PAPERS on Economics & Evolution

MAX-PLANCK-GESELLSCHAFT

0417

**Strategic Interaction and
Externalities:
FD-games and pollution**

by

Reinoud Joosten

Strategic Interaction and Externalities: FD-games and pollution[*]

Reinoud Joosten[†]

Abstract

To analyze strategic interaction which may induce externalities, we designed Bathroom Games with frequency-dependent stage payoffs. Two people regularly use a bathroom, before leaving they can either clean up the mess made, or not. Cleaning up involves an effort, so this option always gives a lower immediate utility than not cleaning up.

The immediate utility of using the bathroom depends on its condition: the cleaner it is, the higher the utility. The pollution at a certain point in time depends on how often the players did not clean up in the past. Furthermore, as the bathroom's condition deteriorates, cleaning up becomes more burdensome, leading to increasing disutilities.

We follow the analysis of repeated games and find that if the agents are sufficiently patient, individually-rational rewards can be supported by (subgame perfect) equilibria involving threats. In almost every such equilibrium, the bathroom is cleaned up regularly.

JEL-Codes: C72, C73

Keywords: frequency-dependent stage payoffs, average rewards, repeated and stochastic games, Folk Theorems

1 Introduction

We study a social dilemma or social trap resulting from strategic interaction which may induce (negative) externalities over time. Such a situation may occur if an economic activity causes externalities which destroys its own profitability, e.g., pollution of an environment exploited for recreational purposes. Dawes [1980] states that two properties define a social dilemma: (a) each individual receives a higher payoff for a socially defecting choice than for a socially cooperative choice, no matter what other individuals in society do, and (b) all individuals are better off if all cooperate than if all defect. A social trap is a social dilemma, but with an effect in the time-dimension.

Certain behavior leads to a small positive outcome which is immediate, and a large negative outcome which is delayed (see Platt [1973]).

Bathroom Games describe rather stylized externality problems concerning pollution. Each player has two actions 'cleaning up' or 'not cleaning up' in each stage game. The strategic interaction between players is such that the latter action gives a strictly higher immediate payoff to either player regardless of the choice made by the opponent. There also exists however, strategic interaction on a different time-scale, through externalities which become evident after some time. The less often the bathroom is cleaned up, the lower the stage payoffs on all actions because of a degradation of the bathroom's condition. An essential feature of the game is that cleaning up becomes more and more burdensome relative to not cleaning up, the more polluted the bathroom is. The reason why this may occur is that the costs of cleaning up increase in the presence of more pollution, but also emotional factors such as disgust, indignation or irritation, may lead to decreasing utilities. In more general settings an increase in pollution may pose more dangers to the person(s) cleaning up.

Bathroom Games are special instances of games with frequency-dependent stage payoffs (Joosten *et al.* [2003]). The idea behind this notion stems from the work of the psychologist and economist Herrnstein on distributed choice in which (experimental) stimuli changed depending on the choices made by the experimental subjects (see e.g., Herrnstein [1997]). Until recently, these effects had been examined primarily in 'one-person games' or 'games against nature' settings. The first contribution using the concept in a (multi-person) strategic-interaction framework was Brenner & Witt [2003].

In Joosten *et al.* [2003] an extensive analysis of games with frequency-dependent stage payoffs was undertaken. Although such games form a sub-class of stochastic games introduced by Shapley [1953], the analysis of infinitely repeated games can very well be adapted to this type of games. Joosten *et al.* [2003] derive several Folk-Theorem-like results. If the agents are sufficiently patient, all individually-rational rewards can be supported by an equilibrium involving threats. An individually-rational reward gives a player at least the reward he can guarantee himself, regardless of his opponent's action.

Here, we demonstrate as well that certain, but not all, individually-rational rewards can be supported by (subgame perfect) equilibria involving threats. In almost every such equilibrium, the bathroom is cleaned up regularly, but not necessarily always. If one player were to punish the other under a 'grim-trigger strategy', one of two situations may arise: either the punishee cleans up forever while the punisher never cleans up, or nobody ever cleans up again. Which situation then arises is determined by a parameter used in the model. The parameter indicates that, for identical levels of pollution of the environment, cleaning up becomes more and more detrimental to the utilities derived for increasing values of this parameter. Replacing

2

'grim-trigger' strategies with strategies which are more forgiving in case of unilateral deviations, and which increase the stream of stage payoffs to the punisher, are shown to induce almost the same set of equilibrium rewards.

The organization of the paper is as follows. In the next section, we present the model. In Section 3, we focus on feasible rewards in a Bathroom Game, especially rewards connected to strategies with certain restrictions. In Section 4, we present sets of rewards which can be connected to equilibria with threats. Section 5 is devoted to reflections on the results found. The Appendix contains proofs of the results presented.

2   The Bathroom Game

We start by giving an illustration of the problem to be analyzed and present the main concepts as we go along. To keep our account tractable, we perform the analysis on the one-parameter family of Bathroom Games below.

Example 1 *The Bathroom Game.* *Aaron and Batsheva frequently use a bathroom. After each visit, each has two options: to clean up the mess made or not. The utility which either person derives from using the bathroom, depends on its condition, the cleaner the bathroom the higher the utilities. Cleaning up is burdensome, while the alternative requires no effort at all, moreover cleaning up a tidy bathroom is less burdensome than cleaning up a soiled bathroom. The following formalizes these ideas. The payoffs at stage t in state s_t are given by:*

$$A(s_t) = \begin{bmatrix} 6 - 4\rho_t^A - 4\rho_t^B & 6 - 4\rho_t^A - 4\rho_t^B \\ 4 - \gamma\rho_t^A - \gamma\rho_t^B & 4 - \gamma\rho_t^A - \gamma\rho_t^B \end{bmatrix} = B(s_t)^\top.$$

