Chapter 12
Stochastic Games with Endogenous Transitions

Reinoud Joosten and Robin Meijboom

12.1 Introduction

We present and subsequently analyze a stochastic game in which transition probabilities at any point in time depend on the history of the play, i.e., players' past action choices, their current choices, and the current state. This development has been inspired by an ambition to incorporate certain empirical phenomena into *Small Fish Wars*[1] [37]. Here, agents possess the fishing rights on a body of water, and the resource can be in either of two states, *High* or *Low*. In the former, the fish are more abundant and therefore catches are larger than in the latter. The agents have two options, to fish with or without restraint. Fishing with restraint by both agents is (assumed to be) sustainable in the long run, as the resource will be (assumed to be) able to recover; unrestrained fishing by both yields higher immediate catches, but damages the resource significantly if continued for prolonged periods of time. This damage becomes apparent in the dynamics of the system as an increase in the probabilities that the system moves from *High* to *Low*, and simultaneously a decrease in the probabilities of the system to move from *Low* to *High*. This causes the system and hence the play, to spend a higher proportion of time in *Low*.

We additionally aim to incorporate hysteresis effects called poaching pits in the field of management of replenishable resources (e.g., Bulte [11], Courchamp et al. [13], Hall et al. [24]). Hysteresis may be caused by biological phenomena induced by the (nature of the) exploitation of the resource. For instance, full-grown cod

---

[1]A word play on Levhari and Mirman [50] who show that strategic interaction in a fishery may induce a "tragedy of the commons" [27].

R. Joosten (✉) · R. Meijboom
IEBIS, BMS, University of Twente, POB 217,
7500 AE Enschede, The Netherlands
e-mail: r.a.m.g.joosten@utwente.nl

spawn a considerably higher number of eggs than younger specimen: Oosthuizen and Daan [57], Armstrong et al. [3] find linear fecundity-weight relations, Rose et al. [63] report exponential fecundity-weight relations. As mature cod are targeted by modern catching techniques such as for instance gill netting, overfishing hurts mainly the cohorts most productive in providing offspring. To regain full reproductive capacity, younger cohorts must reach ages well beyond adulthood. Hence, it may take cod a long while to escape a poaching pit after a recovery plan or program to replenish the stock has been effectuated.

To achieve our goals we engineered a stochastic game[2] as follows. Nature (chance) may move the play from one state to the other dependent on the current action choices of the agents, but also on their past catching behavior. To achieve the above-formulated modeling aims we introduce *endogenously changing* stochastic variation,[3] the evolution of the transition probabilities reflects that the more frequently the agents exploit the resource without restraint, the more it deteriorates. Here, the probability of moving to *High* may decrease in time in each state and for each action combination if the agents show prolonged lack of restraint, i.e., overfish frequently.

Transition probabilities from *Low* to *High* may become zero, resulting in *Low* becoming a *temporarily* absorbing state. If the agents keep overexploiting the resource, this situation does not change in our model. Even if the agents revert to restraint in order to bring about the recovery of the resource, it may take a long time before *High* becomes accessible again. Thus, we endeavor to reproduce effects similar to the ones associated to hysteresis.

The agents are assumed to wish to maximize their long-term average catches. We adopt a Folk Theorem type analysis as in Joosten et al. [42], and validate relevant procedures in this new setting. First, we show how to establish the rewards for any pair of jointly convergent pure strategies. Then, we determine the set of jointly convergent pure-strategy rewards. A more complex issue is then to find for each player the threat point reward, i.e., the highest amount this player can guarantee himself if his opponent tries to minimize his rewards. Finally, we obtain a large set of rewards which can be supported by equilibria using threats, namely all jointly-convergent pure-strategy rewards giving each player more than the threat point reward.

