that, for any such game, any existing folk theorem or result obtained for the set of uniform equilibrium payoffs immediately extends to the stronger notion of strongly uniform equilibrium, *i.e.*, the payoffs can also be sustained by strategies that satisfy several additional desirable features. As a crucial tool to prove this and other results in this paper, we elaborate on a lemma by Lehrer

---

[2]Dal Bó (2007) makes a similar observation in a related setting.

[3]The strategies remain simple even in the case in which in addition to the highly imperfect private monitoring of the standard RARMG setting we add imperfect private monitoring within a match.

[4]Note that, since in our setting there is no public signal, there is no room at all for the use of public strategies and, hence, all the equilibria we obtain are in private strategies.

This version: November 5, 2013

(1990) and introduce a kind of "strategic" Tchebychev inequality, whose main role is to ensure that punishments associated with the strategies we define, are effective not only in expectation, but also almost surely.

Further, we show that defining strongly uniform equilibrium for discounted payoffs would lead to a weaker equilibrium concept.[5] Thus, the strength of our approach in establishing the folk theorems stems from the extra freedom given by epsilon-equilibria, and not from the choice of undiscounted payoffs. It is worth noting that, if applied to a setting where discounted payoffs are considered, the strategies we construct have the important feature that they are robust to changes in the discount factor and, indeed, apply even if players have different discount factors that are private information. Typically, these properties do not hold in equilibria obtained in the analysis of repeated games with imperfect monitoring, where it is common that the strategies have to be tailored to the (common) discount factor. Such robustness properties are particularly important in our setting, where economic agents may not be very sophisticated.

It is well known that there is no hope of supporting the whole set of feasible and individually rational payoffs under exact Nash equilibria in discounted games, not even under perfect monitoring (see Forges et al. (1986) for a simple counter-example). Thus, a modification of the equilibrium concept is needed to get positive results. In this sense, we establish that for RARMG, once we relax the equilibrium notion to epsilon-Nash, we can impose additional requirements such as sequential rationality, uniformity, stability, and robustness and still get folk theorems using simple strategies.

The use of epsilon-equilibria in the analysis of repeated interactions can be traced back to Radner (1980, 1981). We consider the setting of this paper to be specially appropriate for approximate solution concepts. In the anonymous random matching setting, equilibrium behavior is typically interpreted as a social norm. Research in anthropology and theoretical biology (*e.g.*, Boyd and Richerson (2002)) argues that the evolution of cooperation and punishment as social norms are plausibly a side effect of a tendency of people to conform (adopt common behaviors). Group beneficial social norms are sustained by the threat of punishment to violators. In turn, punishing violators is sustained as normative behavior, because the relative disadvantage suffered by those who enforce social norms compared with those who do not is small and easily balanced by even a weak tendency to conform. Boyd and Richerson (2001), discuss the adaptive nature of social norms and hypothesize that people deviate from the norm and adopt alternative behavior only if the payoff from deviating is *significantly* higher. While we do not explicitly model the agent's tendency to conform, it provides a foundation for why approximate optimality is sensible for members in a community, that is, players adhere to prescribed actions as long as they don't foresee large gains from deviating.

We now elaborate a bit on the equilibrium strategies we construct. Consider a RARMG with two communities. The strategies build on the idea of community enforcement, and we show cooperation

---

[5]A related analysis is also carried out in Fudenberg and Levine (1991).

This version: November 5, 2013

can be sustained via reversion to minmax punishments. The game is played in blocks of increasing length. In each block, players play the action profiles that achieve the target payoff. If a player observes a deviation, he switches to the minmax action till the end of the current block. At the start of the new block, players "restart" by again playing the action profiles that sustain the target payoff. Suppose a player plans to deviate at some period that is not near the end of a block. Regardless of his continuation strategy, if the block length is large enough, from some point on, with probability very close to 1, all players in the community would be playing the minmax action. After this, he will be minmaxed for the remainder of the block. If the block is long enough, this future loss in payoff will be larger than any gain he can make by deviating. Near the end of a block, a player might have a strict incentive to deviate, but playing the prescribed equilibrium action is still approximately optimal. Finally, concerning the incentives of a player who is supposed to punish a deviation, if he is patient enough, it is approximately optimal for him to spend the current block disciplining those who did not conform with the norm. As we mentioned above, the main challenge of extending this construction for games with more than two communities is to ensure not only that deviations are detected, but also that deviators are identified.

The use of block strategies to study uniform equilibria is not new. For instance, see the folk theorem results in Lehrer (1990, 1992c), Fudenberg and Levine (1991), and Tomala (1999). When trying to derive folk theorems, there are two main challenges: detecting deviations and identifying the deviator.[6] In the above papers these tasks are facilitated by imposing some structure on the monitoring technology to facilitate detection and, for identification, either two player games are considered or some restrictions are imposed on the set of achievable payoffs. In this paper we show that, within the context of RARMG, folk theorems are possible under virtually no assumptions on the stage game. In this literature it has also been common to assume that payoffs are observable, an assumption that we do not need when we allow for imperfect private monitoring within a match.

The equilibrium strategies also have a desirable stability property. If we take the equilibrium to be a positive description of play in large communities, it is reasonable to require that one mistake by a single player does not destroy the social norm forever. Put differently, we would like social norms to be "globally stable" in the sense that, after any finite history, play finally reverts to cooperative play (Kandori, 1992). The equilibrium we construct has this feature, since equilibrium play involves "restarting" the game at the start of each new block, regardless of the history.

The rest of the paper is organized as follows. In Section 2, we present the model. In Section 3, we present our solution concept and elaborate on its key properties. Section 4 contains the folk theorems. Section 5 contains a discussion. The Appendix contains some technical proofs.

---

[6]The survey paper by Gossner and Tomala (2009) has a deep discussion on this and other challenges that have to be solved in order to get folk theorems. Also the book by Mailath and Samuelson (2006) deeply discusses these issues in the context of discounted payoffs.

5

This version: November 5, 2013

2 Model

2.1 Players, matching technology, actions, and strategies

There is a finite set of communities, denoted by $C = \{1, \ldots, |C|\}$, each community with players indexed by the set $M = \{1, \ldots, |M|\}$. The set of players in society is denoted by $N = C \times M$. Given a player $i \in N$, $i^c \in C$ denotes the community of player i. In each period $t \in \{1, 2, \ldots\}$, the players are randomly matched to play a $|C|$-player stage game, where the role of each player is given by the community he belongs to. A matching is a partition of the $|C| \times |M|$ players into $|M|$ groups of $|C|$ players such that the players in each group play together the $|C|$-player stage game and no two players in the same community are matched together. The matching is anonymous, independent, and uniform over time.

In the stage game each community $c \in C$ has a finite action set, A_c, with $A = \prod_{c \in C} A_c$ denoting the set of action profiles in the stage game between communities. We assume that each community has at least two actions. Let $\hat{A} = A^M$ denote the set of action profiles that can be played at each stage of the random matching game. Given $i \in N$, let $A_i = A_{i^c}$ and $A_{-i} = \prod_{j \in C \setminus \{i^c\}} A_j$. We denote mixed actions by $s_i \in \Delta(A_i)$ and correlated action profiles of players different from i by $s_{-i} \in \Delta(A_{-i})$. Given $s_i \in \Delta(A_i)$, $s_{-i} \in \Delta(A_{-i})$, $a_i \in A_i$ and $a_{-i} \in A_{-i}$, we let $s_i(a_i)$ and $s_{-i}(a_{-i})$ denote the probabilities with which a_i and a_{-i} are chosen.

In the repeated game the players only observe the transactions they are personally engaged in. In particular, a player gets no information about his opponents' identities, or about how other players have been matched, or about the actions chosen by any other group of players. Thus, for each $t \geq 0$, the set of personal period t histories is given by $H^t = A^t$, where $A^0 = \{\emptyset\}$. Let $H = \bigcup_{t=0}^{\infty} H^t$ denote the set of all personal histories. Further, the set of complete period t histories is given by $\hat{H}^t$ and is defined so that each $\hat{h} \in \hat{H}^t$ contains all the information about everything that has happened up to period t, namely, the t realized matchings and the chosen action profiles (an element of $\hat{A}^t$). Let $\hat{H} = \bigcup_{t=0}^{\infty} \hat{H}^t$ denote the set of all complete histories. A (behavior) strategy for player $i \in N$ is defined as a mapping $\sigma_i : H \to \Delta(A_i)$. The set of strategies of player $i \in N$ is denoted by Σ_i. A strategy profile σ is *symmetric* if, for each pair of players i and j such that $i^c = j^c$, we have $\sigma_i = \sigma_j$.

2.2 Payoffs

For each $c \in C$, $u_c : A \to \mathbb{R}$ denotes the stage-game payoffs of community c. We also use u_c to denote the natural extension to mixed actions. Given a player $i \in N$, $u_i = u_{i^c}$. Further, for each $i \in N$, we also define his *ex ante* expected payoffs (across all possible matchings) as $g_i : A^M \to \mathbb{R}$.

Now, we define the expected payoffs in the repeated game. Given the matching technology, a strategy profile σ induces a probability measure $\mathbb{P}_\sigma$ over the set of action profiles to be played. For each $i \in N$, γ_i^T denotes the (expected) undiscounted average utilities up to period T and γ_i^δ denotes

the (expected) discounted average utilities. Formally,

$$\gamma_i^T(\sigma) = \mathbb{E}_{\mathbb{P}_\sigma}\left(\frac{1}{T}\sum_{t=1}^{T} g_i(a^t)\right) \qquad \text{and} \qquad \gamma_i^\delta(\sigma) = \mathbb{E}_{\mathbb{P}_\sigma}\left((1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}g_i(a^t)\right).$$

If a strategy profile is symmetric, then we can talk about the expected payoff of a community. Given a player $i \in N$, $\gamma_{i^c}^T = \gamma_i^T$, $\gamma_{i^c}^\delta = \gamma_i^\delta$.

We extend the above expressions to define expected payoffs conditional on a given history. For each period $\bar{t} \geq 0$, each complete history $\hat{h} \in \hat{H}^{\bar{t}}$, and each $t > \bar{t}$, let $\mathbb{P}_\sigma(a^t \mid \hat{h})$ denote the probability of profile a^t being played in period t when $\hat{h}$ has been realized and play thereafter proceeds according to σ (well defined even if $\mathbb{P}_\sigma(\hat{h}) = 0$). Then, for each $i \in N$ we have

$$\gamma_i^T(\sigma \mid \hat{h}) = \mathbb{E}_{\mathbb{P}_\sigma(\cdot\mid\hat{h})}\left(\frac{1}{T}\sum_{t=\bar{t}+1}^{\bar{t}+T} g_i(a^t)\right) \quad \text{and} \quad \gamma_i^\delta(\sigma \mid \hat{h}) = \mathbb{E}_{\mathbb{P}_\sigma(\cdot\mid\hat{h})}\left((1-\delta)\sum_{t=\bar{t}+1}^{\infty}\delta^{t-\bar{t}-1}g_i(a^t)\right).$$

2.3 Feasibility and individual rationality

Let $F = \text{co}(u(A)) \subset \mathbb{R}^C$ be the feasible set of the stage game. We define now the different minmax payoff notions we use, and the corresponding individually rational payoff sets.

Independent minmax. For each $i \in N$, $v_i = \min\limits_{s_{-i} \in \prod_{j \in C \setminus \{i^c\}} \Delta(A_j)} \max\limits_{a_i \in A_i} u_i(a_i, s_{-i})$.

The corresponding set of individually rational payoffs is $IR = \{x \in \mathbb{R}^C : x \geq v\}$.

Correlated minmax. For each $i \in N$, $w_i = \min\limits_{s_{-i} \in \Delta(A_{-i})} \max\limits_{a_i \in A_i} \sum\limits_{a_{-i} \in A_{-i}} s_{-i}(a_{-i})u_i(a_i, a_{-i})$.

Let $IRC = \{x \in \mathbb{R}^C : x \geq w\}$.

Repeated game minmax.[7] We say that player $i \in N$ can be *forced* to utility $x_i \in \mathbb{R}$ if, for each $\varepsilon > 0$, there are $\sigma_{-i} \in \Sigma_{-i}$ and $T_0 \in \mathbb{N}$ such that,[8] for each $\tau_i \in \Sigma_i$ and each $T \geq T_0$, $\gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \varepsilon$.

Now we are ready to define, for each player $i \in N$, i's repeated game minmax payoff as $v_i^\infty = \inf\{x_i : i \text{ can be forced to } x_i\}$. Let $IR^\infty = \{x \in \mathbb{R}^C : x \geq v^\infty\}$.

Note that w is attained with correlated punishments and v with independent punishments. On the other hand, v_i^∞ lies somewhere in between since the punishing players can use past play to get some degree of coordination. Thus, $w_i \leq v_i^\infty \leq v_i$. Further, in the case of two communities, $w_i = v_i^\infty = v_i$ and and $IR = IRC = IR^\infty$.

---

[7]See Renault and Tomala (1998).

[8]We denote by $\mathbb{N}$ the set $\{1, 2, 3 \ldots\}$.

This version: November 5, 2013

3 Strongly Uniform Equilibrium

In the anonymous random matching setting, equilibrium behavior is typically interpreted as a social norm. In this context, the assumption that agents choose approximate best responses is reasonable, and is consistent with evidence that people do not deviate from norms unless the implied gains are high enough.[9] This observation motivates the use of *uniform equilibrium* (UE), in which players' strategies are (uniformly) approximate best responses to their rivals' strategies.