The row (column) player is Aaron (Batsheva), 'not cleaning up' ('cleaning up') is action 1 (2), $A(s_t)$ ($B(s_t)$) denotes Aaron's (Batsheva's) payoff matrix in state s_t. As usual the top row (left column) corresponds to Aaron's (Batsheva's) action 1. The interpretation of the payoff matrices is that if Aaron uses action 1, i.e., he does not clean up, and Batsheva uses action 2, i.e., she cleans up, the immediate payoff to Aaron is $6 - 4\rho_t^A - 4\rho_t^B$ and the immediate payoff to Batsheva is $4 - \gamma\rho_t^A - \gamma\rho_t^B$. Each player's stage payoff only depends on the own action.
The state in which the game is at stage t is denoted by $s_t \equiv \left(\rho_t^A, \rho_t^B\right)$, where ρ_t^k is the relative frequency with which player $k = A, B$, played action 1 ('not cleaning up') until stage t. More precisely, for $k = A, B$, let $\rho_t^k = \frac{\#\{j_u^k = 1 | u < t\}}{t}$. The bathroom's condition at stage t is measured by the average frequency with which both players have not cleaned up until stage t, i.e., $\rho_t^A + \rho_t^B$.
Since $\gamma \geq 4$, stage payoffs to both players clearly decrease as the bathroom's condition deteriorates. Moreover, the difference between the utilities for cleaning up and not cleaning up increases as the condition of the bathroom deteriorates. Figure 1 illustrates the effects of the pollution on the stage payoffs. ∎

Formally, a Bathroom Game is played by two players at discrete moments in time called stages. Each player has two actions and at each stage each player chooses an action. The payoffs at stage $t' \in \mathbb{N}$ of the play depend on the choices of the players at that stage, and on the relative frequencies with which all actions where chosen until then.

Let $h_{t'}^A = \left(j_1^A, ..., j_{t'-1}^A \right)$ be the sequence of actions chosen by Aaron until stage $t' \geq 2$ and let $q \geq 0$, then define ρ_t^A recursively for $t \leq t'$ by

$$
\begin{aligned}
\rho_1^A &= \rho^A \in [0,1], \\
\rho_t^A &= \begin{cases} \frac{q+t-1}{q+t} \rho_{t-1}^A + \frac{1}{q+t} & \text{if } j_{t-1}^A = 1, \\ \frac{q+t-1}{q+t} \rho_{t-1}^A & \text{if } j_{t-1}^A = 2. \end{cases}
\end{aligned}
\tag{1}
$$

Define ρ_t^B for Batsheva similarly. Taking $q \gg 0$ serves to moderate 'early' effects on the stage payoffs. Recall that j_{t-1}^A denotes the action chosen by Aaron at stage $t-1$, hence, the number ρ_t^A converges in the long run to the relative frequency with which he chose his first action before stage t, regardless of the numbers ρ^A and q.

At each stage a bi-matrix game is played, and the choices of the players realized at that stage determine the stage payoffs to the players, as well as the matrix game to be played at the next stage. With respect to the payoff matrices $(A(s_t), B(s_t))$, we assume the following. Suppose the play is in state $s_t \equiv \left(\rho_t^A, \rho_t^B \right)$ at stage $t \in \mathbb{N}$ and action pair $(i,j) \in J \equiv J^A \times J^B \equiv \{1,2\} \times \{1,2\}$ is chosen, then Aaron receives a stage payoff equal to $a_{ij}(s_t)$ the entry ij of matrix $A(s_t)$ and Batsheva receives $b_{ij}(s_t)$. Entries $a_{ij}(s_t)$ and $b_{ij}(s_t)$ are jointly linear in ρ_t^A, ρ_t^B, i.e.,

$$
\begin{aligned}
a_{ij}(s_t) &= a_{ij} - \alpha_{ij}^A \rho_t^A - \alpha_{ij}^B \rho_t^B, \\
b_{ij}(s_t) &= b_{ij} - \beta_{ij}^A \rho_t^A - \beta_{ij}^B \rho_t^B.
\end{aligned}
\tag{2}
$$

The coefficients in Eq. (2) are fixed and finite, and the following restrictions are assumed to hold in any Bathroom Game for all $i,j \in \{1,2\}$, $k \in \{A,B\}$:

$$
a_{1j} > a_{2j}, \ b_{i1} > b_{i2}, \ \alpha_{2j}^k \geq \alpha_{1j}^k \geq 0, \ \beta_{i2}^k \geq \beta_{i1}^k \geq 0.
$$

These restrictions ensure that in the stage games 'not cleaning up' always strictly dominates 'cleaning up', i.e., always gives a strictly higher immediate utility than the alternative. Furthermore, choosing an action pair in a perfectly clean bathroom, i.e., $\rho_t^A = \rho_t^B = 0$, yields higher utilities than choosing the same action pair for higher values of ρ_t^A and ρ_t^B. This reflects the idea that a deteriorating condition of the bathroom decreases utilities.

The final typical feature of a Bathroom Game is that the more polluted the bathroom is, the bigger the discrepancies between the utilities of 'cleaning up' and 'not cleaning up' become. Aaron and Batsheva may differentiate between who made the mess, even though objectively the condition of the

bathroom might be the same.[1] Figure 1 illustrates the effects of the condition of the bathroom on the stage payoffs. Note that 'cleaning up' does not involve cleaning up the entire bathroom, only a limited effort is made.
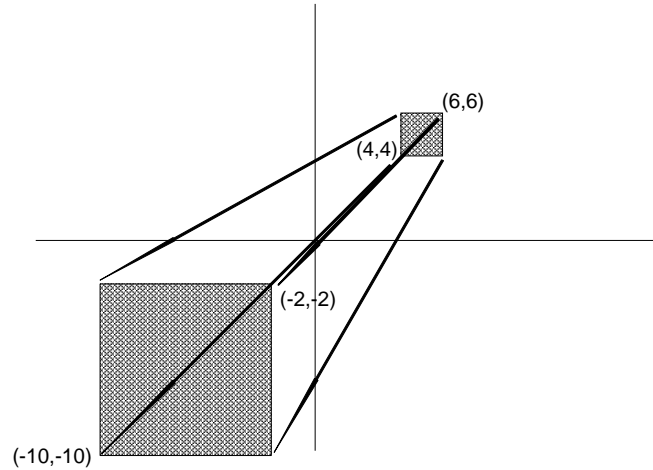

Figure 1: The feasible stage payoffs for the Bathroom Game with $\gamma = 7$. The arrows indicate that the stage payoffs decrease as the condition of the bathroom deteriorates.