In the model analyzed throughout the chapter for expository purposes, we gain insights relevant to the management of the resource. Our findings reveal a potential for compromise between ecological and economic maximalistic goals, thus overcoming the one-sidedness of management policies for natural resources as noted by e.g., Holden [33], Brooks et al. [10], and in turn improving their chances of success cf., e.g., BenDor et al. [5], Sanchirico et al. [64]. Full restraint, an ecological maximalistic goal, yields total rewards which are considerably higher than never-restraint rewards. Yet, a possible economic maximalistic goal, i.e., Pareto-efficient equilibrium rewards resulting from jointly convergent pure strategies with threats, yields a

---

[2] 'Engineered' as in Aumann [4]. Stochastic games were introduced by Shapley [69], see also Amir [1] for links to difference and differential games to which much work on fisheries belongs, cf., e.g., Haurie et al. [29], Long [51] for overviews.

[3] So, the Markov property of standard stochastic games [69] is lost.

sizeable increase of total rewards even over full restraint. We find that the proportion of time spent in such a poaching pit goes to zero in the long run under equilibrium behavior. A whole range of models should be analyzed to obtain general findings providing insights into the full range of fishery management games.

Next, we introduce our model with endogenous transition probabilities. In Sect. 12.3, we focus on strategies and restrictions desirable or resulting from the model. Section 12.4 treats rewards in a very general sense, and equilibrium rewards more specifically. Also some attention is paid to the complexity of computing threat point rewards. Section 12.5 concludes.

12.2   Endogenous Transition Probabilities

A *Small Fish War* is played by row player A and column player B at discrete moments in time called stages. Each player has two actions and at each stage $t \in \mathbb{N}$ the players independently and simultaneously choose an action. Action 1 for either player denotes the action for which some restriction exists allowing the resource to recover, e.g., catching with wide-mazed nets or catching a low quantity. Action 2 denotes the action with little restraint.

We assume catches to vary due to random shocks, which we model by means of a stochastic game with two states at every stage of the play. First, let us capture the past play until stage t, $t > 1$, by the following two matrices:

$$QH^t = \begin{bmatrix} q_1^t & q_2^t \\ q_3^t & q_4^t \end{bmatrix}, \text{ and } QL^t = \begin{bmatrix} q_5^t & q_6^t \\ q_7^t & q_8^t \end{bmatrix}.$$

Here, e.g., q_1^t is the relative frequency with which action pair top-left in *High* has occurred until stage t, and q_7^t is the relative frequency of action pair bottom-left in *Low* having occurred during past play. So, we must have $q^t = \left(q_1^t, \ldots, q_8^t\right) \in \Delta^7 = \{x \in \mathbb{R}^8 | x_i \geq 0 \text{ for all } i = 1, \ldots, 8 \text{ and } \sum_{j=1}^8 x_j = 1\}$. We refer to such a vector as the **relative frequency vector**.

Let the interaction at stage t of the play be represented by the following:

$$H^t = H\left(q^t\right) = \begin{bmatrix} \theta_1, p_1\left(q^t\right) & \theta_2, p_2\left(q^t\right) \\ \theta_3, p_3\left(q^t\right) & \theta_4, p_4\left(q^t\right) \end{bmatrix},$$

$$L^t = L\left(q^t\right) = \begin{bmatrix} \theta_5, p_5\left(q^t\right) & \theta_6, p_6\left(q^t\right) \\ \theta_7, p_7\left(q^t\right) & \theta_8, p_8\left(q^t\right) \end{bmatrix}.$$

Here $H^t\left(q^t\right)$ $\left(L^t\left(q^t\right)\right)$ indicates state *High* (*Low*) at stage t of the play if the play until then resulted in relative frequency vector q^t. Each entry of the two matrices has an ordered pair denoting the pair of payoffs to the players $\theta_i = \left(\theta_i^A, \theta_i^B\right)$ if the corresponding action pair is chosen and the probability $p_i\left(q^t\right)$ that the system moves

to *High* at stage $t + 1$ (and to *Low* with the complementary probability). All functions $p_i : \Delta^7 \to [0, 1]$ are assumed continuous. We now give an example.