Definition 1. A strategy profile σ is a *uniform equilibrium* with payoff $x \in \mathbb{R}^N$ if $\lim_{T \to \infty} \gamma^T(\sigma) = x$ and, for each $\varepsilon > 0$, there is T_0 such that, for each $T \geq T_0$, σ is an ε-Nash equilibrium in the finitely repeated game with T periods, *i.e.*, for each $i \in N$ and each $\tau_i \in \Sigma_i$,

$$\gamma_i^T(\tau_i, \sigma_{-i}) \leq \gamma_i^T(\sigma) + \varepsilon.$$

In this paper we go one step further and introduce a new equilibrium concept, *strongly uniform equilibrium* (SUE), which is a refinement of UE. In addition to the UE requirements, players' strategies satisfy a strong version of sequential rationality after every history. This implies that equilibrium behavior is not only (approximately) optimal *ex ante*, but indeed after any realized history of play. An SUE also has the desirable feature of global stability: if we think of an equilibrium as a positive description of behavior in large communities, we should require some stability, in the sense that a small mistake by a single agent does not destroy the norm forever.

Definition 2. A strategy profile σ is a *strongly uniform equilibrium* with payoff $x \in \mathbb{R}^N$ if we have that $\lim_{T \to \infty} \gamma^T(\sigma) = x$ and, for each $\bar{t} \geq 0$, each $\hat{h} \in \hat{H}^{\bar{t}}$, and each $i \in N$, the following conditions hold:

1) $\lim_{T \to \infty} \dfrac{1}{T} \displaystyle\sum_{t=\bar{t}+1}^{\bar{t}+T} g_i(a^t) = x_i \quad \mathbb{P}_\sigma(\cdot \mid \hat{h})$-a.s.

2) For each $\tau_i \in \Sigma_i$, $\limsup_{T \to \infty} \dfrac{1}{T} \displaystyle\sum_{t=\bar{t}+1}^{\bar{t}+T} g_i(a^t) \leq x_i \quad \mathbb{P}_{(\tau_i, \sigma_{-i})}(\cdot \mid \hat{h})$-a.s.

3) For each $\varepsilon > 0$, there is $T_0 \in \mathbb{N}$ such that, for each $T \geq T_0$ and each $\tau_i \in \Sigma_i$,

$$\gamma_i^T(\sigma_{-i}, \tau_i \mid \hat{h}) \leq \gamma_i^T(\sigma \mid \hat{h}) + \varepsilon.$$

For most of the paper we work with symmetric strategy profiles, so we denote by $E^* \subset \mathbb{R}^C$ and $E \subset \mathbb{R}^C$ the sets of payoffs attainable in symmetric UE and SUE, respectively.

---

[9]See for instance Boyd and Richerson (2002).

8

Clearly, an SUE is also a UE. Condition 1) implies that as T goes to infinity, x is not only the expected payoff of σ, but also, almost surely, the realized payoff. Further, given any history, if we look at the realized payoff from this history onwards, this payoff is almost surely x (as T goes to infinity). This is a strong form of global stability (Kandori, 1992), which is usually stated in expected terms. Condition 2) implies that, after every history, all deviations are almost surely non-profitable (not only in expectation). Condition 3) implies that, for each history $\hat{h} \in \hat{H}^t$, each $\varepsilon > 0$, and each deviation after $\hat{h}$, if T is large enough, the expected profit will not be larger than ε. This constitutes a strong form of approximate sequential rationality since, regardless of the beliefs a player might have at an information set, no deviation would give him a profit larger than ε.[10] For this last observation it is important that SUE is defined with respect to histories in $\hat{H}$, not with respect to personal histories. For instance, when looking at Condition 3), we have that a player would not be able to make more than ε profit by deviating even if he could condition his deviation on the complete history of past play, not only on what he has observed.[11] Using histories in $\hat{H}$ leads to a more demanding equilibrium concept and, thus, the obtained folk theorems are stronger results.

At this point, it may be worth relating SUE to the notion of uniform sequential equilibrium introduced in Fudenberg and Levine (1991). Essentially, their equilibrium notion is captured by Condition 3) above. Further, their definition is of the form "for each T_0 there exists $T \geq T_0$ such that", *i.e.*, under the equilibria they construct it may be that, for each T_0 there is $T \geq T_0$ for which the condition does not hold.[12] They acknowledge this limitation, but say that they were not able to strengthen the results in this direction. Fudenberg and Levine (1991) were also concerned with the frequency of punishments on the equilibrium path, an issue for which Condition 2) in our definition ensures that, in an SUE, punishments almost surely cease.

Below, we adapt the definition of SUE to the context of discounted payoffs and show that we get an equilibrium concept weaker than SUE. Thus, our results are limited only by the fact that we work with approximate equilibria, but not by the choice of undiscounted payoffs.[13]

Definition 3. A strategy profile σ is a *discounted strongly uniform equilibrium* (DSUE) with payoff $x \in \mathbb{R}^n$ if $\lim_{\delta \to 1} \gamma^\delta(\sigma) = x$ and, for each $\bar{t} \in \mathbb{N}$, each $\hat{h} \in \hat{H}^{\bar{t}}$, and each $i \in N$ the following conditions hold:

1) $\lim_{\delta \to 1} (1 - \delta) \sum_{t=\bar{t}+1}^{\infty} \delta^{t-\bar{t}-1} g_i(a^t) = x_i \quad \mathbb{P}_\sigma(\cdot \mid \hat{h})$-a.s.

---

[10]It is worth noting that condition 2) does not imply condition 3). This is because, in the latter, T_0 is the same for all $\tau_i \in \Sigma_i$. This uniformity is not implied by the almost sure convergence in condition 2).

[11]In this sense, SUE has some flavor to belief-free equilibria (Ely and Välimäki, 2002; Piccione, 2002; Ely et al., 2005).

[12]Mathematically, this is as working with the $\liminf$ instead of the $\limsup$, which is significantly weaker. Because of this, uniform sequential equilibrium does not imply uniform equilibrium.

[13]In Fudenberg and Levine (1991) the authors devote a whole section to make a similar point regarding the use of discounted or undiscounted payoffs.

This version: November 5, 2013

2) For each $\tau_i \in \Sigma_i$, $\limsup\limits_{\delta \to 1} (1 - \delta) \sum\limits_{t=\bar{t}+1}^{\infty} \delta^{t-\bar{t}-1} g_i(a^t) \leq x_i \quad \mathbb{P}_{(\tau_i, \sigma_{-i})}(\cdot \mid \hat{h})$-a.s.

3) For each $\varepsilon > 0$, there is $\delta_0 \in (0, 1)$ such that, for each $\delta \in (\delta_0, 1)$ and each $\tau_i \in \Sigma_i$,

$$\gamma_i^\delta(\sigma_{-i}, \tau_i \mid \hat{h}) \leq \gamma_i^\delta(\sigma \mid \hat{h}) + \varepsilon.$$

Proposition 1. *If a strategy profile σ is an SUE then it is a DSUE.*

Proof. Given an arbitrary sequence of real numbers $(x^t)_{t \in \mathbb{N}}$, let $\bar{x}^T = \frac{1}{T} \sum_{t=1}^{T} x^t$ and $\bar{x}^\delta = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} x^t$. Then, it is not hard to show that

$$\limsup_{T \to \infty} \bar{x}^T \geq \limsup_{\delta \to 1} \bar{x}^\delta \geq \liminf_{\delta \to 1} \bar{x}^\delta \geq \liminf_{T \to \infty} \bar{x}^T,$$

and so the convergence of $\bar{x}^T$ implies the convergence of $\bar{x}^\delta$ to the same limit. The formal proof, which can be seen in Renault and Tomala (2011, Lemma 2.15), relies on the fact that $\bar{x}^\delta = \sum_{T=1}^{\infty} T(1 - \delta)^2 \delta^{T-1} \bar{x}^T$.

Suppose that a strategy profile σ satisfies 1) in Definition 2. Then, $\frac{1}{T} \sum_{t=\bar{t}}^{\bar{t}+T} g_i(a^t)$ converges to x_i almost surely, which implies that the corresponding discounted average, $(1 - \delta) \sum_{t=\bar{t}}^{\infty} \delta^{t-\bar{t}} g_i(a^t)$, also converges to x_i almost surely, which establishes that 1) in Definition 2 implies 1) in Definition 3. Since $\limsup_{T \to \infty} \bar{x}^T \geq \limsup_{\delta \to 1} \bar{x}^\delta$, the argument for 2) is also straightforward. Concerning 3), the proof follows from a straightforward adaptation of the second part of the proof of Lemma 2.15 in Renault and Tomala (2011). $\square$

It is not hard to construct counterexamples to show that the converse of Proposition 1 is not true.

3.1 General coincidence of UE and SUE payoff sets

While an SUE is a refinement of a UE in the sense that the strategies that constitute a UE do not necessarily constitute an SUE, it turns out that the two solution concepts are equivalent in terms of payoff sets. Formally, we prove that, if a particular individually rational and feasible payoff of the stage game can be sustained in a UE, then there exists an SUE that achieves the same payoff. In other words, when studying payoff sets, all the refinements contained in the definition of SUE can be obtained at no cost. It is important to note that this equivalence result about payoff sets is not specific to the random matching environment, and actually applies to any repeated game with signals (private or public, deterministic or stochastic), with no other assumption than a finite set of players N, finite sets of actions (in the stage game), and finite sets of signals for each player.

Theorem 1. *For each infinitely repeated game with finite players, actions, and signals, the sets of UE and SUE payoffs coincide.*

Because of its generality, this result is of independent interest, in order to facilitate the flow of the paper, we relegate the proof to the Appendix. Crucially for this paper, the SUE constructed in the proof has the following feature.

Corollary 1. *In the anonymous random matching setting, if a symmetric strategy profile σ is a uniform equilibrium, then there is a symmetric SUE that attains the same payoff.*

3.2 A strategic Tchebychev inequality

Although we have relegate the proof of Theorem 1 to the Appendix, we present below a lemma that is crucial for its proof (and other proofs in this paper). This result is a sort of generalization of Tchebychev inequality that allows to show that the different punishments we will construct are effective not only in expectation, but also almost surely.[14] The result and its proof are slightly adapted from Lemma 5.6 in Lehrer (1990). For the sake of completeness, we present below a short proof that uses the standard Tchebychev inequality.[15]

Consider any finite two-player repeated game with a finite set of signals. Let A_1 and A_2 be the sets of actions an g_1 be player 1's payoff function in the stage game. No further assumption is made on the signals. Let s_2 in $\Delta(A_2)$ be a mixed action of player 2 in the stage game. Assume that player 2 plays the strategy σ_2 that consists of playing s_2 at each period independently of everything else. Let R be an upper bound on $|g_1|$.

The idea underlying the lemma below is that, since player 2's randomizations are independent across time, regardless of the realized history, there is no way in which player 1 can strategically use the information about past realizations of player 2's actions to affect the distribution of his current and future payoffs. Note that the result below does not depend on the quality of information available to player 1 and, hence, it is independent of the signal structure.

Lemma 1 (A strategic Tchebychev inequality). *Let σ_2 be the strategy of player 2 that consists of playing $s_2 \in \Delta(A_2)$ in every period. For each strategy σ_1 of player 1, each length $T \in \mathbb{N}$, and each $\varepsilon > 0$,*

$$\mathbb{P}_{(\sigma_1,\sigma_2)}\left(\left|\frac{1}{T}\sum_{t=1}^{T} g_1(a_1^t, a_2^t) - \frac{1}{T}\sum_{t=1}^{T} g_1(a_1^t, s_2)\right| \geq \varepsilon\right) \leq \frac{R^2}{\varepsilon^2 T}$$

and, as a consequence,

$$\mathbb{P}_{(\sigma_1,\sigma_2)}\left(\frac{1}{T}\sum_{t=1}^{T} g_1(a_1^t, a_2^t) \geq \max_{a_1 \in A_1} g_1(a_1, s_2) + \varepsilon\right) \leq \frac{R^2}{\varepsilon^2 T}.$$

---

[14]This result is needed to rule out, for instance, situations in which a player, using past information, can ensure that in the repeated game punishment is effective with low probability (although very harsh when effective so that, in expectation, he does not gain more than ε).

[15]The same arguments could be used with other types of inequalities, *e.g.*, the Hoeffding inequality would give an upper bound on $\exp(-\varepsilon T)$.

This version: November 5, 2013

Proof. Expectations and probabilities in this proof are with respect to $\mathbb{P}_{(\sigma_1,\sigma_2)}$. For each $t \in \{1,\ldots,T\}$, let $\mathcal{F}_t$ be the σ-field generated by history $h^{t-1} = (a_1^{\bar{t}}, a_2^{\bar{t}})_{\bar{t}<t}$, and define the random variable $Z_t = g_1(a_1^t, a_2^t) - g_1(a_1^t, s_2)$. So defined, Z_t is measurable with respect to $\mathcal{F}_{t+1}$. Further, since a_2^t is independent of h^{t-1} we have that, for each $\bar{t} \le t$, $\mathbb{E}(Z_t \mid \mathcal{F}_{\bar{t}}) = 0$. Then, for each $\bar{t} < t$,

$$\mathbb{E}(Z_t Z_{\bar{t}}) = \mathbb{E}(\mathbb{E}(Z_t Z_{\bar{t}} \mid \mathcal{F}_{\bar{t}+1})) = \mathbb{E}(Z_{\bar{t}} \mathbb{E}(Z_t \mid \mathcal{F}_{\bar{t}+1})) = 0.$$

Let $Z = \sum_{t=1}^T Z_t$. Then, $\mathbb{E}(Z) = 0$ and, since the Z_t are uncorrelated, $V(Z) = \sum_{t=1}^T V(Z_t) \le T R^2$.[16] By Tchebychev inequality,

$$\mathbb{P}(|Z| \ge T\varepsilon) \le \frac{V(Z)}{\varepsilon^2 T^2} \le \frac{R^2}{\varepsilon^2 T}.$$

Therefore, $\mathbb{P}(|\frac{1}{T}\sum_{t=1}^T g_1(a_1^t, a_2^t) - \frac{1}{T}\sum_{t=1}^T g_1(a_1^t, s_2)| \ge \varepsilon)$ is, at most, $\frac{R^2}{\varepsilon^2 T}$. The last observation in the statement follows from the inequality $\frac{1}{T}\sum_{t=1}^T g_1(a_1^t, a_2^t) \le \max_{a_1 \in A_1} g_1(a_1, s_2)$. $\qquad\square$

It is important to note that Lemma 1 will be used several times in the following way: We consider that a given player i is being minimaxed for T periods. Then, when invoking Lemma 1, this player i will play the role of player 1 and all other players together will play the role of player 2.