3 Strategies and rewards

At every stage t, both players know the current state and the history of play,[2] i.e., the state visited and actions chosen at each stage before. A **strategy** prescribes at all stages, for any state and history, a mixed action to be used by a player. The sets of all strategies for Aaron respectively Batsheva will be denoted by $\mathcal{X}^A$ respectively $\mathcal{X}^B$, and $\mathcal{X} \equiv \mathcal{X}^A \times \mathcal{X}^B$. The (stochastic) payoff to player k, $k = A, B$, at stage t, depends on the strategy-pair $(\pi, \sigma) \in \mathcal{X}$, the **expected stage payoff** is denoted by $R_t^k(\pi, \sigma)$.

The players receive an infinite stream of stage payoffs during the play, and they are assumed to wish to maximize their average rewards. For a given pair of strategies (π, σ), player k's **average reward,** $k = A, B$, is given by $\gamma^k(\pi, \sigma) = \liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} R_t^k(\pi, \sigma)$; $\gamma(\pi, \sigma) \equiv (\gamma^A(\pi, \sigma), \gamma^B(\pi, \sigma))$.

---

[1] Many people are considerably more annoyed by the mess made by others!

[2] Allowing for instance that Aaron can not observe Batsheva's action(s) directly, adds a layer of reality, but also several layers of complexity. For repeated games with incomplete information, see e.g. Hart [1984], Forges [1986].

It may be quite hard to determine the **set of feasible (average) rewards** F, directly. It is not uncommon in the analysis of repeated or stochastic games to limit the scope of strategies on the one hand, and to focus on rewards on the other. Here, we will do both, we focus on rewards from strategies which are pure and jointly convergent. Then, we extend our analysis from there to obtain more feasible rewards.

A strategy is **pure**, if at *each* stage a **pure action** is chosen, i.e., the action is chosen with probability 1. The set of pure strategies for player k is $\mathcal{P}^k$, and $\mathcal{P} \equiv \mathcal{P}^A \times \mathcal{P}^B$. The strategy pair $(\pi, \sigma) \in \mathcal{X}$ is **jointly convergent** if and only if $z \in \Delta^{m \times n}$ exists such that for all $\varepsilon > 0$:

$$\limsup_{t \to \infty} \mathrm{Pr}_{\pi,\sigma} \left[\left| \frac{\#\{j_u^A = i \text{ and } j_u^B = j | \ 1 \leq u \leq t\}}{t} - z_{ij} \right| \geq \varepsilon \right] = 0 \text{ for all } (i,j) \in J,$$

where $\Delta^{m \times n}$ denotes the set of all nonnegative $m \times n$-matrices such that the entries add up to 1, hence $z_{ij} \in [0,1]$; $\mathrm{Pr}_{\pi,\sigma}$ denotes the probability under strategy-pair (π, σ). $\mathcal{JC}$ denotes the set of jointly-convergent strategy pairs. Under a pair of jointly-convergent strategies, the relative frequency of action pair $(i,j) \in J$ converges with probability 1 to z_{ij} in the terminology of Billingsley [1986, p.274]. Moreover, the empirical distribution of the past play by Aaron under such a pair of strategies converges with probability 1 to the vector given by the row-sums of the matrix z. Hence, ρ_t^A converges with probability 1 to Z^A, i.e., the sum of the first row of the matrix z. Similar remarks hold with respect to the other player.

The **set of jointly-convergent pure-strategy rewards** is given by

$$P^{\mathcal{JC}} \equiv cl \left\{ \left(x^1, x^2\right) \in \mathbb{R}^2 | \ \exists_{(\pi,\sigma) \in \mathcal{P} \cap \mathcal{JC}} : \left(\gamma^k\left(\pi, \sigma\right), \gamma^k\left(\pi, \sigma\right)\right) = \left(x^1, x^2\right) \right\},$$

where $cl \ S$ is the closure of the set S. The interpretation of this definition is that for any pair of rewards in this set, we can find a pair of jointly-convergent pure strategies that yield rewards arbitrarily close to the original pair of rewards. $P^{\mathcal{JC}}$ can be determined rather conveniently, as we will show now. With respect to jointly-convergent strategies, Eq. (2) and the arguments presented imply that $\lim_{t \to \infty} (a_{ij}(s_t), b_{ij}(s_t)) = (a_{ij} - \alpha_{ij}^A Z^A - \alpha_{ij}^B Z^B, b_{ij} - \beta_{ij}^A Z^A - \beta_{ij}^B Z^B)$. So, the matrices $A(s_t), B(s_t)$ 'converge' in the long run, too.

Let $\varphi(z) \equiv \sum_{(i,j) \in J} z_{ij} \left(a_{ij} - \alpha_{ij}^A Z^A - \alpha_{ij}^B Z^B, \ b_{ij} - \beta_{ij}^A Z^A - \beta_{ij}^B Z^B\right)$. The interpretation of $\varphi(z)$ is that under jointly-convergent strategy pair (π, σ) the relative frequency of action pair $(i,j) \in J$ being chosen is z_{ij} and each time this occurs the players receive $(a_{ij}(s_t), b_{ij}(s_t))$ in the long run. Hence, the players receive an average amount of $\varphi(z)$. So, $\gamma(\pi, \sigma) = \varphi(z)$.

The following result has been proven in Joosten *et al.* [2003] for games with frequency-dependent stage payoffs. Less general ideas had been around earlier for the analysis of repeated games with vanishing actions (cf., Joosten [1996, 2001], Schoenmakers *et al.* [2002]).