Example 1 In this Small Fish War we assume that in both states Action 1, i.e., catching with restraint, is dominated by the alternative.[4] Let, for given relative frequency vector $q^t \in \Delta^7$, the transition functions $p_i : \Delta^7 \to [0, 1]$, $i = 1, \ldots, 8$, governing the transition probabilities, be given by

$$p_1(q^t) = \left[\tfrac{8}{10} - \tfrac{11}{24}q_4^t - \tfrac{11}{12}q_8^t\right]_+$$
$$p_2(q^t) = p_3(q^t) = \left[\tfrac{6}{10} - \tfrac{11}{20}q_4^t - \tfrac{11}{10}q_8^t\right]_+$$
$$p_4(q^t) = \left[\tfrac{3}{10} - \tfrac{11}{16}q_4^t - \tfrac{11}{8}q_8^t\right]_+$$
$$p_5(q^t) = \left[\tfrac{6}{10} - \tfrac{11}{12}q_4^t - \tfrac{11}{6}q_8^t\right]_+$$
$$p_6(q^t) = p_7(q^t) = \left[\tfrac{4}{10} - \tfrac{11}{8}q_4^t - \tfrac{11}{4}q_8^t\right]_+$$
$$p_8(q^t) = \left[\tfrac{1}{10} - \tfrac{11}{4}q_4^t - \tfrac{11}{2}q_8^t\right]_+ .$$

Here, $[x]_+$ is short hand for $\max\{x, 0\}$. These equations capture the following deliberations. Two-sided full restraint is assumed to cause not more damage to the resource in both states than if exactly one player catches with restraint. Hence, the probability that during the next stage play is in *High* if the first case arises is at least equal to the corresponding probability in the second case. We also assume symmetry, hence $p_2(q^t) = p_3(q^t)$ and $p_6(q^t) = p_7(q^t)$. Furthermore, we assume that exactly one player catching without restraint is not more harmful to the resource than two players catching without restraint. The inequalities $p_i(q^t) \geq p_{i+4}(q^t)$ for $i = 1, \ldots, 4$, are assumed to hold because if the play is in *Low*, the system is assumed at least as more vulnerable to overfishing as in *High*. We refer to e.g., Kelly et al. [46] for an empirical underpinning of these modeling choices.

Now, we show that renewable resources may recuperate slowly after a program of recovery has been taken up. Suppose both agents play Action 1 twice followed by 2 for a sufficiently long period of time until stage t^*. Clearly, $q_4^{t^*} + q_8^{t^*} = \tfrac{t^*-2}{t^*}$. Now, for $t^* \to \infty$, $p_5(q^{t^*}) = p_6(q^{t^*}) = p_7(q^{t^*}) = p_8(q^{t^*}) = 0$, because

$$\tfrac{6}{10} - \tfrac{11}{12}q_4^{t^*} - \tfrac{11}{6}q_8^{t^*} = \tfrac{6}{10} - \tfrac{11}{12}\left(\tfrac{t^*-2}{t^*} - q_8^{t^*}\right) - \tfrac{11}{6}q_8^{t^*} =$$
$$\tfrac{6}{10} - \tfrac{11}{12}\left(1 - \tfrac{2}{t^*} - q_8^{t^*}\right) - \tfrac{11}{6}q_8^{t^*} = -\tfrac{19}{60} + \tfrac{11}{6t^*} - \tfrac{11}{12}q_8^{t^*} < 0.$$

Then, $p_5(q^{t^*}) = 0$ and by the relation to the other transition probability functions, $p_6(q^{t^*}) = p_7(q^{t^*}) = p_8(q^{t^*}) = 0$ as well. Take $t^* = 16$, clearly

$$-\tfrac{19}{60} + \tfrac{11}{6t^*} - \tfrac{11}{12}q_8^{t^*} < -\tfrac{19}{60} + \tfrac{11}{6t^*} < 0.$$

If both agents switch to playing sequences of $(1, 1, 1, \ldots)$ from then on, it will take a while before $p_5(q^t)$ becomes positive again. Since

---

[4]Right now, we do not need the actual payoffs and focus on the transition probabilities.

$$\frac{6}{10} - \frac{11}{12}q_4^{t^*+k} - \frac{11}{6}q_8^{t^*+k} = \frac{6}{10} - \frac{11}{12}\frac{t^*-2}{t^*+k} - \frac{11}{12}q_8^{t^*+k} =$$
$$\frac{6}{10} - \frac{11}{12}\left(1 - \frac{k+2}{t^*+k}\right) - \frac{11}{12}q_8^{t^*+k} = -\frac{19}{60} + \frac{11}{12}\frac{k+2}{t^*+k} - \frac{11}{12}q_8^{t^*+k} =$$
$$< -\frac{19}{60} + \frac{11}{12}\frac{k+2}{t^*+k},$$

the first expression cannot be positive for $k < \frac{19t^*-110}{36}$. So, for $t^* = 16$ it takes **at least** six stages for the play to be able to return to *High*.