3.3 Moral cost and sequential equilibrium

As discussed earlier, equilibria in the anonymous random matching setting are often viewed as social norms. To see an alternative interpretation of strategies that constitute an SUE, consider a setting with a discounted anonymous random matching game, in which agents suffer a (small) moral cost each time they deviate from a prescribed social norm. It is easy to show that a strategy profile that constitutes an SUE would constitute a standard (exact) sequential equilibrium in an environment with (arbitrarily) small moral costs and discounting.

Formally, suppose there is $\lambda > 0$, no matter how small, such that each time a player deviates from the prescribed social norm he faces a moral cost λ.

Lemma 2. *If a strategy profile σ is an SUE then, for each $\lambda > 0$, there is δ_0 such that, for each $\delta \in (\delta_0, 1)$, σ is a sequential equilibrium of the infinitely repeated game with discount δ and moral cost λ.*

---

[16]To see that $V(Z_t) \le R^2$ just note the following. Let $X = Z_t$. Then, $V(X) = \mathbb{E}(X^2)$ and so

$$\mathbb{E}(X^2) = \sum_{a_i \in A_i} \mathbb{P}(a_i^t = a_i)\mathbb{E}(X^2 \mid a_i^t = a_i) = \sum_{a_i \in A_i} \mathbb{P}(a_i^t = a_i)V(g_1(a_i, a_j^t)) \le \sum_{a_i \in A_i} \mathbb{P}(a_i^t = a_i)R^2 = R^2.$$

This version: November 5, 2013

4 Folk Theorems in the Anonymous Random Matching Setting

In this section, we present the three key results of this paper. First, we establish a folk theorem for two communities by showing that $E = F \cap IR$. We prove the result constructively, by describing explicitly the equilibrium strategies and then showing that they constitute an SUE. A nice feature of our construction is that the strategies are conceptually very simple. Similar to the seminal papers by Kandori (1992) and Ellison (1994), any target payoff is sustained in equilibrium using community enforcement. In other words, if an agent observes a deviation, he responds by punishing his future rivals, with the punishment action spreading the information that there has been a deviation.

Second, we show that the folk theorem extends to environments with more than two communities. The equilibrium strategies still retain the flavor of community enforcement. However, sustaining cooperation is more subtle with more than two communities, because in case of a deviation, players now need to not only spread the information that there has been a deviation, but also need to communicate the identity of the community from which a player deviated. We construct equilibrium strategies that involve such credible punishments and communication.

Finally, we go back to the case of two communities, and consider an environment in which there is imperfect monitoring. Even within a match, players do not perfectly observe the action of their rival, but rather receive a noisy (possibly private) signal about their rival's action. We present a weak sufficient condition on the information structure under which the folk theorem obtains.

4.1 Folk theorem with two communities

Recall that Corollary 1 ensures that the sets of symmetric UE and SUE payoffs coincide. Thus, in order to prove that $E = F \cap IR$ we can just prove the result for the set E^* of uniform equilibrium payoffs and then rely on Corollary 1 to get the result for SUE. Yet, in this section we have chosen to construct the strongly uniform equilibrium profile explicitly, so that the reader can appreciate that the equilibrium strategies are indeed quite natural. The idea of the strategy we construct is as follows: The game is played in blocks of increasing length. In each block, on the equilibrium path, players play the action profiles that achieve the target payoff. If a player observes a deviation, he switches to the minmax action until the end of the current block. At the start of the new block, players "restart" by again playing the action profiles that sustain the target equilibrium payoff.

Clearly, such a strategy profile achieves the target payoff. Further, since after a deviation the game restarts from the next block, we have a strong form of global stability. Concerning the incentives, if a player plans to deviate early in a block, from some point on, with probability very close to one, he will be minmaxed for a large number of periods. On the other hand, deviations late in a block will not significantly affect the payoffs. The proof shows how to handle intermediate deviations. Importantly, when a player is asked to revert to the minmax action, it is approximately

optimal for him to spend the current block disciplining those who did not conform with the norm.

Theorem 2 (Folk Theorem for Two Communities). *Suppose $|C| = 2$. Then, $E = F \cap IR$.*

Proof. Let $x \in F \cap IR$. Let $(\bar{a}^t)_t$ be a sequence of pure action profiles whose average payoff converges to x, *i.e.*, $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \bar{a}^t = x$.

Let $0 < \alpha \leq 1$ be the maximum probability a pure action receives in either of the two minmax mixed actions. We distinguish two cases in the definition of the strategy σ.

Case 1: $\alpha < 1$. The strategy σ is played in consecutive blocks of increasing length, B^l, with $l \in \mathbb{N}$ and starting with $l = 3$. For each l, block B^l consists of l^4-periods. The strategy σ prescribes playing grim trigger in each block. When a new block starts, regardless of the past, the players play according to $(\bar{a}^t)_t$ (starting at $t = 1$) and, during this block, after every history in which a player has observed at least one deviation, he minmaxes his opponent. Therefore, within each block, a player can be either on the equilibrium path or in the punishment phase. For the sake of brevity, we refer to these two states of a player as "uninfected" and "infected," respectively.

Case 2: $\alpha = 1$. The equilibrium strategy is the same one, except for the following modification: each player of a community whose minmax action is a pure one, upon observing a deviation, starts the punishment by randomizing uniformly over all his actions during l^2 periods and then switches to the minmax action.

We show that σ is an SUE. Clearly, $\lim_{T \to \infty} \gamma^T(\sigma) = x$. Now, take $\bar{t} \geq 0$, $\hat{h} \in \hat{H}^{\bar{t}}$, and $i \in N$. We divide each block B^l in three subblocks: $B^{l,1}$ and $B^{l,3}$ containing the first and last l^3 periods, respectively, and $B^{l,2}$ containing the $l^4 - 2l^3$ intermediate periods.

Condition 1) $\lim_{T \to \infty} \frac{1}{T} \sum_{t=\bar{t}+1}^{\bar{t}+T} g_i(a^t) = x_1$ $\mathbb{P}_\sigma(\cdot \mid \hat{h})$-a.s. This follows from the fact that, regardless of the history $\hat{h}$, the play is restarted at the end of each block, going back to $(\bar{a}^t)_t$.

Condition 2) Suppose, without loss of generality, that a player i in community 1 deviates. We want to show that, for each $\tau_i \in \Sigma_i$, $\limsup_{T \to \infty} \frac{1}{T} \sum_{t=\bar{t}+1}^{\bar{t}+T} g_i(a^t) \leq x_1$ $\mathbb{P}_{(\tau_i, \sigma_{-i})}(\cdot \mid \hat{h})$-a.s. All probabilities and expectations in this proof are with respect to $\mathbb{P}_{(\tau_i, \sigma_{-i})}$. Since play is "restarted" at the end of each block, it suffices to establish the above inequality with respect to the empty history, *i.e.*, $\bar{t} = 0$. First, note that, as soon as a player in community 2 observes a deviation, he gets infected and starts punishing. Then, contagion starts and more people get infected and start punishing.

Case 1: $\alpha < 1$. Given an infected player, unless all players in the other community are infected as well, the probability that he infects a new player in the current period is, at least, $\frac{1-\alpha}{M}$. To see why, note that, regardless of what is supposed to be played on the equilibrium path, the infected player will mismatch it with probability at least $1 - \alpha$ and $\frac{1}{M}$ is a lower bound on the probability of meeting an uninfected player. After the first deviation occurs, the deviating player infects his current opponent. Then, the probability that one of the remaining $2M - 2$ players remains uninfected after l^2 periods of contagion can be bounded above using a binomial distribution $\text{Bi}(l^2, \frac{1-\alpha}{M})$. Let Y_l denote

14

the number of infected players after l^2 periods and let $p = \frac{1-\alpha}{M}$. Thus,

$$P(Y_l < 2M - 2) \leq P(\text{Bi}(l^2, p) \leq 2M) \overset{\text{large } l}{\leq} P(|\text{Bi}(l^2, p) - pl^2|) \geq l^{\frac{8}{5}}),$$

and, by Tchebychev inequality, $P(|\text{Bi}(l^2, p) - pl^2| \geq l^{\frac{8}{5}}) \leq \dfrac{p(1-p)l^2}{l^{\frac{16}{5}}} \leq \dfrac{1}{l^{\frac{6}{5}}}.$

Case 2: $\alpha = 1$. Suppose that the minmax action requires the players in at least one of the communities to use a pure action and that $(\bar{a}^t)_t$ involves frequent use of it. In such a case, using only the minmax action after getting infected might make the contagion spread very slowly. This is the reason for the randomization during the first l^2 stages after getting infected. Then, we can parallel the above arguments with $\alpha = 1/2$ (if the punishing player has at least two actions, the probability of mismatching the equilibrium path one during these l^2 periods is at least $\frac{1}{2}$).

So far we have shown that, in both cases, the following property holds: if a player deviates in period $\hat{t}$ of a given block and period $\hat{t} + l^2$ still belongs to the same block, then the probability that all players different from player i are in the punishment phase after period $\hat{t} + l^2$ is, at least, $1 - 1/l^{\frac{6}{5}}$.

By definition of $(\bar{a}^t)_t$, given $\varepsilon \geq 0$ there is T_0 such that, for each $T \geq T_0$, $|\sum_{t=1}^{T} \frac{g(\bar{a}^t)}{t} - x| \leq \varepsilon$. We now study the gains that the deviating player can make in each block. We distinguish several cases, depending on the subblock at which the first deviation occurs. Let $l \geq \frac{T_0}{\varepsilon}$; in particular, $\frac{1}{l} \leq \varepsilon$. Let $R = \max_{a \in A, j \in \{1,2\}} |u_j(a)|$.

Late deviation. The first deviation occurs in subblock $B^{l,3}$. Then,

$$\sum_{t=1}^{l^4} \frac{g_i(a^t)}{l^4} = \sum_{t=1}^{l^4 - l^3} \frac{g_i(\bar{a}^t)}{l^4} + \sum_{t=l^4-l^3+1}^{l^4} \frac{g_i(a^t)}{l^4} \leq \frac{(l^4 - l^3)(x_1 + \varepsilon)}{l^4} + \frac{l^3 R}{l^4} \leq x_1 + \varepsilon + R\varepsilon.$$

Therefore, late deviations never lead to a gain higher than $(R+1)\varepsilon$.

Intermediate deviation. The first deviation occurs at period $\tilde{t}$, in subblock $B^{l,2}$. Then,

$$\sum_{t=1}^{l^4} \frac{g_i(a^t)}{l^4} \leq \sum_{t=1}^{\tilde{t}-1} \frac{g_i(\bar{a}^t)}{l^4} + \sum_{t=\tilde{t}}^{\tilde{t}+l^2+1} \frac{R}{l^4} + \sum_{t=\tilde{t}+l^2+1}^{l^4} \frac{g_i(a^t)}{l^4} \leq \frac{\tilde{t}(x_1 + \varepsilon)}{l^4} + R\varepsilon + \sum_{t=\tilde{t}+l^2+1}^{l^4} \frac{g_i(a^t)}{l^4}.$$

After period $\tilde{t} + l^2 + 1$ there have been already l^2 periods of contagion so, with probability at least $1 - 1/l^{\frac{6}{5}}$ all the players have seen a deviation. For player 1 to make some gains after period $\tilde{t} + l^2 + 1$ one of two things has to happen: either the contagion was not successful, which happens with probability at most $1/l^{\frac{6}{5}}$ or the contagion was successful but not the ensuing punishment. By construction, from period $\tilde{t} + l^2 + 1$ on, there are at least $l^3 - l^2$ periods of punishment in $B^{l,3}$. Let $\hat{\sigma}$ be any strategy profile where player 1 is being minmaxed. Then, using the strategic Tchebychev inequality (Lemma 1), we have that the probability that player 1 makes a gain larger than ε while

15

being punished during $l^3 - l^2$ periods is bounded by

$$\mathbb{P}_{\hat{\sigma}}\Big(\sum_{t=1}^{l^3-l^2} \frac{g_i(a^t)}{l^3 - l^2} - v_1 \geq \varepsilon\Big) \leq \frac{R^2}{\varepsilon^2(l^3 - l^2)}.$$

Recall that, since $x \in IR$, $x_1 \geq v_1$. Then, with probability at least $(1 - \frac{1}{l^{6/5}})(1 - \frac{R^2}{\varepsilon^2(l^3-l^2)})$, we have

$$\sum_{t=1}^{l^4} \frac{g_i(a^t)}{l^4} \leq \frac{\tilde{t}(x_1 + \varepsilon)}{l^4} + R\varepsilon + \sum_{t=\tilde{t}+l^2+1}^{l^4} \frac{v_1 + \varepsilon}{l^4} \leq \frac{(l^4 - l^2)(x_1 + \varepsilon)}{l^4} + R\varepsilon \leq x_1 + \varepsilon + R\varepsilon.$$

Therefore, the probability that player 1 gets more than $(R+1)\varepsilon$ in block l is, at most, $\frac{1}{l^{6/5}} + \frac{R^2}{\varepsilon^2(l^3-l^2)}$.