Lemma 2 *In the Bathroom Game, we have* $P^{\mathcal{JC}} = \bigcup_{z \in \Delta^{m \times n}} \varphi(z)$. *More-over, each pair of rewards in the convex hull of* $P^{\mathcal{JC}}$ *is feasible.*

This lemma provides a convenient algorithm to determine $P^{\mathcal{JC}}$ and its convex hull. Below, we have depicted the sets $P^{\mathcal{JC}}$ for different cases with respect to the parameter γ.
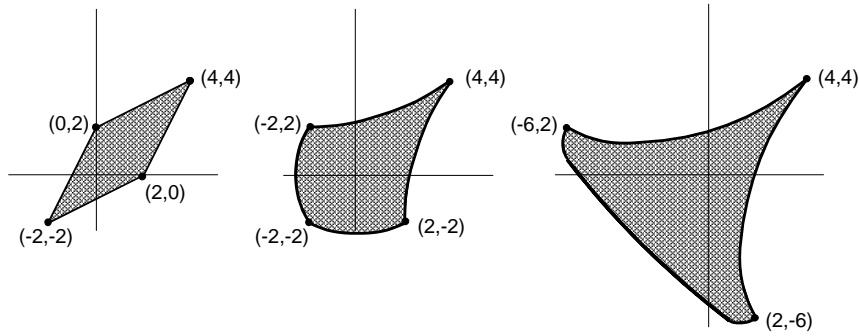

Figure 2: The sets of jointly-convergent pure-strategy rewards for $\gamma = 4$ (left), $\gamma = 6$ (middle) and $\gamma = 10$ (right).

To give the intuition on how to construct strategies which yield a convex combination of both rewards, we provide the following example.

Example 3 *Note that in Figure 2 (where* $\gamma = 7$*) no strict convex combination of rewards* $(-3, 2)$ *and* $(4, 4)$ *can be achieved by a pair of jointly-convergent pure strategies. Let* (π, σ) *be given by*

$$
\begin{array}{lll}
\pi_t = \sigma_t = \left(\frac{1}{2}, \frac{1}{2}\right) & \text{for } t = 1, 2 & \\
\pi_t = \sigma_t = 2 & \text{for } t \geq 3 & \text{if } j_1^A = j_1^B \text{ and } j_2^A = j_2^B \\
\pi_t = 2, \ \sigma_t = 1 & \text{for } t \geq 3 & \text{otherwise.}
\end{array}
$$

At the first two stages, each player randomizes with equal probability on both actions. Then, the play continues with probability $\frac{1}{4}$ *with* $((2, 2), (2, 2), (2, 2), ...)$ *and with the complementary probability with* $((2, 1), (2, 1), (2, 1), ...)$*, because the event* $j_1^A = j_1^B$ *and* $j_2^A = j_2^B$ *has probability* $\frac{1}{4}$*. Furthermore, the first sequence gives long run average payoffs of* $(4, 4)$ *and the second one* $(-3, 2)$*. Hence,* $\gamma(\pi, \sigma)$ *yields rewards* $\frac{1}{4}(4, 4) + \frac{3}{4}(-3, 2)$*. Randomizing over two periods in this manner*

and proceeding according to the outcome similarly gives all convex combinations of $(-3, 2)$ *and* $(4, 4)$ *being multiples of* $\frac{1}{4}$. *Moreover, following the same procedure for a randomization of three periods we can generate all multiples of* $\frac{1}{8}$; *in general, for* T *periods one gets all multiples of* $\frac{1}{2}^T$. $\blacksquare$

Joosten *et al.* [2004] provide a constructive method to obtain pure-strategy rewards which may be outside of the convex hull of $P^{\mathcal{JC}}$. In order to present an analogy, we first establish the players' worst outcomes in $P^{\mathcal{JC}}$.

Lemma 4 *The worst outcome in* $P^{\mathcal{JC}}$ *for a player is* $\min\{-2, 4 - \gamma + \frac{1}{4-\gamma}\}$.

The intuition behind the constructive method mentioned before the lemma is as follows. For $\gamma = 7$, let $\pi^1 = (1, 1, 1, ...)$ and $\sigma^1 = (2, 2, 2, ...)$; similarly, let $\pi^2 = (2, 2, 2, ...)$ and $\sigma^2 = (1, 1, 1, ...)$. Then, $\gamma\left(\pi^1, \sigma^1\right) = (2, -3)$ and $\gamma\left(\pi^2, \sigma^2\right) = (-3, 2)$. Now, construct a pair of strategies (π, σ) such that the average stage payoffs approach $(2, -3)$ by playing according to $\left(\pi^1, \sigma^1\right)$, next by playing according to $\left(\pi^2, \sigma^2\right)$ to approach $(-3, 2)$ more closely, followed by playing according to $\left(\pi^1, \sigma^1\right)$ again to approach $(2, -3)$ even more closely, and so on. Then, clearly (π, σ) is not jointly-convergent and $\liminf_{t \to \infty} \frac{1}{T} \sum_{t=1}^{T} R_t^k(\pi, \sigma) \leq -3$, $k = 1, 2$. Obviously, $(-3, -3)$ is not in the convex hull of $P^{\mathcal{JC}}$.

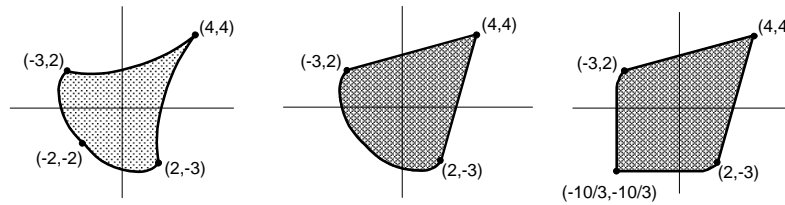Figure 3: An illustration for $\gamma = 7$. Left: the set of jointly-convergent pure-strategy rewards $P^{\mathcal{JC}}$; middle: the convex hull of $P^{\mathcal{JC}}$; right: the set P'.