12.3   Strategies and Restrictions

A strategy is a game plan for the entire infinite time horizon, allowing it to depend on any condition makes an extensive analysis of infinitely repeated games quite impossible. Most restrictions in the literature put requirements on what aspects the strategies are conditional upon. For instance, a *history-dependent* strategy prescribes a possibly mixed action to be played at each stage conditional on the current stage and state, as well as on the full history until then, i.e., all states visited and all action combinations realized before.

Less general strategies are for instance, *action independent* ones which condition on all states visited before, but not on the action combinations chosen [31]. Markov strategies condition on the current state and the current stage, and stationary strategies only condition on the present state (cf., e.g., Filar and Vrieze [20], Flesch [21]).

The challenge in the present framework is to find restrictions on strategies which are helpful in the analysis. Although Markov and stationary strategies have proven their value in the analysis of finite state stochastic games with fixed transition probabilities, it is quite unclear what their contribution can be in the present framework.

Essentially, (at least) two points of view can be adopted to analyze the present framework. The one we favor is the one in which *High* and *Low* are seen as the states with the transitions between these states being a function of the history of the play as captured by the relative frequency vector q^t. Stationary strategies are easily formulated here, but probably much too simple for analytical purposes as some link with q^t must be assumed to be useful. An alternative is to define the states according to the relative frequency vector in which there exist infinitely many states $H(q^t)$ and $L(q^t)$. Here, the practical problem is the enormity of the task of infinitely many stationary or Markov strategies to be defined.

Let $\mathcal{X}^k$ denote the set of history-dependent strategies of player $k = 1, 2$. A strategy is **pure**, if at *each* stage a **pure action** is chosen, i.e., an action is chosen with probability 1. The set of pure strategies for player k is $\mathcal{P}^k$, and $\mathcal{P} \equiv \mathcal{P}^A \times \mathcal{P}^B$. Let us define the following notions, introduced before in a rather informal manner, a bit more formally. For $j = 1, 2, t > 1$

$$q_j^t \equiv \frac{\#\left\{\left(j_u^{A,H}, j_u^{B,H}\right) \mid j_u^{A,H}=1, \ j_u^{B,H}=j, \ 1 \leq u < t\right\}}{t-1},$$

$$q_{j+2}^t \equiv \frac{\#\left\{\left(j_u^{A,H}, j_u^{B,H}\right) \mid j_u^{A,H}=2, \ j_u^{B,H}=j, \ 1 \leq u < t\right\}}{t-1},$$

$$q_{j+4}^t \equiv \frac{\#\left\{\left(j_u^{A,L}, j_u^{B,L}\right) \mid j_u^{A,H}=1, \ j_u^{B,H}=j, \ 1 \leq u < t\right\}}{t-1},$$

$$q_{j+6}^t \equiv \frac{\#\left\{\left(j_u^{A,L}, j_u^{B,L}\right) \mid j_u^{A,H}=2, \ j_u^{B,H}=j, \ 1 \leq u < t\right\}}{t-1}.$$

Here, $j_u^{A,X}$ ($j_u^{B,X}$) denotes the action taken by player A (B) while being in state $X = H, L$ at stage u. So, for instance q_4^t is the relative frequency of action pair $(2, 2)$ in state H being chosen until stage t.