Early deviation. The first deviation occurs at period $\tilde{t}$, in subblock $B^{l,1}$. In this case we have l^2 periods of contagion followed by more than l^3 periods of punishment and we can easily get that, with probability at least $(1 - \frac{1}{l^{6/5}})(1 - \frac{R^2}{\varepsilon^2 l^3})$,

$$\sum_{t=1}^{l^4} \frac{g_i(a^t)}{l^4} \leq \frac{(\tilde{t} + l^2 + 1)R}{l^4} + \frac{(l^4 - \tilde{t} - l^2 - 1)(x_1 + \varepsilon)}{l^4} \leq x_1 + (R+1)\varepsilon.$$

Therefore, in all three cases, player 1 can make a gain larger than $x_1 + (R+1)\varepsilon$, with probability at most $\frac{1}{l^{6/5}} + \frac{R^2}{\varepsilon^2(l^3-l^2)}$. Since $\sum_{l \geq 1}(\frac{1}{l^{6/5}} + \frac{R^2}{\varepsilon^2(l^3-l^2)}) < \infty$, by Borel-Cantelli lemma, with probability 1 there is l_0 such that, for each $\bar{l} \geq l_0$, the payoff to player 1 in block $\bar{l}$ is smaller than $x_1 + (R+1)\varepsilon$. Since this is true for every $\varepsilon > 0$, $\lim_{T\to\infty} \sum_{t=1}^{T} \frac{g_i(a^t)}{T} \leq x_1$ $\mathbb{P}_{(\tau_i, \sigma_{-i})}$-a.s.

Condition 3) For each $\varepsilon > 0$, there is $T_0 \in \mathbb{N}$ such that, for each $T \geq T_0$ and each $\tau_i \in \Sigma_i$, $\gamma_i^T(\sigma_{-i}, \tau_i \mid \hat{h}) \leq \gamma_i^T(\sigma \mid \hat{h}) + \varepsilon$. This easily follows from the analysis above. For each block, in case of a late deviation, the payoff is essentially given by the sequence $(\bar{a}^t)_t$. The payoff in case of an intermediate deviation is essentially given by $(\bar{a}^t)_t$ and the minmax phase. For an early deviation, the payoff is essentially given by the minmax phase. $\qquad\square$

4.2 Folk theorem with more than two communities

The main complication when working with an arbitrary number of communities is that, in order to make punishments as effective as possible, one needs to identify the community of a deviator. This is a common problem in repeated games with imperfect monitoring which is deeply discussed, for instance, in Gossner and Tomala (2009) where the authors present a simple three-player game in which a payoff profile at which deviations are detectable cannot be sustained in equilibrium because the deviator cannot be identified. Fudenberg and Levine (1991) abstract away from the identification problem by restricting attention to the set of "mutually punishable payoffs"; each such payoff x has

the property that there is some mixed action of the stage game that simultaneously forces all players to a payoff below x. The main effort here is to show that, in our setting, deviators can be identified.

Since the explicit construction of an SUE is more involved than in the previous section, we construct a the proof for UE payoffs and then rely on Corollary 1 to get the desired result. Given any payoff in $F \cap IR^\infty$, we construct a UE that achieves it. In addition to v_i^∞ we define, for each $T \in \mathbb{N}$, v_i^T as the minmax payoff of player i in the T-period repeated game.

$$v_i^T = \min_{\sigma_{-i} \in \Sigma_{-i}} \max_{\sigma_i \in \Sigma_i} \gamma_i^T(\sigma_i, \sigma_{-i}).$$

The larger T is, the more coordination the punishing players can achieve. Indeed, v_i^T converges to $\inf_{T \in \mathbb{N}} v_i^T = v_i^\infty$.

Theorem 3. *Suppose $|C| \geq 2$. Then, $E = F \cap IR^\infty$.*

The case $|C| = 2$ corresponds with Theorem 2. Thus, we assume that $|C| > 2$. Similarly to the case of two communities, we present strategies in which play proceeds in blocks. However, we now introduce, between any two blocks of play, what we call *communication blocks*. During these blocks, players use their actions to spread information about past play.In particular, the information that a deviation has occurred and the identity of the community of the deviating player. Below we describe the communication blocks and show that, if they are long enough, they enable detection of deviations and identification of deviators.

Communication blocks. Assume that players are supposed to play a given sequence of pure actions in some block B. At the end of B, we say that each player $i \in N$ is in a state $\mu(i, c) \in \{0, 1\}$ with community c in C, with the interpretation that $\mu(i, c) = 1$ if and only if player i has seen a deviation by someone of community c at some period in B. Fix two distinct actions a_c and $\hat{a}_c$ for each community c. We define a communication block B^L, to be played after block B, as follows. B^L has two consecutive subblocks $B^{L,1}$ and $B^{L,2}$, each subblock having $L|C|$ periods.

The subblocks $B^{L,1}$ and $B^{L,2}$ are identical. Each comprises L consecutive periods targeting community 1, followed by L consecutive periods targeting community 2, and so on. At each period in which the target is community $c \in C$, each player i in N plays $a_i = a_{i^c}$ if $\mu(i, c) = 0$ and $\hat{a}_i = \hat{a}_{i^c}$ if $\mu(i, c) = 1$. The interpretation of playing $\hat{a}_i$ is that player i accuses someone from community c of having deviated during block B. At the end of the L periods targeting community c, each player i updates his state as follows. If player i started block B^L with $\mu(i, c) = 0$ for all communities c, and observed players of at least 2 *different* communities accusing community c (at the same or at different periods), then player i sets $\mu(i, c) = 1$; it is possible that $\mu(i, i^c) = 1$. Otherwise, player i's state does not change.

It is worthwhile to clarify why we need two subblocks in B^L. Suppose that $|C| = 3$ and that player $i \in N$ has deviated in one period of block B. Now, if there is no other deviation in B or B^L,

only the players in i^c can learn about player i's deviation during $B^{L,1}$. Then, in $B^{L,2}$, since players in i^c will accuse their own community, all other communities will learn about the deviation as well.

Let σ be a strategy profile such that a given path of pure actions is to be played at some block B, followed by a communication block B^L. Deviations are temporarily ignored during block B. Let $\mathcal{Z}_i$ be the event: "player i has deviated from the path of pure actions in some period of B". Let $\bar{\mathcal{Z}}_i$ be the complement of $\mathcal{Z}_i$. Let $\mu_L(i,c)$ be the final state of i regarding community c at the end of B^L.

Lemma 3 (Communication lemma). *For each player $i \in N$ and each strategy $\tau_i \in \Sigma_i$, we have*

  1) "False accusations are ignored".

$$\mathbb{P}_{(\tau_i, \sigma_{-i})}(\forall i' \neq i, \forall c \neq i^c, \mu_L(i', c) = 0) = 1 \ \text{ and } \ \mathbb{P}_{(\tau_i, \sigma_{-i})}(\forall i' \neq i, \mu_L(i', i^c) = 0 \mid \bar{\mathcal{Z}}_i) = 1.$$

  2) "Deviations are detected". There is $f(L, |M|)$ such that $\lim_{L \to \infty} f(L, |M|) = 1$ and

$$\mathbb{P}_{(\tau_i, \sigma_{-i})}(\forall i' \neq i, \mu_L(i', i^c) = 1 \mid \mathcal{Z}_i) \geq f(L, |M|).$$

Proof of Lemma 3. Statement 1) is straightforward, since accusations by only one community are unsuccessful. On the other hand, if a player i deviates in block B, $|C| - 1 \geq 2$ players from different communities observe this deviation and inform in block B^L. Thus, given a community size $|M|$, the probability of the states $\mu_L(i', i^c) = 1$ goes to 1 as L goes to infinity. $\qquad\square$

In the proof of Theorem 3 we rely on the following well known characterization of uniform equilibrium payoffs.[17]

Lemma 4. *Let $x \in \mathbb{R}^N$. The following conditions are equivalent:*

  1) x is a uniform equilibrium payoff.

  2) For each $k \in \mathbb{N}$, there are $\sigma \in \Sigma$ and $T \in \mathbb{N}$ such that, for each $i \in N$ and each $\tau_i \in \Sigma_i$,

$$\gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \frac{1}{k} \quad \text{and} \quad \gamma_i^T(\sigma) \geq x_i - \frac{1}{k}.$$

Proof of Theorem 3 (when $|C| > 2$). The inclusion $E \subset F \cap IR^\infty$ is straightforward. We now prove that $F \cap IR^\infty \subset E$. Let x in $F \cap IR^\infty$ and let $\varepsilon > 0$. Our objective is to find $T \in \mathbb{N}$ and a strategy profile σ of the T-period repeated game such that, for each $i \in N$ and each $\tau_i \in \Sigma_i$, $\gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \varepsilon$ and $\gamma_i^T(\sigma) \geq x_i - \varepsilon$. By Lemma 4, this will imply that x is a UE payoff and then we can use Corollary 1 to get that $x \in E$.

---

[17] See Proposition 4.5 in Mertens et al. (1994, page 196) or Lemma 2.13 in Renault and Tomala (2011).

Definition of σ. The strategy σ is played in L consecutive blocks, where L is a large number to be fixed later. Each block l comprises a "payoff block" B^l of length L^2 followed by a "communication block" D^l of length $2L|C|$. The total number of periods is $T = L(L^2 + 2L|C|)$. For each player $i \in N$, let $\bar{\sigma}^i = \bar{\sigma}^{i^c}$ denote a punishment strategy against player i in an L^2-period game, *i.e.*, for each $\tau_i \in \Sigma_i$, $\gamma_i^{L^2}(\tau_i, \bar{\sigma}^i_{-i}) \leq v_i^{L^2}$.[18] At the start of the game, for each player i and community c, the corresponding state is $\mu(i, c) = 0$. This state will never be updated in payoff blocks.

During the first payoff block, B^1, the players play a sequence $(a^t)_t$ of fixed pure actions approximating the target payoff x. If L is large enough we can ensure that, for each $i \in N$, $|\frac{1}{L}\sum_{t \in B^1} g_i(a^t) - x_i| \leq \varepsilon$.

During the first communication block, D^1, the players communicate about what was played in B^1. During block D^1, each player i, for each community c, updates his state variable $\mu(i, c)$ as described before Lemma 3. This updating is such that $\mu(i, c) = 1$ if and only if player i is convinced that someone from community c has deviated from σ. We say that player i is "(as if) on the main path" if $\mu(i, c) = 0$ for each community c (including i^c). We say that player i is in the "punishing community c phase" if $\mu(i, c) = 1$ and $\mu(i, c') = 0$ for $c' \neq c$. Once a player is in a "punishing community c phase" he no longer updates his states.

During each subsequent payoff block B^l, with $l \geq 2$, play proceeds as follows. A player who starts the block "on the main path" plays as in B^1. A player i who starts the block in the "punishing community c phase", plays according to $\bar{\sigma}^c$ (starting with the empty history at the beginning of the block). During each subsequent communication block D^l, with $l \geq 2$, a player i who starts D^l in a "punishing community c phase" just communicates, without updating his states. A player who was on the main path at the block B^l communicates and updates his states. In cases not described above, prescribe play arbitrarily.

Now, let L be large enough so that the following conditions hold:

1) $f(L, |M|) \geq 1 - \varepsilon$, where f is the function whose existence is guaranteed by Lemma 3.

2) For each player i, $v_i^{L^2} \leq v_i^\infty + \varepsilon$ (recall that $\lim_{T \to \infty} v_i^T = \inf_T v_i^T = v_i^\infty$).

3) The lengths of the communication blocks are negligible with respect to the lengths of the payoff blocks, in the sense that $2L|C| \leq \varepsilon L^2$. This implies $L\varepsilon \geq 1$, so that the length of any block is large.

4) $R \leq \varepsilon L$, where $R = \max_{a \in A, c \in C} |u_c(a)|$.

Next, we examine the payoffs from playing σ and from deviating.

---

[18]Since everything is anonymous, punishing a player in a community is equivalent to punishing his community.

Suppose that σ is played. Then all players always are "on the main path". By the definition of $(a^t)_t$ and using 3) above, we have

$$|\gamma_i^T(\sigma) - x_i| \leq \frac{L(\varepsilon \cdot L^2 + 2R \cdot 2L|C|)}{L(L^2 + 2L|C|)} \leq \frac{\varepsilon L^2 + 2R\varepsilon L^2}{L^2 + 2L|C|} \leq \frac{\varepsilon L^2(1 + 2R)}{L^2} = \varepsilon(1 + 2R).$$

Suppose that some player $i \in N$ deviates to $\tau_i \in \Sigma_i$. All probabilities and expectations here are with respect to $\mathbb{P}_{(\tau_i, \sigma_{-i})}$. Denote by $\mathcal{Z}$ the event "player i always follows the path of pure actions in payoff blocks" and by $\bar{\mathcal{Z}}$ the complementary event.

First, suppose that $\mathcal{Z}$ holds. Then, all players except i will always stay on the main path because, by Lemma 3, a unilateral deviation by player i during a communication block cannot induce another player to go off the main path. As a consequence, play in payoff blocks proceeds as if σ was being played and we get

$$\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}(g_i(a^t)) \mid \mathcal{Z}) \leq x_i + \varepsilon(1 + 2R).$$

Next, assume that $\bar{\mathcal{Z}}$ holds. Let $\bar{l}$ be the first block number in $\{1, \ldots, L\}$ such that all players except i are on the main path at B^l and player i deviates at some period in B^l. With probability at least $(1 - \varepsilon)$, player i's deviation is detected in block $D^{\bar{l}}$ by all other players, so that his future expected payoff is below $v_i^\infty + \varepsilon$ in each future payoff block. Since $x \in IR^\infty$, $v_i^\infty + \varepsilon \leq x_i + \varepsilon$. There are three situations in which player i may have high payoffs: (i) during the payoff block $B^{\bar{l}}$, (ii) during communication blocks (proportion of periods at most ε), and (iii) if his deviation is not detected (which occurs with probability at most ε). We thus have

$$\mathbb{E}(\frac{1}{T}\sum_{t=1}^{T}(g_i(a^t)) \mid \bar{\mathcal{Z}}) \leq \overbrace{\varepsilon R}^{\text{iii)}} + (1-\varepsilon)\left(\overbrace{\varepsilon R}^{\text{ii)}} + (1-\varepsilon)\frac{(L-1)(x_i + \varepsilon) + \overbrace{R}^{\text{i)}}}{L}\right) \leq x_i + 2\varepsilon(1+R).$$

Thus, we have found a strategy profile σ and $T = L(L^2 + 2L|C|)$ so that, for each $i \in N$ and each $\tau_i \in \Sigma_i$, $\gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \varepsilon(1 + 2R)$ and $\gamma_i^T(\sigma) \geq x_i - \varepsilon(1 + 2R)$. Therefore, by Lemma 4, x can be sustained in a UE and, by Corollary 1, also in an SUE. $\qquad \square$

4.3   Folk theorem with imperfect private monitoring within a match

So far, we assumed that players perfectly observe the actions played in their own matches. An important question is whether this result is robust to monitoring imperfections. In this section, we go back to the two communities setting, but allow for imperfect monitoring within a match.