Theorem 5 *Let* $w \in \mathbb{R}^2$ *be the pair of worst outcomes in* $P^{\mathcal{JC}}$, *and let* P' *be the convex hull of* w *and* $P^{\mathcal{JC}}$. *Then every reward in* P' *is feasible.*

4   Threats and equilibria

The strategy pair (π^*, σ^*) is an **equilibrium**, if no player can improve by unilateral deviation, i.e.,

$$\gamma^A(\pi^*, \sigma^*) \geq \gamma^A(\pi, \sigma^*), \ \gamma^B(\pi^*, \sigma^*) \geq \gamma^B(\pi^*, \sigma) \text{ for all } \pi \in \mathcal{X}^A, \sigma \in \mathcal{X}^B.$$

An equilibrium is called **subgame perfect** if for each possible state and history the subsequent play corresponds to an equilibrium, i.e., no player can improve by deviating unilaterally from then on. In the construction of equilibria for repeated games, 'threats' play an important role. A threat specifies the conditions under which one player will punish the other, as well as the subsequent measures.

We call $v = \left(v^A, v^B\right)$ the **threat point**, where $v^A = \min_{\sigma \in \mathcal{X}^B} \max_{\pi \in \mathcal{X}^A} \gamma^A(\pi, \sigma)$, and $v^B = \min_{\pi \in \mathcal{X}^A} \max_{\sigma \in \mathcal{X}^B} \gamma^B(\pi, \sigma)$. So, v^A is the highest amount Aaron can get if Batsheva tries to minimize his average payoffs. Under a pair of **individually rational** (feasible) rewards each player receives at least the threat-point reward. We have the following result for the Bathroom Game.

Lemma 6 *We have* $v = (\max\{-2, 4 - \gamma\}, \max\{(-2, 4 - \gamma\})$.

Example 7 *Consider the pair of strategies in which Aaron and Batsheva clean up alternately, starting with Aaron. This implies that $\rho^A = \rho^B = \frac{1}{2}$ in the long run. As Batsheva's long run stage payoffs (for $\gamma = 7$) are alternately $6 - 4\left(\rho^A + \rho^B\right) = 2$ and $4 - 7\left(\rho^A + \rho^B\right) = -3$, her average stage payoffs converge to $-\frac{1}{2}$. A similar statement holds for Aaron.*

Suppose Aaron were to deviate unilaterally, and clean up exactly thrice every four stages. Then, $\rho^A = \frac{1}{4}$ and $\rho^B = \frac{1}{2}$ in the long run. Aaron then receives once $6 - 4\left(\frac{1}{4} + \frac{1}{2}\right) = 3$ and thrice $4 - 7\left(\frac{1}{4} + \frac{1}{2}\right) = -\frac{5}{4}$ in every four stages. This leads to average stage payoffs of $\frac{1}{4} \cdot 3 + \frac{3}{4} \cdot -\frac{5}{4} = -\frac{3}{16} > -\frac{1}{2}$. Hence, Aaron can improve unilaterally against Batsheva's strategy.

Consider however, the following pair of strategies. Aaron and Batsheva clean up alternately, starting with Aaron, as long as both players stick to cleaning up alternately. If one player deviates from this course of action, the other player is never to clean up again. This pair of strategies leads to exactly the same sequence of play: on odd stages Aaron cleans up and on even stages Batsheva cleans up, inducing rewards $-\frac{1}{2}$ to both. If Aaron were to deviate only once, then Batsheva would 'punish' him by never cleaning up again, and Aaron's reward is at most -2. Hence, Aaron can not improve unilaterally.

The 'threat' is that Batsheva will never clean up again if Aaron deviates even once from cleaning up on odd-numbered stages. Observe that in such an equilibrium involving threats, the play is such that the threats are never carried out. ∎

To present the general idea of the next result of Joosten *et al.* [2003], we adopt terms from Hart [1985] and Forges [1986]. First, there is a 'master

plan' which is followed by each player as long as the other does too; then there are 'punishments' which come into effect if a deviation from the master plan occurs. The master plan is a sequence of 'intra-play communications' between the players, the purpose of which is to decide by which equilibrium the play is to continue. The outcome of the communication period is determined by a 'jointly controlled lottery', i.e., at each stage of the communication period the players randomize with equal probability on the first two actions; at the end of the communication period one sequence of pairs of action choices materializes. Detection of deviation from the master plan *after* the communication period is easy as both players use pure actions on the equilibrium path from then on. Deviation in the communication period by using another action than one of the first two actions can also be easily detected, but deviation by using an *alternative randomization* on the first two actions is impossible to detect. It can be shown that no alternative unilateral randomization yields a higher reward. So, the outcome of the procedure is an equilibrium. For more details, we refer to Joosten *et al.* [2003]. We restate here the major result which applies to general games with frequency-dependent stage payoffs.

Theorem 8 *(Joosten, Brenner & Witt [2003]) Each pair of rewards in the convex hull of all individually-rational pure-strategy rewards can be supported by an equilibrium. Moreover, each pair of rewards in the convex hull of all pure-strategy rewards giving each player strictly more than the threat-point reward, can be supported by a subgame-perfect equilibrium.*

Corollary 9 *Let $E' = \{(4 - \gamma, 4 - \gamma), (4 - \gamma, 2), (2, 4 - \gamma), (4, 4)\}$ for $\gamma \in [4, 6]$ and $E' = \{(-2, -2), \left(-2, -2 + \frac{48(\gamma - 3)}{\gamma^2}\right), \left(-2 + \frac{48(\gamma - 3)}{\gamma^2}, -2\right), (4, 4)\}$ for $\gamma \geq 6$ Then, each pair of rewards in the set E' can be supported by an equilibrium. Moreover, all rewards in E' giving Aaron more than v^A and Batsheva more than v^B can be supported by a subgame-perfect equilibrium.*