The strategy pair $(\pi, \sigma) \in \mathcal{X}^A \times \mathcal{X}^B$ is **jointly convergent** if and only if $q \in \Delta^7$ exists such that for all $\varepsilon > 0$, $i \in \{1, 2, \ldots, 8\}$:

$$\limsup_{t\to\infty} \Pr_{\pi,\sigma}\left[\left|q_i^t - q_i\right| \geq \varepsilon\right] = 0. \tag{12.1}$$

$\Pr_{\pi,\sigma}$ denotes the probability under strategy pair (π, σ). $\mathcal{JC}$ denotes the set of jointly convergent strategy pairs. Under such a pair of strategies, the relative frequency of each action pair in both states as play goes to infinity converges to a fixed number with probability 1 in the terminology of Billingsley [8, p. 274]). The **set of jointly-convergent pure-strategy rewards** $P^{\mathcal{JC}}$ is then the set of pairs of rewards obtained by using a pair of jointly-convergent pure strategies.

For a pair of jointly convergent pure strategies, let $p_i \equiv \lim_{t\to\infty} p_i\left(q^t\right) = p_i(q)$ for $i = 1, \ldots, 8$. These notions are well defined as the relevant functions are continuous (cf., e.g., Billingsley [8]). We distinguish the following restrictions to be explained below:

$$0 < \sum_{i=1}^4 q_i\,(1 - p_i) = \sum_{i=5}^8 q_i\,p_i \text{ and } 0 < \sum_{i=1}^4 q_i < 1, \tag{12.2}$$

$$\sum_{i=5}^8 q_i = 1, \text{ and } q_i > 0 \Longrightarrow p_i = 0, \ i = 5, \ldots, 8, \tag{12.3}$$

$$\sum_{i=1}^4 q_i = 1, \text{ and } q_i > 0 \Longrightarrow p_i = 1, \ i = 1, \ldots, 4. \tag{12.4}$$

Restriction (12.2) is a conservation of flow equation: play takes place on both states infinitely often, therefore, due to the law of large numbers the actual instances of leaving *High* must be proportional to the long run probability of leaving it and the latter must be equal to the probability of returning.

If the long run play occurs in *Low* exclusively, (12.3) must hold. The former part is obvious, if $q_i p_i > 0$ for some $i = 5, \ldots, 8$, then play would visit the corresponding entry infinitely often as time goes to infinity, hence with probability at least $q_i p_i$ state *High* would occur. Similar reasoning applies to the other case that play occurs only in *High*, hence (12.4). We now show the implications for jointly-convergent pure strategies.

Example 2 Now, (12.3) can only hold if $p_i = 0$ or $q_i = 0$ for all $i = 5, \ldots, 8$. Similarly, (12.4) can only hold if $1 - p_i = 0$ or $q_i = 0$ for all $i = 1, \ldots, 4$. So, *if a state is absorbing, then positive mass on a component of the relative frequency vector q can only occur if the associated probability of leaving that state is zero*. Observe
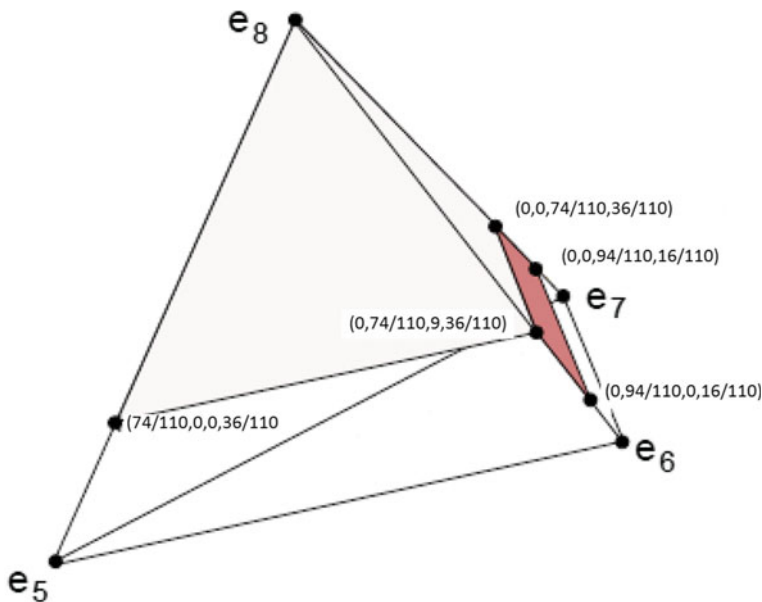
Fig. 12.1 If play concentrates on *Low*, $q_1 = \cdots = q_4 = 0$ and $q_5 + \cdots + q_8 = 1$. We depict this face of Δ^7 as a "projection" unto Δ^3. Extreme point e_i has component $i - 4$ equal to one. The admissible q's, are sketched as the three-dimensional set on top, and the two-dimensional boundary set

that therefore only *Low* can be absorbing. From the ranking of probabilities, we may distinguish the following three subcases.