Recall that action sets are denoted by A_1 for community 1 and A_2 for community 2, and payoff functions by u_1 and u_2. Now we add sets of signals Y_1 and Y_2 for communities 1 and 2, respectively,

as well as a signaling function $\pi : A \longrightarrow \Delta(Y)$, where $Y = Y_1 \times Y_2$. Within a match, if the player from community 1 plays a_1 in A_1 and the player from community 2 plays a_2 in A_2, then a joint signal $(y_1, y_2) \in Y$ is drawn according to $\pi(a_1, a_2)$. The player from community 1 observes y_1 and the player from community 2 observes y_2. As usual, we assume that each player remembers his own actions. Signals may be stochastic and correlated across communities.

We want to characterize the subset E of $\mathbb{R}^2$ consisting of symmetric equilibrium payoffs of the repeated game. Recall that, with 2 communities, all minmax payoff vectors coincide and $IR = IRC = IR^\infty$. We provide a mild sufficient condition on the signal structure under which the folk theorem obtains.[19] We introduce the following observability condition.

Condition C: For each community $c \in \{1, 2\}$, each pure action $a_c \in A_c$, and each mixed action $s_c \in \Delta(A_c)$, if a_c and s_c are different, then there is a pure action $a_{\bar{c}}$ of the other community such that $\pi(a_c, a_{\bar{c}})$ and $\pi(s_c, a_{\bar{c}})$ have distinct marginals on the set of signals $Y_{\bar{c}}$ of community $\bar{c}$.

This condition requires that, for each pure action $a_c \in A_c$, the players in the other community can manage to distinguish a_c from any other (possibly mixed) action.[20] It turns out that Condition C above is sufficient to establish a folk theorem. Note that it does not imply that the players from a given community can statistically distinguish any two mixed actions of the other community.[21] In the context of standard repeated games, Kandori and Matsushima (1998) introduce a series of similar conditions, also related to the possibility of statistically distinguish pairs of actions. Also slightly related is the notion of "more informative actions" used, for instance, in Lehrer (1992c).

Theorem 4. *Suppose that Condition C is satisfied. Then, for all utility functions and community sizes, $E = F \cap IR$.*

Note that Theorem 2 is a particular case of this result. We present here the equilibrium construction and the intuition underlying the proof, which we develop formally in the Appendix. Recall that, because of Corollary 1, it suffices to show that $E^* = F \cap IR$, *i.e.*, it suffices to establish the result for uniform equilibrium payoffs.

Equilibrium construction: For simplicity, we consider a pure action $\bar{a} = (\bar{a}_1, \bar{a}_2) \in A$ such

---

[19]Note that some condition is needed to get a folk theorem. Think, for instance, of "trivial observation," when players observe nothing about their rivals' actions. Formally, Y_1 and Y_2 are singletons. In such games, E is the convex hull of the payoffs of the one-shot game. In order to sustain payoffs beyond the static Nash outcome, players need to observe something about the behavior of other players.

[20]For a simple example in which this condition is violated, suppose that $A_1 = \{T, M, B\}$, and players in community 2 get two possible signals y and y' and observe these independently of their own action. They observe y if $a_1 = M$, y' if $a_1 = B$ and the stochastic signal $\frac{1}{2}y + \frac{1}{2}y'$ if $a_1 = T$. Here, a player in community 2 cannot determine whether a player in community 1 is playing the pure action T or the mixed action $\frac{1}{2}M + \frac{1}{2}B$. The folk theorem may fail in this case.

[21]For instance, consider the following example. Suppose that $A_1 = \{T, M_1, M_2, B\}$. Players from community 2 observe three possible signals y_1, y_2 and y_3 (independently of their own action): they observe y_1 if $a_1 = T$, stochastic signal $\frac{1}{2}y_1 + \frac{1}{2}y_2$ if $a_1 = M_1$, $\frac{1}{2}y_1 + \frac{1}{2}y_3$ if $a_1 = M_2$, and $\frac{1}{2}y_2 + \frac{1}{2}y_3$ if $a_1 = B$. Here, Condition C is satisfied, and so the folk theorem will hold. However, the mixed actions $s_1 = \frac{1}{2}T + \frac{1}{2}B$ and $s_2 = \frac{1}{2}M_1 + \frac{1}{2}M_2$ both induce the signal $\frac{1}{2}y_1 + \frac{1}{4}y_2 + \frac{1}{4}y_3$ and thus are statistically indistinguishable for the players in community 2.

that $u(\bar{a}) \in IR$ and prove that $u(\bar{a}) \in E^*$.[22] The strategy σ is played in L' consecutive blocks, where all B^l blocks have the same length L. Thus, σ is defined in a repeated game with LL' periods. We define a "main path" and a "punishment phase." All players start on the main path. Let $R = \max_{a \in A, c \in C} |u_c(a)|$.

Punishment phase. Once a player is in the punishment phase, he minmaxes his opponents at every period independently of everything else, and never returns to the main path. Let $\hat{s}_i \in \Delta(A_i)$ denote the action that minmaxes community j, with $j \neq i$.

Main path. We now consider a player i from community 1. The strategy of a player in community 2 is defined analogously.

$$s_1 = (1 - \frac{1}{K}) \, \bar{a}_1 + \frac{1}{K} \, \mathcal{U}_1,$$

where $\mathcal{U}_1$ is the uniform distribution over A_1 and K is a constant to be fixed later. Let $L' = K$, $L = K^8$, and $\eta = \frac{1}{K^2}$. A player playing s_1 first selects $\bar{a}_1$ or $\mathcal{U}_1$ with respective probabilities $(1 - \frac{1}{K})$ and $\frac{1}{K}$, and in the latter case then selects an action according to the uniform law. Let $\mathbf{1}_i^t$ be the random variable that takes value one if player i has first selected $\mathcal{U}_1$ and value zero otherwise.

While on the main path, at the end of each block B^l, player i performs a statistical test to decide whether to enter the punishment phase or not. The test is described as follows. Let π_1 and π_2 denote the respective marginals of π over Y_1 and Y_2. Take $a_1 \in A_1$ and a signal $y \in Y_1$. Player i can compute the theoretical frequency of "playing a_1 via $\mathcal{U}_1$ and observing y when the opponent is playing s_2" as

$$TH(a_1, y) = \frac{1}{K|A_1|} \pi_1(y \mid a_1, s_2).$$

Player i can also compute the empirical frequency of "playing a_1 via $\mathcal{U}_1$ and observing y":

$$OB_i^l(a_1, y) = \frac{1}{L} |\{t \in B^l, a_i^t = a_1, \mathbf{1}_i^t = 1, y_i^t = y\}|.$$

Given this, player i uses the following simple test. If there are $a_1 \in A_1$ and $y \in Y_1$ such that $|TH(a_1, y) - OB_i^l(a_1, y)| \geq \eta$, then he enters the punishment phase. Otherwise, he remains on the main path.

This concludes the definition of σ. We need to establish that it constitutes a UE. In the proof we proceed in two steps. First, given that all players start out on the main path, we can use Tchebychev inequality to find a lower bound for the probability with which all players will remain on the main path and play according to s_1 or s_2 for all L' blocks. Indeed, we can show that, for K large enough, with high probability all players will remain on the main path and play according to s_1 or s_2 for all L' blocks, and therefore, the average expected payoff is approximately the target payoff. Second,

---

[22]In Footnote 28 in the Appendix we comment on the case in which the target payoff is not achievable in pure actions.

we show that no player can do much better by deviating. The idea here is that a deviating player cannot make a significant gain without being detected. For this part it is crucial that this player cannot condition on past history to influence the probabilities with which he is detected while still making some profit. The reason for this is that, within a block, regardless of what he plays, the behavior he faces is coming from the composition of i.i.d. draws.

It is worthwhile to highlight that Theorem 4 only provides a sufficient condition for the folk theorem. Finding a necessary and sufficient condition on the signal structure for the folk theorem to hold for any payoff function is a difficult problem. Indeed, the question is unanswered even in the case of 2 players (in our setting, 2 communities with a single player each), although Condition C is known not to be necessary (see the notion of more informative actions in Lehrer (1992a,b,c)).

5 Discussion

In this section, we present some interesting, qualitative observations based on our study of the set of SUE payoffs in the anonymous random matching setting.

5.1 Sustaining asymmetric payoffs

So far, we restricted attention to symmetric strategy profiles, which deliver the same expected payoff to all players within a community. Here we explore whether there is room for equilibria or social norms that sustain a certain number of free riders in a society. In such equilibria, social surplus may not be maximized, but this may still be better than full break down of cooperation. We study this possibility by considering the prisoner's dilemma, and show that it is possible to sustain equilibria in which a fixed group of people in the community are allowed to defect while the others cooperate. Formally, this means that while the average payoffs sustainable in symmetric SUE lie in $F \cap IR$, it is possible to get non-symmetric equilibria in which some players to get payoffs outside the set of feasible and individually rational payoffs of the stage game.[23]

$$
\begin{array}{cc}
 & \begin{array}{cc} C_2 & D_2 \end{array} \\
\begin{array}{c} C_1 \\ D_1 \end{array} & \left(\begin{array}{cc} (1,1) & (-1,2) \\ (2,-1) & (0,0) \end{array} \right)
\end{array}
$$

In the Appendix we prove the following result for the above version of the prisoner's dilemma.

---

[23]A similar observation was made by Dal Bó (2007), who obtains a folk theorem with respect to the $2M$-dimensional feasible set of "society payoffs". Dal Bó (2007) assumes either perfect information or local information transmission (Kandori, 1992; Okuno-Fujiwara and Postlewaite, 1995), whereas we consider a setting in which a player gets no information about matches not involving him.

This version: November 5, 2013

Theorem 5. *Let $M \in \mathbb{N}$ be the community size and let $K \in \mathbb{N}$, $K \leq \frac{M}{2}$. There is a non-symmetric SUE in which each player in community 1 gets payoff $1 - 2K/M$, K players in community 2 get payoff 2 and the remaining players in community 2 get payoff 1.*

It is worthwhile to point out that, though the result above establishes that some asymmetry in payoffs can be supported in equilibrium, we believe that it is unlikely that a full folk theorem can be achieved. The main difficulty in sustaining asymmetric payoffs is preventing the players who are supposed to get low payoffs from "impersonating" those players in the same community who get higher payoffs. In our setting, this entails two challenges: to identify that a deviation has occurred, and to punish appropriately. Since punishments occur even on the equilibrium path, we also have to ensure that players do not have strong incentives to defect.

To prove the result, for each $K \leq \frac{M}{2}$, we construct an equilibrium in which K players in one community are allowed to defect while the rest of society cooperates. Review blocks are needed to prevent (or control) deviations (impersonations). Notice that the higher K is, the greater is the loss in "social surplus". The aggregate payoffs go down from $2M$ if everybody cooperates, to $2M - K$ when K players in one of the two communities defect.

5.2 Use of correlated punishments

In Theorem 3 we established that the set of SUE payoffs is $F \cap IR^\infty$. Yet, we offer no explicit characterization of v^∞, which defines the relevant feasible and individually rational set. In general repeated games, computing this value is a hard problem, and is much beyond the scope of this paper.[24] Interestingly, we can show how in our setting of repeated anonymous random matching games it may happen that $v^\infty < v$ and, therefore, $F \cap IR$ is a strict subset of $F \cap IR^\infty$. Put differently, there is room for some correlation in punishments. Knowing how close we can get to perfectly correlated punishments is an open question.

We present below two examples to illustrate that $v^\infty < v$. In particular, we show that we can *force* a player $i \in N$ to a payoff x_i strictly below v_i (see Section 2.3). More precisely, there is $x_i < v_i$ such that, for each $\varepsilon > 0$, there are $\sigma_{-i} \in \Sigma_{-i}$ and $T_0 \in \mathbb{N}$ such that, for each $\tau_i \in \Sigma_i$ and each $T \geq T_0$, $\gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \varepsilon$.

Example 1. Suppose that there is perfect monitoring within a match and take the three player game

---

[24]It has been studied for special cases in Gossner and Tomala (2007); see also Gossner and Hörner (2010).

whose payoff function is given by the following matrices:

$$
\begin{array}{cc}
 & \begin{array}{cc} E_2 & W_2 \end{array} \\
\begin{array}{c} E_1 \\ W_1 \end{array} & \left(\begin{array}{cc} (0,0,-1) & (0,0,0) \\ (0,0,0) & (0,0,0) \end{array} \right) \\
 & E_3
\end{array}
\qquad
\begin{array}{cc}
 & \begin{array}{cc} E_2 & W_2 \end{array} \\
 & \left(\begin{array}{cc} (0,0,0) & (0,0,0) \\ (0,0,0) & (0,0,-1) \end{array} \right) \\
 & W_3
\end{array}
$$

Clearly, in this game the correlated minmax is $w = (0, 0, -1/2)$ and the independent minmax is $v = (0, 0, -1/4)$, with v_3 achieved when players 1 and 2 randomize uniformly between their two actions. Since $w_3 < v_3$, it may be possible to bring players in community 3 below payoff v_3 if players in communities 1 and 2 manage to use correlated punishments. In order to force a player i in community 3 to a payoff below v_3, the other players in his community will help players in communities 1 and 2 to correlate. Suppose that $M = 5$. Consider the following strategy profile σ, played in blocks of size $l + 1$, where l is to be determined.