5   Interpretations and reflections

The final result of the previous section can be seen as a Folk Theorem. Appreciation for such results varies widely among the profession. On the negative side we find Gintis [2000] for instance, where it is stated (p.129): 'By explaining practically anything, the model in fact explains nothing'. Gintis then expresses a cautious preference for Pareto-efficiency as a selection criterion, as well as a preference for more realistic punishments in case of unilateral deviation from an equilibrium path, which should be more forgiving, e.g., allow for repair. On the positive side, Osborne & Rubinstein [1994] point out that equilibria of the infinitely repeated game exist which are Pareto-superior to any equilibrium of the associated one-shot game.
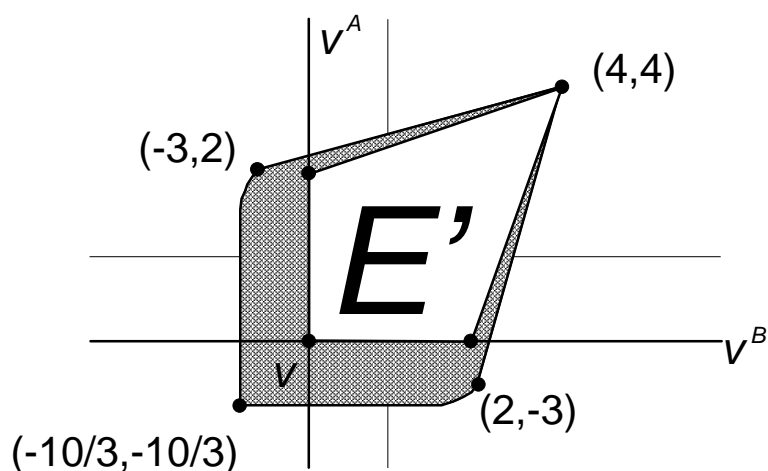
Figure 4: Corollary 9 illustrated for $\gamma = 7$. The shaded area represents P'; all rewards in E' are equilibrium rewards, all rewards in its interior can be supported by a subgame perfect equilibrium.

An outcome X is called Pareto-superior to outcome Y, if under X at least one agent is better off than under Y, while all others are equally well-off. A feasible outcome is Pareto-efficient if there exists no feasible outcome which is Pareto-superior to it. Here, it is quite meaningless to compare equilibrium rewards of the infinitely repeated game with equilibrium payoffs of a stage game, because we have infinitely many stage games. Therefore, we will apply these notions exclusively to rewards in the infinitely repeated game. The unique Pareto-efficient outcome is the pair of rewards $(4, 4)$ and it is also an equilibrium reward. What is particularly striking is that these rewards can only be obtained if the Pareto-inferior action in the stage games is played with a long-run relative frequency going to one by both players. Moreover, for every pair of equilibrium rewards except $(4, 4)$ an outcome can be found which is Pareto-superior to it, but which can not be supported by an equilibrium.

From a strictly theoretical point of view no value judgement seems possible on the issue of punishing unilateral deviations. Threats and punishments are quite possible in repeated games, removing these options seems artificial. However, the idea of punishments is hard to sell to the more practically oriented, despite the fact that in equilibrium *threats are never acted out*. The type of punishment which is often used in the construction of equilibria, is of the grim-trigger type (see e.g., Van Damme [1991]). Under such a strategy one deviation, even an inadvertent one, triggers the grimmest punishment

possible inducing the threat-point reward to the player being punished, no matter what the consequences are for the rewards of the punisher. Of course, subgame-perfectness aims to take away this objection. After a unilateral deviation, the play is to continue with an equilibrium giving the deviating player a lower reward. Often however, the threat-point rewards coincide with equilibrium rewards, so the objection may still be present.

The richness of the strategy space allows us to accomodate however, several of the objections connected to equilibria involving threats. The following shows that we can construct subgame-perfect equilibria which in case of a unilateral deviation are 'forgiving', i.e., they allow (not too many) deviations, and if a deviator is punished the player punishing is better off afterwards. However, this does not reduce the set of equilibria significantly.

Theorem 10 *For any pair* $(a, b) \in int\left(E' \cap P^{\mathcal{JC}}\right)$ *a subgame-perfect equilibrium* (π, σ) *exists yielding rewards* (a, b) *which induces play such that*

- *if Aaron deviates from the equilibrium path 'too often', then play proceeds according to an equilibrium such that Aaron receives strictly less than the amount* a *but more than* v^A *and Batsheva receives at least* b;

- *if Batsheva deviates from the equilibrium path 'too often', then play proceeds according to an equilibrium such that Batsheva receives strictly less than* b *but more than* v^B *and Aaron receives at least* a.

The 'too often'-s are well specified in the proof. The convexification result of Joosten *et al.* [2003] can also be applied to $int\left(E' \cap P^{\mathcal{JC}}\right)$.

6   Appendix

Proof of Lemma 4. We only prove the case $w^A = \min\{-2, 4 - \gamma + \frac{1}{4-\gamma}\}$. Aaron's rewards for a given pair of jointly-convergent pure strategies are

$$R(\rho, \varphi) = (\gamma - 4)\rho^2 + (\gamma\varphi - \gamma - 4\varphi + 2)\rho + (4 - \gamma\varphi).$$

Here, $\rho = \lim_{t\to\infty} \rho_t^1$ and $\varphi = \lim_{t\to\infty} \rho_t^2$. As $\partial R(\rho, \varphi)/\partial\varphi = (\gamma - 4)\rho - \gamma < 0$, $\varphi = 1$ minimizes Aaron's reward. Observe that $\partial R(\rho, 1)/\partial\rho \leq 0$ for $\rho \leq \frac{1}{\gamma-4}$ and $\partial R(\rho, 1)/\partial\rho \geq 0$ for $\rho \geq \frac{1}{\gamma-4}$ and $\partial^2 R(\rho, 1)/\partial^2\rho = 2(\gamma - 4) > 0$ for $\gamma > 4$. So, $R(\rho, \varphi)$ is minimized for $\rho = \min\left\{1, \frac{1}{\gamma-4}\right\}$ and $\varphi = 1$. Then straightforward substitution yields the statement of the case.   ∎

Proof of Theorem 5. We will prove the theorem for $\gamma = 7$ and for w. The general case is similar but notationally very involved. From the previous proof we know that $\left(-3\frac{1}{3}, \frac{2}{3}\right)$ is reached for strategies such that $\rho = \frac{1}{3}$ and $\varphi = 1$, $\left(\frac{2}{3}, -3\frac{1}{3}\right)$ is reached for strategies such that $\rho = 1$ and $\varphi = \frac{1}{3}$. Our proof is based on the

construction of a pair of pure strategies (π, σ) under which for certain stretches of play the average payoffs converge to a neighborhood of $\left(-3\frac{1}{3}, \frac{2}{3}\right)$, alternated by stretches of play where the average payoffs converge to $\left(\frac{2}{3}, -3\frac{1}{3}\right)$.