$$q_8 = 1 \text{ and } p_8 = \tfrac{1}{10} - \tfrac{11}{4}q_4 - \tfrac{11}{2}q_8 \le 0 \text{ or}$$
$$\textstyle\sum_{i=6}^{8} q_i = 1 \text{ and } p_6 = p_7 = \tfrac{4}{10} - \tfrac{11}{8}q_4 - \tfrac{11}{4}q_8 \le 0 \text{ or}$$
$$\textstyle\sum_{i=5}^{8} q_i = 1 \text{ and } p_5 = \tfrac{6}{10} - \tfrac{11}{12}q_4 - \tfrac{11}{6}q_8 \le 0.$$

Clearly, $q_4 = 0$. The first case is easily checked reducing analysis to

$$\textstyle\sum_{i=6}^{8} q_i = 1 \text{ and } \tfrac{16}{110} \le q_8 \le \tfrac{36}{110}, \text{ or}$$
$$\textstyle\sum_{i=5}^{8} q_i = 1 \text{ and } q_8 \ge \tfrac{36}{110}.$$

leading to $q_5 = 0$ and $\tfrac{16}{110} \le q_8 \le \tfrac{36}{110}$, and $q_8 \ge \tfrac{36}{110}$ and $q_5, \ldots, q_7 \ge 0$.

Figure 12.1 visualizes these restrictions for *Low* being absorbing. The upper three-dimensional subset of Δ^3, is connected to the final inequality; the parallelogram on the face of Δ^3 is connected to the former.

12.4 On Rewards and Equilibrium Rewards

The players receive an infinite stream of stage payoffs, they are assumed to wish to maximize their average rewards. For a given pair of strategies (π, σ), $R_t^k(\pi, \sigma)$ is the expected payoff to player k at stage t under strategy combination (π, σ), then player k's **average reward,** $k = A, B$, is $\gamma^k(\pi, \sigma) = \liminf_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} R_t^k(\pi, \sigma)$, and $\gamma(\pi, \sigma) \equiv (\gamma^A(\pi, \sigma), \gamma^B(\pi, \sigma))$. Moreover, for vector $q \in \Delta^7$, the q**-averaged payoffs** $(x, y)_q$ are given by

$$(x, y)_q = \sum_{i=1}^{8} q_i \theta_i.$$

The strategy pair $(\pi^*, \sigma^*) \in \mathcal{X}^A \times \mathcal{X}^B$ is **an equilibrium** if and only if

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

The rewards $\gamma(\pi^*, \sigma^*)$ associated with an equilibrium (π^*, σ^*) will be referred to as equilibrium rewards.

In the analysis of repeated games, another helpful measure to reduce complexity is to focus on rewards instead of strategies. It is more rule than exception that one and the same reward combination can be achieved by several distinct strategy combinations. Here, we focus on rewards to be obtained by jointly-convergent pure strategies.

12.4.1 Jointly Convergent Pure-Strategy Rewards

The next result connects notions introduced in the previous sections.

Proposition 1 *Let strategy pair* $(\pi, \sigma) \in \mathcal{JC}$ *and let* $q \in \Delta^7$ *for which* (12.1) *is satisfied, then the average payoffs are given by* $\gamma(\pi, \sigma) = (x, y)_q$.