- Each player in community 3 other than player i plays $W_3, \ldots, W_3$ or $E_3, \ldots, E_3$ with equal probabilities, in periods $1, \ldots, l$. This is unobservable to player i. Their choice for period $l+1$ is irrelevant.

- In the first l periods of each block, players in communities 1 and 2 minmax community 3, *i.e.*, they randomize uniformly between their two actions. In period $l+1$ a player in community $c \in \{1, 2\}$ plays W_c if he has seen W_3 more often than E_3 and, otherwise, plays E_c.

It is easy to check that, during each of the first l periods in each block, setting aside player i, with probability $5/16$, three or four players in community 3 will be playing E_3; with the same probability three or four players will be playing W_3; and, with probability $6/16$, two players will play E_3 and two players will play W_3. Unless we are in the latter case, as l becomes large, the players in communities 1 and 2 will successfully correlate to punish player i in period $l + 1$ (using the actions of players in community 3 to coordinate); regardless of what player i does. However, there is a trade-off. A larger l improves correlation, but increases the number of periods in which player i is "only" punished to his minmax level v_i. Clearly, the best reply of player i is to stick to one action, say W_3, during the first l periods (this increases the likelihood of communities 1 and 2 correlating on (W_1, W_2)) and then switch to the opposite action, E_3, in period $l + 1$. Yet, player i can only influence the choices in period $l + 1$ if the players in his own community were equally split between W_3 and E_3. Hence, as l increases, the probability of successful correlation goes to $10/16$, in which case player i would expect a payoff of $-1/2$ in period $l + 1$. Therefore, the expected payoff of

player i in each block is bounded above by

$$r(l) = \frac{l \cdot \frac{-1}{4} + 1 \cdot \frac{-1}{2} \frac{10}{16} f(l)}{l + 1}, \text{ where } \lim_{l \to \infty} f(l) = 1.$$

Hence, there is l_0 such that, for each $l > l_0$, $r(l) < -\frac{1}{4} = v_3$. In particular, one of the best choices of l in this example would be 81, leading to a utility for player i of at most $-\frac{501}{2000} < -\frac{1}{4}$. $\diamond$

Example 2. The stage game in this example is a kind of generalization of the one above, again assuming that there is perfect monitoring within a match. In the stage game there are three players and each player i has 5 actions $a_i^1, \ldots, a_i^5$. Payoffs to all players are $(0, 0, 0)$, except when all players chose the action with the same number, leading to a payoff of $(0, 0, -1)$. It is not hard to see that $w = (0, 0, -1/5)$ and $v = (0, 0, -1/25)$, so $w_3 < v_3$. Let $M = 2$.

We want to punish a player i in community 3. The strategy profile is played in blocks of only two periods. In the first period, players in communities 1 and 2 completely randomize and, in the second period, players in community 1 repeat the same action and players in community 2 copy the action observed in period 1. According to these strategies, in period 1, player i is minmaxed, and his expected payoff is $-1/25$. In period 2, with probability $1/4$ player i is matched with the two players that did not play against him on period 1. In this case, these two players will perfectly correlate and player i has no information about this correlation. Therefore, no matter what player i does, his expected payoff is bounded by $\frac{v_3 + \frac{1}{4} w_3}{2} = -\frac{9}{200} < -\frac{1}{25} = v_3$. $\diamond$

References

Boyd, R., P. J. Richerson. 2001. *Bounded Rationality: The Adaptive Toolbox*, chap. Norms and Bounded Rationality, 281–296. MIT Press, Cambridge.

———. 2002. Group Beneficial Norms can Spread Rapidly in a Structured Population. *Journal of Theoretical Biology* **215** 287–296.

Dal Bó, P. 2007. Social norms, cooperation and inequality. *Economic Theory* **30** 89–105.

Deb, J. 2012. Cooperation and Community Responsibility: A Folk Theorem for Random Matching Games with Names. Mimeo.

Deb, J., J. González-Díaz. 2013. Community Enforcement Beyond the Prisoner's Dilemma. Mimeo.

Ellison, G. 1994. Cooperation in the Prisoner's Dilemma with Anonymous Random Matching. *Review of Economic Studies* **61** 567–88.

Ely, J. C., J. Hörner, W. Olszewski. 2005. Belief-Free Equilibria in Repeated Games. *Econometrica* **73** 377–415.

Ely, J. C., J. Välimäki. 2002. A Robust Folk Theorem for the Prisoner's Dilemma. *Journal of Economic Theory* **102** 84–105.

Forges, F., J. F. Mertens, A. Neyman. 1986. A counterexample to the folk theorem with discounting. *Economics Letters* **20** 7.

Fudenberg, D., D. Levine. 1991. An approximate Folk theorem with imperfect private information. *Journal of Economic Theory* **54** 26–47.

Gossner, O., J. Hörner. 2010. When is the lowest equilibrium payoff in a repeated game equal to the minmax payoff? *Journal of Economic Theory* **145** 63–84.

Gossner, O., T. Tomala. 2007. Secret Correlation in Repeated Games with Imperfect Monitoring. *Mathematics of Operations Research* **32** 413–424.

———. 2009. Repeated Games with Complete Information. In R. Meyers, ed., *Encyclopedia of Complexity and Systems Science*. Springer.

Hasker, K. 2007. Social norms and choice: a weak folk theorem for repeated matching games. *International Journal of Game Theory* **36** 137–146.

Kandori, M. 1992. Social Norms and Community Enforcement. *Review of Economic Studies* **59** 63–80.

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

Lehrer, E. 1990. Nash Equilibria of n-Player Repeated Games with Semi-Standard Information. *International Journal of Game Theory* **19** 191–207.

———. 1992a. Correlated Equilibria in Two-Player Repeated Games with Nonobservable Actions. *Mathematics of Operations Research* **17** 175–199.

———. 1992b. On the Equilibrium Payoffs Set of Two Player Repeated Games with Imperfect Monitoring. *International Journal of Game Theory* **20** 211–226.

———. 1992c. Two-Player Repeated Games with Nonobservable Actions and Observable Payoffs. *Mathematics of Operations Research* **17** 200–224.

This version: November 5, 2013

Mailath, G. J., L. Samuelson. 2006. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press.

Mertens, J. F., S. Sorin, S. Zamir. 1994. Repeated Games. CORE discussion paper 9420. Lovain-la-Neuve.

Okuno-Fujiwara, M., A. Postlewaite. 1995. Social Norms and Random Matching Games. *Games and Economic Behavior* **9** 79–109.

Piccione, M. 2002. The Repeated Prisoner's Dilemma with Imperfect Private Monitoring. *Journal of Economic Theory* **102** 70–83.

Radner, R. 1980. Collusive Behavior in Noncooperative Epsilon-Equilibria of Oligopolies with Long but Finite Lives. *Journal of Economic Theory* **22** 136–154.

———. 1981. Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship. *Econometrica* **49** 1127–1148.

Renault, J., T. Tomala. 1998. Repeated proximity games. *International Journal of Game Theory* **27** 539–559.

———. 2011. General Properties of Long-Run Supergames. *Dynamic Games and Applications* **1** 319–350.

Sorin, S. 1986. On repeated games with complete information. *Mathematics of Operations Research* **11** 147–160.

Takahashi, S. 2010. Community enforcement when players observe partners' past play. *Journal of Economic Theory* **145** 42–62.

Tomala, T. 1999. Nash Equilibria of Repeated Games with Observable Payoff Vectors. *Games and Economic Behavior* **28** 310324.

A Appendix

A.1 Coincidence of UE and SUE payoff sets

This section is devoted to the proof of Theorem 1, which states that, for every infinitely repeated game with finite actions and signals, the sets of UE and SUE payoffs coincide.

Proof of Theorem 1. It is sufficient to prove that every UE payoff is an SUE payoff. Fix a uniform equilibrium payoff $x \in \mathbb{R}^N$ and define, for each length T and strategy profile σ,

$$\varepsilon_T(\sigma) = \max_{i \in N} \left\{ |\gamma_i^T(\sigma) - x_i|, \max_{\tau_i \in \Sigma_i}\{\gamma_i^T(\tau_i, \sigma_{-i}) - x_i\} \right\} \geq 0.$$

The condition $\varepsilon_T(\sigma) \leq \varepsilon$ implies that σ is a 2ε-Nash equilibrium of the T-period game with payoff ε close to x. Let $\varepsilon_T = \min_{\sigma \in \Sigma} \varepsilon_T(\sigma)$.[25] In a $(T+T')$-repeated game it is possible to follow a strategy σ for T periods and then, independently, a strategy σ' for T' periods, and by this concatenation principle we have that $(T + T')\varepsilon_{T+T'} \leq T\varepsilon_T + T'\varepsilon_{T'}$. The sequence $(\varepsilon_T)_{T \in \mathbb{N}}$ being bounded, this implies its convergence to $\inf_{T \geq 1} \varepsilon_T$. Further, since x is a uniform equilibrium payoff, we have $\inf_{T \geq 1} \varepsilon_T = 0$. For each $L \in \mathbb{N}$, let σ^L be a profile achieving the minimum in the definition of ε_L. The payoff to player i if σ^L is played for L periods is $\gamma_i^L(\sigma^L) \in [x_i - \varepsilon_L, x_i + \varepsilon_L]$.

We now construct σ, an SUE with payoff x. Play proceeds in consecutive blocks of increasing length, B^L. Each block B^L is itself divided into L^2 consecutive subblocks $B^{L,1}, B^{L,2}, \ldots, B^{L,L^2}$. Each of these subblocks has length L, so that the length of the block B^L is L^3. At each subblock $B^{L,l}$, σ^L is played, independently of what has happened before.

We now prove that σ is an SUE. Let $R = \max_{a \in A, c \in C} |u_c(a)|$. Fix a player i in N and define, for each subblock $B^{L,l}$, its average payoff by

$$X_{L,l} = \frac{1}{L} \sum_{t \in B^{L,l}} g_i(a^t) \text{ and, similarly, let } X_L = \frac{1}{L^2} \sum_{l=1}^{L^2} X_{L,l}.$$

Both $X_{L,l}$ and X_L are bounded random variables defined in $[-R, R]$. Now we prove the three conditions in Definition 2.[26]

Condition 1) Assume that σ is played by all players. At a given block B^L, the strategy σ^L is repeated L^2 times, so that $(X_{L,l})_{l \in \{1, \ldots, L\}}$ is a sequence of i.i.d. random variables with expectation $\gamma_i^L(\sigma^L) \in [x_i - \varepsilon_L, x_i + \varepsilon_L]$. By Tchebychev inequality, we obtain, for each $\varepsilon > 0$,

$$\mathbb{P}_\sigma(|X_L - \gamma_i^L(\sigma^L)| \geq \varepsilon) \leq \frac{R^2}{\varepsilon^2 L^2} \quad \text{and, thus,} \quad \mathbb{P}_\sigma(|X_L - x_i| \geq \varepsilon + \varepsilon_L) \leq \frac{R^2}{\varepsilon^2 L^2}.$$

Since ε_L goes to 0 and $\sum_{L \geq 1} \frac{R^2}{\varepsilon^2 L^2} < \infty$, by Borel-Cantelli lemma, with probability 1 there is L_0 such that, for each $L \geq L_0$, $|X_L - x_i| \leq 2\varepsilon$. As a consequence, $\lim_{L \to \infty} X_L = x_i$ $\mathbb{P}_\sigma$-a.s. Now, it easily follows that $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T g_i(a^t) = x_i$ $\mathbb{P}_\sigma$-a.s. Further, since the payoffs are bounded, we can apply Lebesgue's convergence theorem to get $\lim_{T \to \infty} \gamma_i^T(\sigma) = x_i$.

---

[25] The strategy sets are compact (endowed with the product topology) and the payoff functions are continuous.

[26] Note that Condition 1) already implies $\lim_{T \to \infty} \gamma^T(\sigma) = x$; it suffices to take the empty history in 1) and then, since the random variables are bounded, almost sure convergence to x_i implies convergence on average.

This version: November 5, 2013

Condition 2) Assume now that player i deviates to τ_i, whereas the other players follow σ. In what follows, all expectations or probabilities are computed with respect to the probability $\mathbb{P}_{(\tau_i, \sigma_{-i})}$.

Let G^L denote the L-period repeated game with average payoffs. G^L is nothing but a finite game with payoffs in $[-R, R]$, and the block B^L may be viewed as the repetition of the game G^L during L^2 "stages" (each stage now corresponding to a subblock of B^L, hence to L original periods). At each "stage" in which game G^L is played, the players different from player i play i.i.d. a mixed action profile corresponding to the first L periods of the strategy σ_{-i}^L. By definition of σ^L, this mixed action forces the expected payoff of player i to be at most $x_i + \varepsilon_L$ at each stage of B^L. As a consequence, by the strategic Tchebychev inequality (Lemma 1) we have that, for each $\varepsilon > 0$,

$$\mathbb{P}_{(\tau_i, \sigma_{-i})}(X_L \geq x_i + \varepsilon_L + \varepsilon) \leq \frac{R^2}{\varepsilon^2 L^2}.$$

Again by Borel-Cantelli lemma, with probability 1 there is L_0 such that, for each $L \geq L_0$, $X_L \leq x_i + 2\varepsilon$ and, thus, $\limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} g_i(a^t) \leq x_i$ $\mathbb{P}_{(\tau_i, \sigma_{-i})}$-a.s.