First, we introduce a couple of convenient notations, where the notation involving π always refers to Aaron's strategy and σ always to Batsheva's. Let $\pi^1 = \sigma^1 = (1, 1, 1, 1, 1, 1, \ldots)$, and $\pi^{\frac{1}{3}} = \sigma^{\frac{1}{3}} = (1, 2, 2, 1, 2, 2, \ldots)$. Furthermore, let $\rho_t = \rho_t^1 + \rho_t^2$, and let $\mu^A(t)$, $\mu^B(t)$ denote the average stage payoffs for Aaron respectively Batsheva during the first t stages. Take $\varepsilon \in \left(0, \frac{1}{2}\right)$. Now, we construct (π, σ) as follows.

$$
(\pi, \sigma) = \begin{cases}
\left(\pi^{\frac{1}{3}}, \sigma^1\right) & \text{till} & \left|\mu^A(t) + 3\frac{1}{3}\right| \leq \varepsilon & \text{followed by} \\
\left(\pi^1, \sigma^{\frac{1}{3}}\right) & \text{till} & \left|\mu^B(t) + 3\frac{1}{3}\right| \leq \frac{\varepsilon}{2} & \text{followed by} \\
\left(\pi^{\frac{1}{3}}, \sigma^1\right) & \text{till} & \left|\mu^A(t) + 3\frac{1}{3}\right| \leq \frac{\varepsilon}{3} & \text{followed by} \\
\left(\pi^1, \sigma^{\frac{1}{3}}\right) & \text{till} & \left|\mu^B(t) + 3\frac{1}{3}\right| \leq \frac{\varepsilon}{4} & \text{followed by} \\
\vdots & \vdots & \vdots & \vdots
\end{cases}
$$

Given this pair of strategies, define sequence $t_1 < t_2 < t_3 < t_4 < \ldots$ where t_k for odd (even) k denotes the first stage for which $\mu^A(t) \in \left[-3\frac{1}{3} - \frac{\varepsilon}{k}, -3\frac{1}{3} + \frac{\varepsilon}{k}\right]$ $\left(\mu^B(t) \in \left[-3\frac{1}{3} - \frac{\varepsilon}{k}, -3\frac{1}{3} + \frac{\varepsilon}{k}\right]\right)$. So, $\lim_{k \to \infty}\left(\mu^A(t_{2k}), \mu^B(t_{2k})\right) = \left(\frac{2}{3}, -\frac{10}{3}\right)$, $\lim_{k \to \infty}\left(\mu^A(t_{2k+1}), \mu^B(t_{2k+1})\right) = \left(-\frac{10}{3}, \frac{2}{3}\right)$. Hence, $\gamma(\pi, \sigma) \leq \left(-3\frac{1}{3}, -3\frac{1}{3}\right)$.

Take $n \gg 0$ be odd, let $t^* = t_n$ and $t^{**} = t_{n+1}$, and let integer $k \in [1, t^{**} - t^* - 1]$. We will now show that Aaron's average rewards do not decrease.

$$
\begin{aligned}
\mu^A(t^* + k + 1) - \mu^A(t^* + k) &= \frac{-\mu^A(t^* + k) + \left(6 - 4\rho_{t^*+k+1}\right)}{t^* + k + 1} \geq \\
\frac{6 - \mu^A(t^* + k) - 4 \cdot \frac{4}{3}}{t^* + k + 1} &= \frac{\frac{2}{3} - \mu^A(t^* + k)}{t^* + k + 1}.
\end{aligned}
$$

This implies that $\mu^A(t^* + k) \geq -3\frac{1}{3}$.

Furthermore, the 'expected change' in average payoffs for Batsheva is

$$
\begin{aligned}
&E\left\{\mu^B(t^* + k + 1) - \mu^B(t^* + k)\right\} \\
&= \frac{\left(-\mu^B(t^* + k) + \frac{1}{3}\left(6 - 4\rho_{t^*+k+1}\right) + \frac{2}{3}\left(4 - 7\rho_{t^*+k+1}\right)\right)}{t^* + k + 1} \\
&= \frac{\left(\frac{14}{3} - \mu^B(t^* + k) - 6\rho_{t^*+k+1}\right)}{t^* + k + 1} = \frac{\left(-\frac{10}{3} - \mu^B(t^* + k) + 6\nu_k\right)}{t^* + k + 1},
\end{aligned}
$$

where $\nu_k = \frac{4}{3} - \rho_{t^*+k+1}$. It can be easily verified that $\nu_k > 0$ and that ν_k is strictly decreasing in $k \in [k^*, t^{**} - t^* - 1]$. So, $E\left\{\mu^B(t^* + k + 1) - \mu^B(t^* + k)\right\} < 0$ if and only if $\mu^B(t^* + k) > -\frac{10}{3} + 6\nu_k$. This implies that $\left\{\mu^B(t)\right\}_{t = t^*}^{t^{**}}$ is decreasing but bounded away from $-\frac{10}{3}$. Therefore, $\mu^B(t^* + k) \geq -3\frac{1}{3}$.