Proof Let $(\pi, \sigma) \in \mathcal{JC}$ and $E\{\theta_u^{\pi, \sigma}\} \equiv (R_u^1(\pi, \sigma), R_u^2(\pi, \sigma))$, then

$$\lim_{t \to \infty} \frac{1}{t} \sum_{u=1}^{t} E\{\theta_u^{\pi, \sigma}\} = \lim_{t \to \infty} E\left\{\frac{1}{t} \sum_{u=1}^{t} \theta_u^{\pi, \sigma}\right\} =$$
$$\lim_{t \to \infty} E\left\{\sum_{i=1}^{8} q_i^t \theta_i\right\} = \lim_{t \to \infty} \sum_{i=1}^{8} E\{q_i^t\} \theta_i = \sum_{i=1}^{8} q_i \theta_i = (x, y)_q.$$

The second equality sign involves a change in counting: on the left-hand side we sum over all periods, on the right-hand side over all eight entries of the two bi-matrices weighed by their relative frequencies. Equalities one and three are standard, the penultimate one follows from (12.1), cf., e.g., Billingsley [8, p. 274], the final one by the definition given above. Since $\lim_{t \to \infty} \frac{1}{t} \sum_{u=1}^{t} E\{\theta_u^{\pi, \sigma}\}$ equals $(x, y)_q$, it follows that $\gamma(\pi, \sigma) = (x, y)_q$.

Example 3 To continue the example, we add stage payoffs

$$H\left(q^t\right) = \begin{bmatrix} (4,4), p_1\left(q^t\right) & \left(\frac{7}{2},6\right), p_2\left(q^t\right) \\ \left(6,\frac{7}{2}\right), p_3\left(q^t\right) & \left(\frac{11}{2},\frac{11}{2}\right), p_4\left(q^t\right) \end{bmatrix},$$

$$L\left(q^t\right) = \begin{bmatrix} (2,2), p_5\left(q^t\right) & \left(\frac{7}{4},3\right), p_6\left(q^t\right) \\ \left(3,\frac{7}{4}\right), p_7\left(q^t\right) & \left(\frac{11}{4},\frac{11}{4}\right), p_8\left(q^t\right) \end{bmatrix}.$$

Observe that $\theta_i = \frac{1}{2}\theta_{i-4}$ for $i = 5, \ldots, 8$. The specifics for the probabilities $p_1\left(q^t\right), \ldots, p_8\left(q^t\right)$ were already presented earlier. Note that in both states, the first action is dominated by the second for both players.

Figure 12.2 shows the rewards consistent with *Low* being absorbing and note that this hexagon is not convex.[5] The link between rewards in Fig. 12.2 and the strategy restrictions visualized in Fig. 12.1 is that the extreme points in Fig. 12.2 have the following coordinates (i.e., rewards)

$$\frac{74}{110}(2,2) + \frac{36}{110}\left(\frac{11}{4},\frac{11}{4}\right) = \left(\frac{247}{110},\frac{247}{110}\right)$$
$$\frac{74}{110}\left(3,\frac{7}{4}\right) + \frac{36}{110}\left(\frac{11}{4},\frac{11}{4}\right) = \left(\frac{642}{220},\frac{457}{220}\right)$$
$$\frac{74}{110}\left(\frac{7}{4},3\right) + \frac{36}{110}\left(\frac{11}{4},\frac{11}{4}\right) = \left(\frac{457}{220},\frac{642}{220}\right)$$
$$\frac{94}{110}\left(3,\frac{7}{4}\right) + \frac{16}{110}\left(\frac{11}{4},\frac{11}{4}\right) = \left(\frac{652}{220},\frac{417}{220}\right)$$
$$\frac{94}{110}\left(\frac{7}{4},3\right) + \frac{16}{110}\left(\frac{11}{4},\frac{11}{4}\right) = \left(\frac{417}{220},\frac{652}{220}\right).$$

The first three rewards coincide with the lower three vertices of the shaded simplex of dimension 3 within Δ^3 in Fig. 12.1. The latter two coincide with the lower two vertices of the quadrangle on the face of Δ^3 in Fig. 12.1. Finally, the reward $\left(\frac{11}{4},\frac{11}{4}\right)$ coincides with the vertex e_8 in Fig. 12.1.