Condition 3) We establish this property of σ for the initial history; because of the structure of σ, the proof after any other history is analogous. For each L we have, by definition of σ^L, that $\mathbb{E}(X_L) \leq x_i + \varepsilon_L$. Fix $\varepsilon > 0$ and let L_0 be such that, for each $\bar{L} \geq L_0$, $\varepsilon_{\bar{L}} \leq \varepsilon$. Now, let L_1 be such that, for $\hat{L} \geq L_1$,[27]

$$\sum_{L=1}^{L_0-1} L^3 \leq \varepsilon \sum_{L=L_0}^{\hat{L}-1} L^3 \text{ and } \hat{L}^3 \leq \varepsilon \sum_{L=L_0}^{\hat{L}-1} L^3.$$

Let T be large enough so that period T belongs to a block $B^{\hat{L}}$ with $\hat{L} \geq L_1$. Then,

$$T\gamma_i^T(\tau_i, \sigma_{-i}) \leq \sum_{L=1}^{L_0-1} L^3 R + \sum_{L=L_0}^{\hat{L}-1} L^3(x_i + \varepsilon) + R\hat{L}^3 \text{ and, hence, } \gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \varepsilon(1 + 2R). \quad \square$$

A.2 Proof of Theorem 4

Proof. Given a pure action profile $\bar{a} = (\bar{a}_1, \bar{a}_2) \in A$ such that $u(\bar{a}) \in IR$, in Section 4.3 we defined a strategy profile σ. We now rely on Lemma 4 to show that σ is a UE with payoff $u(\bar{a})$. Let $R = \max_{a \in A, c \in C} |u_c(a)|$.[28] Recall that the number of blocks is $L' = K$, the length of each block is $L = K^8$, and $\eta = \frac{1}{K^2}$, where K is a large integer to be determined.

Payoff from playing σ. Suppose that σ is played, and consider a block B^l in which all players

---

[27]To see that such an $\hat{L}$ exists, just recall that $\sum_{L=1}^{\hat{L}} L^3 = \left(\frac{\hat{L}(\hat{L}+1)}{2}\right)^2$.

[28] In case the target payoff in $F \cap IR$ is not achievable in a pure action profile, one can approximate it by a cycle of pure actions, and then the proof will essentially go through, doing statistical tests independently for each element of the cycle, and minmaxing at every single period while in the punishment phase.

This version: November 5, 2013

are on the main path. For each player i in community 1, each action $a_1 \in A^1$, and each signal y we have, by Tchebychev inequality, that $\mathbb{P}_\sigma(|TH(a_1, y) - OB_i^l(a_1, y)| \geq \eta) \leq \frac{1}{\eta^2 L}$. We can obtain an analogous inequality for players in community 2. Then, we have that, with probability at least $1 - \frac{M(|A_1||Y_1| + |A_2||Y_2|)}{\eta^2 L}$, all players start block B^{l+1} on the main path. Thus, when playing according to σ, all players remain on path throughout the game with probability at least $1 - \alpha$, where $\alpha = \frac{(L'-1)M(|A_1||Y_1| + |A_2||Y_2|)}{\eta^2 L}$. Take a player i in community c. Then, for each block l,

$$\left| \mathbb{E}_{\mathbb{P}_\sigma}\left(\frac{1}{L} \sum_{t \in B^l} g_i(a^t)\right) - u_i(s_1, s_2) \right| \leq 2\alpha R, \text{ so } \left| \mathbb{E}_{\mathbb{P}_\sigma}\left(\frac{1}{LL'} \sum_{t=1}^{LL'} g_i(a^t)\right) - u_i(\bar{a}_1, \bar{a}_2) \right| \leq \left(2\alpha + \frac{4}{K}\right)R.$$

Thus, for K large enough, we have $\left| \mathbb{E}_{\mathbb{P}_\sigma}\left(\frac{1}{LL'} \sum_{t=1}^{LL'} g_i(a^t)\right) - u_i(\bar{a}_1, \bar{a}_2) \right| \leq \frac{5R}{K}$.

Payoff from deviating from σ. Suppose that a player i in community 1 deviates to some strategy τ_i (the argument for community 2 is analogous). All probabilities and expectations here are with respect to $\mathbb{P} = \mathbb{P}_{(\tau_i, \sigma_{-i})}$. We now define a couple of random variables. Let $m(t) \in \{1, \ldots, M\}$ denote the identity of the player in community 2 facing player i in period t. Let β^l denote the number of players from community 2 in the punishment phase when block B^l starts. Clearly, β^l takes values in $\{0, \ldots, M\}$ and $\beta^1 = 0$. We denote by $\bar{s}_i^l$ the empirical frequency played by player i in block B^l, that is, $\bar{s}_i^l = \frac{1}{L} \sum_{t \in B^l} a_i^t \in \Delta(A_1)$. Let $X_l = \frac{1}{L} \sum_{t \in B^l} g_i(a^t) - u_i(\bar{a})$. The main goal in the rest of the proof is to control X_l. Given $\varepsilon > 0$, by the strategic Tchebychev inequality (Lemma 1) we have that, with probability at least $1 - \frac{R^2}{\varepsilon^2 L}$,

$$\left| \frac{1}{L} \sum_{t \in B^l} g_i(a^t) - u_i(\bar{s}_i^l, s_2^l) \right| \leq \varepsilon, \text{ where } s_2^l = \frac{(M - \beta^l)s_2 + \beta^l \hat{s}_2}{M}. \tag{1}$$

The above inequality means that player i's payoff in block B^l can be approximately computed as if he was playing the mixed action $\bar{s}_i^l$ i.i.d. during the block.

We now apply Lemma 1 again in a more sophisticated way. We want to show that, no matter how player i is supposed to play (according to τ_i), he cannot do much better than what he would by playing according to $\bar{s}_i^l$ i.i.d. during the block; non only in terms of payoffs on this block, but also regarding future payoffs, *i.e.*, regarding the number of players in the other community who will enter the punishment phase. Thus, to some extent, all player i can choose is the empirical frequency $\bar{s}_i^l$: there is no significant gain to be made by using the history of play. The reason for this is that the behavior he faces is coming from the composition of i.i.d. draws.

To see this, fix a player $\bar{m}$ in community 2 who is on the main path at block B^l, an action a_2 in A_2, and a signal y_2 in Y_2. Consider a (non-strategic) fictitious "superplayer" 2 who just carries out all the randomizations that may be needed at each period of the block. He selects the identity $m(t)$ of the player from community 2 meeting player i at period t, the action in A_2 he plays as σ_2

31

would, and also a realization of π_2 given the chosen actions; randomizations across periods are i.i.d. Consider a new two-player game in which player i acts as player 1 and the fictitious superplayer acts as player 2. The payoff function for player i is an indicator function that takes on value 1 if "the superplayer selects $m(t) = \bar{m}$, a_2 via the uniform law, and y_2," and takes on value 0 otherwise. Applying Lemma 1 to these strategies and payoff function, we get[29]

$$\mathbb{P}\left(\left|\frac{1}{L}\sum_{t\in B^l}\mathbf{1}_{m(t)=\bar{m},\mathbf{1}^t_{\bar{m}}=1,a^t_{\bar{m}}=a_2,y^t_{\bar{m}}=y_2} - \frac{1}{MK|A_2|}\pi_2(y_2\mid \bar{s}^l_i,a_2)\right|\geq\varepsilon\right)\leq\frac{1}{\varepsilon^2 L}. \quad (2)$$

The above inequality means that the law of the number of stages in B^l where: "player i meets player $\bar{m}$, player $\bar{m}$ selects the uniform distribution $\mathcal{U}_2$, plays a_2, and observes signal y_2" can also be approximately computed as if player i was playing $\bar{s}^l_i$ i.i.d. during the block.

Let α^l denote the number of players in community 1, different from i, who are in the punishment phase at block B^l. This random variable takes values in $\{0,\ldots,M-1\}$ and, whenever player $\bar{m}$ does not meet the deviating player i, he faces $M-1-\alpha^l$ players playing s_1 and α^l players using $\hat{s}_1$. As a consequence, if we let $Q = \alpha^l\pi_2(y_2\mid \hat{s}_1,a_2)+(M-1-\alpha^l)\pi_2(y_2\mid s_1,a_2)$, we have, by Tchebychev inequality,

$$\mathbb{P}\left(\left|\frac{1}{L}\sum_{t\in B^l}\mathbf{1}_{m(t)\neq\bar{m},\mathbf{1}^t_{\bar{m}}=1,a^t_{\bar{m}}=a_2,y^t_{\bar{m}}=y_2} - \frac{Q}{MK|A_2|}\right|\geq\varepsilon\;\Big|\;\alpha^l\right)\leq\frac{1}{\varepsilon^2 L}. \quad (3)$$

Summing up equations (2) and (3) we have that, with probability at least $1-\frac{2|Y_2||A_2|}{\varepsilon^2 L}$, given α^l

$$\forall a_2, \forall y_2, \left|OB^l_{\bar{m}}(a_2,y_2) - \frac{1}{K|A_2|}\pi_2\left(y_2\;\Big|\;\frac{\alpha^l}{M}\hat{s}_1+\frac{(M-1-\alpha^l)}{M}s_1+\frac{1}{M}\bar{s}^l_i,a_2\right)\right|\leq\varepsilon. \quad (4)$$

For each $\alpha\in\{0,\ldots,M-1\}$, let $\hat{\varepsilon}_\alpha$ be the function defined, for each $r\geq 0$, by

$$\hat{\varepsilon}_\alpha(r) = \max\left\{u_i(s_1,s_2)-u_i(\bar{a}_1,s_2), s_2\in\Delta(A_2),\; s_1\in\Delta(A_1)\text{ s.t. }\forall y_2\in Y_2,\forall a_2\in A_2,\right.$$

$$\left.\left|\pi_2(y_2\mid\bar{a}_1,a_2)-\pi_2\left(y_2\;\Big|\;\frac{\alpha\hat{s}_1+(M-1-\alpha)\bar{a}_1+s_1}{M},a_2\right)\right|\leq r\right\}.$$

Notice that $\hat{\varepsilon}_\alpha$ does not depend on the parameters K, L, L', η. The function $\hat{\varepsilon}_\alpha$ is non-decreasing, and because Condition C holds and $\bar{a}_1$ is a pure action, we have $\lim_{r\to 0}\hat{\varepsilon}_\alpha(r) = 0$. Assume now player $\bar{m}$ does not enter punishment phase after block B^l. Then, for all a_2 and y_2, $|TH(a_2,y_2) - $

---

[29]Recall that $\mathbf{1}^t_{\bar{m}}$ denotes the random variable that takes value one if player $\bar{m}$ has first selected $\mathcal{U}_2$ and value zero otherwise.

This version: November 5, 2013

$OB^l_{\bar{m}}(a_2, y_2)| \leq \eta$ and by Inequality (4) and the definition of $TH(a_2, y_2)$ we have

$$\frac{1}{K|A_2|}\left|\pi_2\left(y_2 \mid s_1, a_2\right) - \pi_2(y_2 \mid \frac{\alpha^l \hat{s}_1 + (M - 1 - \alpha^l)s_1 + \bar{s}^l_i}{M}, a_2\right)\right| \leq \eta + \varepsilon.$$

We now use the definition of s_1 to get

$$\left|\pi_2(y_2 \mid \bar{a}_1, a_2) - \pi_2\left(y_2 \mid \frac{\alpha^l \hat{s}_1 + (M - 1 - \alpha^l)\bar{a}_1 + \bar{s}^l_i}{M}, a_2\right)\right| \leq K|A_2|(\eta + \varepsilon) + 2/K.$$

This implies, by definition of $\hat{\varepsilon}_\alpha$, that, for each $s_2 \in \Delta(A_2)$,

$$u_i(\bar{s}^l_i, s_2) - u_i(\bar{a}_1, s_2) \leq \hat{\varepsilon}_{\alpha^l}(K|A_2|(\eta + \varepsilon) + 2/K).$$

We use $s_2 = s^l_2 = \frac{(M - \beta^l)s_2 + \beta^l \hat{s}_2}{M}$ and get from Equation (1) that, with probability at least $1 - \frac{R^2}{\varepsilon^2 L}$,

$$\frac{1}{L}\sum_{t \in B^l} g_i(a^t) \leq \varepsilon + u_i(\bar{s}^l_i, s^l_2) \ \leq \ u_i(\bar{a}_1, s^l_2) + \varepsilon + \hat{\varepsilon}_{\alpha^l}(K|A_2|(\eta + \varepsilon) + 2/K)$$

$$\leq u_i(\bar{a}_1, \bar{a}_2) + 2R/K + \varepsilon + \sup_\alpha \hat{\varepsilon}_\alpha(K|A_2|(\eta + \varepsilon) + 2/K)$$

where the last inequality uses the facts that $\hat{s}_2$ minmaxes community 1 and $\bar{a}$ is IR.