The case $n \gg 0$ and n is even, is similar because of the symmetry. So, $\liminf_{t \to \infty} \mu^A(t) = \liminf_{t \to \infty} \mu^B(t) = -3\frac{1}{3}$. Hence, $\gamma(\pi, \sigma) = \left(-3\frac{1}{3}, -3\frac{1}{3}\right)$. ∎

Proof of Lemma 6. We only prove the case $v^A = \max\{-2, 4 - \gamma\}$. If Aaron always plays his first action, then his stage payoffs converge to $\left(2 - 4\rho_t^B\right)$ in the long run and his rewards are at least -2; if he always plays the other action his stage payoffs converge to $\left(4 - \gamma\rho_t^B\right)$ in the long run and his rewards are at least $(4 - \gamma)$. Now, Batsheva can keep Aaron's maximal reward at $\max\{-2, 4 - \gamma\}$ by playing action 1 forever. In that case, his long run stage payoffs are $\left(2 - 4\rho_t^A\right)$ for action 1, and $\left(4 - \gamma - \gamma\rho_t^A\right)$ for action 2. Hence, Aaron's long run average payoffs converge to $\rho_t^A(2 - 4\rho_t^A) + (1 - \rho_t^A)(4 - \gamma - \gamma\rho_t^A) = (\gamma - 4)\left(\rho_t^A\right)^2 - 2\rho_t^A + 4 - \gamma$. It may be readily confirmed that on $[0, 1]$ this function has a maximum of $(4 - \gamma)$ for $\rho_t^A = 0$ if $4 \leq \gamma \leq 6$, and -2 for $\rho_t^A = 1$ if $\gamma \geq 6$. So, Aaron's rewards are at most $\max\{-2, 4 - \gamma\}$. Therefore, $v^A = \max\{-2, 4 - \gamma\}$. ∎

Proof of Theorem 10. Let (x, y) be a pair of rewards in the interior of E'. Then, numbers $\underline{x} < x \leq \overline{x}$ and $\underline{y} < y \leq \overline{y}$ exist such $\left(\underline{x}, \overline{y}\right), \left(\overline{x}, \underline{y}\right) \in E'$. So, equilibrium strategies $(\pi^x, \sigma^y), \left(\pi^{\underline{x}}, \sigma^{\overline{y}}\right), \left(\pi^{\overline{x}}, \sigma^{\underline{y}}\right)$ exist such that $\gamma(\pi^x, \sigma^y) = (x, y), \gamma\left(\pi^{\underline{x}}, \sigma^{\overline{y}}\right) = \left(\underline{x}, \overline{y}\right), \gamma\left(\pi^{\overline{x}}, \sigma^{\underline{y}}\right) = \left(\overline{x}, \underline{y}\right)$. Let $T^* \geq 0$ denote the length of the communication period of strategy pair (π^x, σ^y). For $T > 1$, let $AD_{\pi^x}^A(T) = \frac{\#\{j_t^A \neq \pi_t^x | t \leq T\}}{T}$, $AD_{\sigma^y}^B(T) = \frac{\#\{j_t^B \neq \sigma_t^y | t \leq T\}}{T}$. Define (π, σ) as follows:

$$\pi_t = \pi_{t-T'}^{\underline{x}} \quad \sigma_t = \sigma_{t-T'}^{\overline{y}} \qquad \begin{array}{l} \text{if for some } T' > T^*\text{: } AD_{\pi^x}^A(T') > 1/\sqrt{T'} \\ \text{and } AD_{\pi^x}^A(T') \geq AD_{\sigma^y}^B(T') \end{array}$$

$$\pi_t = \pi_t^{\overline{x}} \qquad \sigma_t = \sigma_t^{\underline{y}} \qquad \begin{array}{l} \text{if for some } T'' > T^*\text{: } AD_{\sigma^y}^B(T'') > 1/\sqrt{T''} \\ \text{and } AD_{\sigma^y}^B(T'') > AD_{\pi^x}^A(T'') \end{array}$$

$$\pi_t = \pi_t^x \qquad \sigma_t = \sigma_t^y \qquad \text{otherwise.}$$

Then, $\gamma(\pi, \sigma) = \gamma(\pi^x, \sigma^y) = (x, y)$, because in the long run deviations from the equilibrium path of (π^x, σ^y) go to zero in relative frequency. Now, (π, σ) is an equilibrium because if, e.g., Batsheva deviates more than Aaron does such that for some $T'' : AD_{\sigma^y}^B(T'') > 1/\sqrt{T''}$, then the play continues according to equilibrium $\left(\pi^{\overline{x}}, \sigma^{\underline{y}}\right)$. Batsheva gets $\underline{y} < y$ in that case, while Aaron receives $\overline{x} \geq x$. A similar statement holds for Aaron as well. Hence, neither player can improve his rewards by deviating unilaterally as described. ∎

7 References

Billingsley P, 1986, "Probability and Measure", John Wiley & Sons, New York, NY.

Brenner T, & U Witt, 2003, Melioration learning in games with constant and frequency-dependent payoffs, *J Econ Behav Organ* **50**, 429-448.

Dawes RM, 1980, Social dilemmas, *Ann Rev Psychology* **21**, 169-193.

Forges F, 1986, An approach to communication equilibria, *Econometrica* **54**, 1375-1385.

Gintis H, 2001, "Game Theory Evolving", Princeton University Press, Princeton, NJ.

Hart S, 1985, Nonzero-sum two-person repeated games with incomplete information, *Math Oper Res* **10**, 117-153.

Herrnstein RJ, 1997, "The Matching Law: Papers in Psychology and Economics", Harvard University Press, Cambridge, MA.

Joosten R, 1996, "Dynamics, Equilibria, and Values", Ph.D.-thesis, Maastricht University.

Joosten R, 2001, A note on repeated games with vanishing actions, forthcoming in *Int Game Theory Rev.*

Joosten R, T Brenner & U Witt, 2003, Games with frequency-dependent stage payoffs, *Int J Game Theory* **31**, 609-620.

Joosten R, G Schoenmakers, H Peters & F Thuijsman, 2004, Equilibria in repeated games with vanishing actions, working paper.

Osborne MJ, & A Rubinstein, 1994, "A Course in Game Theory", MIT Press, Cambridge, MA.

Platt J, 1973, Social traps, *American Psychologist* **28**, 641-651.

Schoenmakers G, J Flesch & F Thuijsman, 2002, Coordination games with vanishing actions, *Int Game Theory Rev* **4**, 119-126.

Shapley L, 1953, Stochastic games, *Proceedings of the National Academy of Sciences U.S.A.* **39**, 1095-1100.

Van Damme EEC, 1991, Stability and Perfection of Nash Equilibria, Springer, Berlin.