Summing up, we have proved that for each block number l and each $\varepsilon > 0$,

$$\mathbb{P}\big(\beta^{l+1} = \beta^l, X_l > 2R/K + \varepsilon + \sup_\alpha \hat{\varepsilon}_\alpha(K|A_2|(\eta + \varepsilon) + 2/K)\big) \leq \frac{R^2 + 2|Y_2||A_2|}{\varepsilon^2 L} \quad (5)$$

We now fix $\varepsilon = \eta$ and use $\eta = \varepsilon = \frac{1}{K^2}$, $L = K^8$, $K = L'$. Then, given an error $\varepsilon' > 0$, one can find K_0 such that for each value of $K \geq K_0$, Equation (5) implies that, for each block l,

$$\mathbb{P}(\beta^{l+1} = \beta^l, X_l > \varepsilon') \leq \varepsilon'. \quad (6)$$

Now consider K large enough so that $K \geq K_0$ and $\frac{5R}{K} \leq \varepsilon'$ (recall the no deviation case). Then, for each block l,

$$\mathbb{E}(X_l \mathbf{1}_{\beta^{l+1}=\beta^l}) = \mathbb{P}(X_l > \varepsilon', \beta^{l+1} = \beta^l)\mathbb{E}(X_l \mid X_l \geq \varepsilon', \beta^{l+1} = \beta^l)$$

$$+ \ \mathbb{P}(X_l \leq \varepsilon', \beta^{l+1} = \beta^l)\mathbb{E}(X_l \mid X_l < \varepsilon', \beta^{l+1} = \beta^l) \ \leq \ (2R + 1)\varepsilon'.$$

Notice now that the size of a community is fixed, so that $\sum_{l=1}^{L'} \mathbf{1}_{\beta^{l+1} > \beta^l} \leq M$.

$$\mathbb{E}(\frac{1}{L'}\sum_{l=1}^{L'} X_l) = \mathbb{E}(\frac{1}{L'}\sum_{l=1}^{L'} X_l \mathbf{1}_{\beta^{l+1}=\beta^l} + \frac{1}{L'}\sum_{l=1}^{L'} X_l \mathbf{1}_{\beta^{l+1}>\beta^l}) \leq (2R + 1)\varepsilon' + \frac{RM}{K}$$

33

If $\frac{RM}{K} \leq \varepsilon'$, the strategy profile σ is a symmetric $2(R+1)\varepsilon'$-Nash equilibrium of the finitely repeated game with LL' periods, with payoff ε'-close to $u(\bar{a})$. By Lemma 4, $u(\bar{a})$ is a UE payoff. $\square$

A.3 Proof of Theorem 5

Proof. By Lemma 4, in order to establish that a payoff x is a UE payoff, it suffices to find, for each $\varepsilon > 0$, $T \in \mathbb{N}$ and a strategy profile σ of the T-period repeated game such that, for each $i \in N$ and each $\tau_i \in \Sigma_i$, $\gamma_i^T(\tau_i, \sigma_{-i}) \leq x_i + \varepsilon$ and $\gamma_i^T(\sigma) \geq x_i - \varepsilon$. Then, Theorem 1 ensures that x is also an SUE payoff.

Denote by $N^{C,1}$ the set of players in community 1, by $N^{D,2}$ a fixed set of K players in community 2 and by $N^{C,2}$ the set of remaining players in community 2. Let L be a large number. We construct a strategy profile σ to be played in consecutive blocks, alternating B^l and D^l blocks, $l \in \{1, \ldots, L\}$. Each block B^l lasts L^5 periods, while each block D^l lasts L periods. Thus, the total length of the game is $T = L(L^5 + L)$.

Each player $i \in N$ is in state $\mu_i \in \{0, 1\}$. The interpretation of $\mu_i = 0$ is that player i is "on the main path", *i.e.*, he thinks that no one has deviated yet, and $\mu_i = 1$ means that player i is "in the punishment phase", *i.e.*, he thinks that it is likely that a deviation has occurred. All players start with $\mu_i = 0$. A player in the punishment phase never returns to the main path.

The B^l blocks are used for payoffs. In such a block, while on the main path, all players cooperate except for those in $N^{D,2}$, who defect in every period. In the punishment phase, every player defects at every period of B^l. The D^l blocks are used for communication. At each period of such a block, players on the main path cooperate and players in the punishment phase defect.

It remains to describe how players update their states. A player i in community 2 switches to $\mu_i = 1$ if and only if he faces a defection. A player i in community 1 switches to $\mu_i = 1$ in two cases. First, he switches to $\mu_i = 1$ if he faces a defection during a communication block. Second, at the end of a payoff block B^l, he computes the frequency of observed defections at B^l and switches to $\mu_i = 1$ if this frequency is higher than $K/M + 1/L$. Note that, if no player deviates, this frequency should to be close to K/M.

We show that, given $\varepsilon > 0$, for large enough L, σ is an ε-Nash equilibrium of the T-period repeated game. We impose the following two conditions on L:

1) Suppose that at most one player i deviates from σ and some player $j \neq i$ starts a communication block D^l in the punishment phase, *i.e.*, $\mu_j = 1$. Then, L is large enough so that at the end of block D^l, with probability at least $1 - \varepsilon$, every player $m \in N \backslash \{i\}$ satisfies that $\mu_m = 1$.

2) Length L is such that $M/L \leq \varepsilon$. In particular, this implies that $L/L^5 = 1/L^4 \leq \varepsilon$, *i.e.*, the length of the communication blocks is small with respect to the total number of periods. It also implies a low probability of ever reaching a punishment phase if σ is played.

Payoff from playing σ. Consider a block B^l in which all players are on the main path and fix a player $i \in N^{C,1}$. Let X_i^t be the random variable given by $X_i^t = 1$ if player i faced a defection at period t and $X_i^t = 0$ otherwise. The sequence $(X_i^t)_{t \in B^l}$ is an i.i.d. sequence of Bernoulli random variables with parameter K/M. Hence, by Tchebychev inequality,

$$\mathbb{P}_\sigma \left(\left| \frac{1}{L^5} \sum_{t \in B^l} X_i^t - K/M \right| \geq 1/L \right) \leq \frac{1}{L^3}.$$

As a consequence, each player $i \in N^{C,1}$ will be on the punishment phase after B^l with probability at most $1/L^3$. Thus, with probability at least $1 - M/L^3$, all players will start block B^{l+1} on the main path. Further, with probability at least $1 - M/L^2$, all players will be on the main path throughout the T periods of the repeated game. Since $1 - M/L^2 \geq 1 - \varepsilon$, the payoffs induced by σ are close to the target payoff. More precisely, close to 2 for players in $N^{D,2}$, close to 1 for players in $N^{C,2}$ and close to $1 - 2K/M$ for players in $N^{C,1}$. Below, we explicitly compute the payoffs for a player $i \in N^{C,1}$. We omit the other cases, as they are similar and simpler.

$$\gamma_i^T(\sigma) \leq (1 - \frac{M}{L^2}) \cdot L \frac{L^5 \left(-1 \cdot (\frac{K}{M} - \frac{1}{L}) + 1 \cdot (1 - \frac{K}{M} + \frac{1}{L}) \right) + L}{L(L^5 + L)} + 2 \cdot \frac{M}{L^2}$$

$$= (1 - \frac{M}{L^2}) \frac{L^5(1 - \frac{2K}{M} + \frac{2}{L}) + L}{L^5 + L} + \frac{2M}{L^2} \leq \frac{(L^5 + L)(1 - \frac{2K}{M} + \frac{2}{L}) + L}{L^5 + L} + \frac{2M}{L^2}$$

$$= 1 - \frac{2K}{M} + \frac{2}{L} + \frac{2M}{L^2} \leq 1 - \frac{2K}{M} + 3\varepsilon,$$

and

$$\gamma_i^T(\sigma) \geq (1 - \frac{M}{L^2}) \cdot L \frac{L^5 \left(-1 \cdot (\frac{K}{M} + \frac{1}{L}) + 1 \cdot (1 - \frac{K}{M} - \frac{1}{L}) \right) + L}{L(L^5 + L)} - 1 \cdot \frac{M}{L^2}$$

$$= (1 - \frac{M}{L^2}) \frac{L^5(1 - \frac{2K}{M} - \frac{2}{L}) + L}{L^5 + L} - \frac{M}{L^2} \geq \frac{L^5(1 - \frac{2K}{M} - \frac{2}{L})}{L^5 + L} - \frac{M}{L^2} - \frac{M}{L^2}$$

$$= \frac{(L^5 + L)(1 - \frac{2K}{M}) - L\frac{2K}{M} - 2L^4}{L^5 + L} - \frac{M}{L^2} - \frac{M}{L^2} = 1 - \frac{2K}{M} - 5\varepsilon.$$

Payoff from deviating from σ. Suppose that some player $i \in N$ deviates to $\tau_i \in \sigma_i$. All probabilities and expectations here are with respect to $\mathbb{P}_{(\tau_i, \sigma_{-i})}$.

Case 1. Since 2 is the highest payoff in the game, no deviation by a player in $N^{D,2}$ is profitable.

Case 2. Suppose $i \in N^{C,1}$. Since deviations of player i are immediately detected, no statistical test is needed. Let Z be the random variable denoting the first payoff block in which player i defected at some period; if no such block exists, $Z = L + 1$. In the latter case, the payoff to player i

This version: November 5, 2013

cannot be high:

$$\mathbb{E}_{\mathbb{P}_{(\tau_i, \sigma_{-i})}}\Big(\frac{1}{T}\sum_{t=1}^{T} g_i(a^t) \mid Z = L+1\Big) \le \frac{L^5(1 - \frac{2K}{M}) + 2L}{L^5 + L} \le 1 - \frac{2K}{M} + 2\varepsilon.$$

In general, at each payoff block before block Z, the expected payoff of player i is at most $1 - \frac{2K}{M} + 2\varepsilon$. Recall that L has been chosen to ensure that, with probability at least $1 - \varepsilon$, at each block after B^Z, all players different from player i defect in every period, thus giving a non-positive payoff to player i. Then, for each $z \in \{1, \ldots, L\}$, $\mathbb{E}_{\mathbb{P}_{(\tau_i, \sigma_{-i})}}\big(\frac{1}{T}\sum_{t=1}^{T} g_i(a^t) \mid Z = z\big)$ is bounded above by

$$\underbrace{2\varepsilon}_{\substack{\text{com.}\\\text{blocks}}} + (1 - \varepsilon)\Big(\underbrace{2\varepsilon}_{\substack{\text{no proper}\\\text{punishment}}} + (1 - \varepsilon)\frac{1}{L}\Big(\underbrace{(z-1)(1 - \frac{2K}{M} + 2\varepsilon)}_{\substack{\text{blocks before}\\\text{deviation}}} + \underbrace{2}_{\substack{\text{dev.}\\\text{block}}} + \underbrace{(L-z)0}_{\substack{\text{blocks after}\\\text{deviation}}}\Big)\Big).$$

The above expression is maximized for $z = L$ and, therefore,

$$\gamma_i^T(\tau_i, \sigma_{-i}) = \mathbb{E}_{\mathbb{P}_{(\tau_i, \sigma_{-i})}}\Big(\frac{1}{T}\sum_{t=1}^{T} g_i(a^t)\Big) \le 1 - \frac{2K}{M} + 8\varepsilon.$$

Case 3. Suppose $i \in N^{C,2}$. Deviations by players in $N^{C,2}$ may be profitable and hard to detect. During the first payoff block, B^1, all players in $N\setminus\{i\}$ are on the main path. Thus, players in $N^{C,1}$ play C_1 throughout this block. Let $a_i^1, a_i^2, \ldots, a_i^{L^5}$ be the (possibly random and correlated) sequence of actions played by i in B^1. For the sake of exposition, hereafter we use $a_i^t = 1$ to denote that i plays D_2 at period t and use $a_i^t = 0$ otherwise. Fix $j \in N^{C,1}$ and let X_j^t be the random variable such that $X_j^t = 1$ if j has observed D_2 at period t, and $X_j^t = 0$ otherwise. Now, by the strategic Tchebychev inequality (Lemma 1),[30] we have that, for each $\varepsilon > 0$,

$$\mathbb{P}_{(\tau_i, \sigma_{-i})}\Big(\Big|\frac{1}{L^5}\sum_{t\in B^1} X_j^t - \frac{1}{L^5}\sum_{t\in B^1}\frac{(K + a_i^t)}{M}\Big| \ge \varepsilon\Big) \le \frac{1}{\varepsilon^2 L^5}.$$

Now, suppose that player i makes a significant profit by deviating at block B^1. More precisely, assume that $\frac{1}{L^5}\sum_{t\in B^1}\frac{a_i^t}{M} \ge 2/L$. By taking $\varepsilon = 1/L^2$ in the above equation we get that, with probability at least $1 - 1/L$, $\big|\frac{1}{L^5}\sum_{t\in B^1} X_j^t - \frac{1}{L^5}\sum_{t\in B^1}\frac{(K+a_i^t)}{M}\big| < 1/L^2$. Hence, $\frac{1}{L^5}\sum_{t\in B^1} X_j^t \ge K/M + 2/L - 1/L^2 > K/M + 1/L$. Therefore, we have shown that, if player i defects at least $2L^4 M$ times during block B^1, then, with probability at least $1 - 1/L$, player j will be in the punishment phase at the end of B^1. Therefore, with probability at least $(1 - 1/L)(1 - \varepsilon) \ge (1 - \varepsilon)^2$, all players in $N\setminus\{i\}$ will be in the punishment phase from block B^2 onwards.

---

[30] We cannot directly apply Tchebychev inequality because the a_i^t actions of player 1 may not be independent of each other and, thus, the X_j^t variables may not be independent either.

The analysis just done for block B^1 can naturally be applied to any block B^l in which all players in $N\backslash\{i\}$ are on the main path. The rest of the argument is similar to the one in Case 2. Let Z be the random variable denoting the first payoff block in which player i has defected at least $2L^4M$ times; if no such period exists, $Z = L + 1$. The expected average payoff to player i in a payoff block strictly larger than Z is at most $2(1 - (1 - \varepsilon)^2) \leq 4\varepsilon$. At any block payoff strictly smaller than Z, his average payoff is not more than $1 + \frac{2L^4M}{L^5} \leq 1 + 2\varepsilon$. Given these bounds, the most favorable case for player i is $Z = L$ and we obtain

$$\gamma_i^T(\tau_i, \sigma_{-i}) \leq \underbrace{2\varepsilon}_{\substack{\text{com.} \\ \text{blocks}}} + (1 - \varepsilon)\frac{1}{L}\Big(\underbrace{(L - 1)(1 + 2\varepsilon)}_{\substack{\text{blocks before} \\ \text{deviation}}} + \underbrace{2}_{\substack{\text{dev.} \\ \text{block}}} \Big) \leq 1 + 6\varepsilon.$$

We have proved that σ is an 8ε-Nash equilibria of the T-period repeated game with payoff 5ε-close to the target payoff. $\qquad\square$

This version: November 5, 2013