So, e_5 corresponds to the situation that in the long run the relative frequency of play on action pair $(1,1)$ in *Low* is 1 (*if that were possible*). The left-hand lowest vertex of the shaded simplex in Fig. 12.1 has coordinates $(74/110, 00, 36/100)$, so the corresponding rewards are obtained by the linear combination of both $(2,2)$ and $\left(\frac{11}{4},\frac{11}{4}\right)$ with the associated weights.

Similarly, all interior points of the shaded simplex in Fig. 12.1 correspond to the interior of the shaded parallelogram in Fig. 12.2. The interior points of the boundary quadrangle in Fig. 12.1 correspond to the interior of the trapezium in Fig. 12.2.

We must also find rewards such that (12.2) is satisfied. Figure 12.3 shows all jointly-convergent pure-strategy rewards. For instance, rewards $\left(\frac{7}{2},\frac{7}{2}\right)$ correspond to mutual full restraint; furthermore, the Pareto-efficient line segment connecting $\left(\frac{22}{6},\frac{23}{6}\right)$ and $\left(\frac{23}{6},\frac{22}{6}\right)$ is achieved by playing Top-Right in *High* and by playing the off-diagonal action pairs in *Low* exclusively.

---

[5]Figures 12.2 and 12.3 are based on Matlab graphs generated by an algorithm yielding 6 million pairs of rewards which took several days. Memory restrictions corrupt image quality as we experienced. The algorithm and output are available on request.
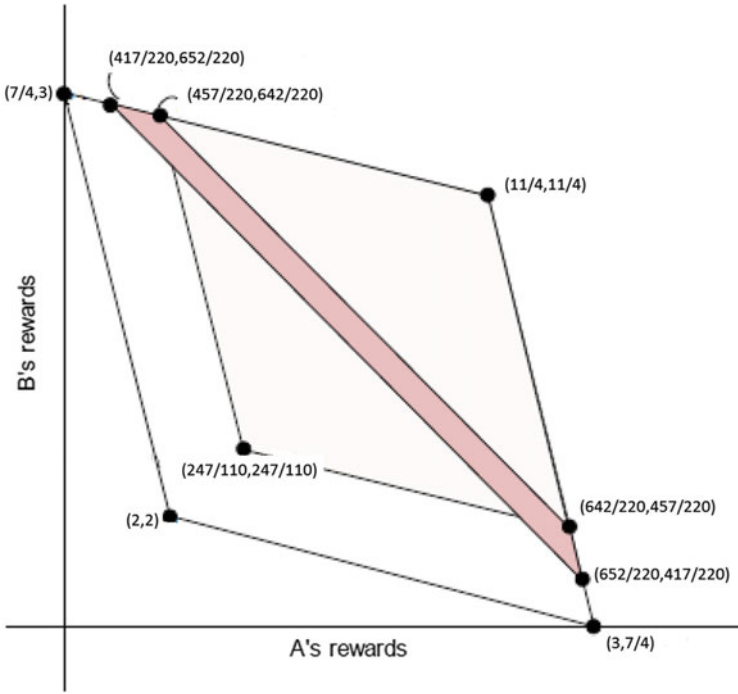
Fig. 12.2 A sketch of the hexagon being the union of the lightly shaded parallelogram and the darker trapezium. The former corresponds to the three-dimensional set, the latter to the two-dimensional boundary set in Fig. 12.1. The other rewards, corresponding to the convex hull of the four entries associated with *Low* are **not feasible by jointly-convergent pure strategies**

12.4.2 Equilibrium Rewards

We now focus on rewards from equilibria involving threats. Our approach is similar to a well-established one in the repeated games literature (cf., e.g., Hart [28], Forges [23]), linked to the Folk Theorem (see e.g., Van Damme [74]) and applied to stochastic games as well (cf., e.g., Thuijsman and Vrieze [71], Joosten et al. [42], Schoenmakers [67]).

We call $v = (v^A, v^B)$ 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 A can get if B tries to minimize A's average payoffs. Under a pair of **individually rational** (feasible) rewards each player receives at least the threat-point reward.

Let $E = \{(x, y) \in P^{\mathcal{JC}} \mid x > v^A \text{ and } y > v^B\}$ be the set of all individually rational jointly convergent pure-strategy rewards giving each player strictly more than his threat point reward. We can now present the following formal result:
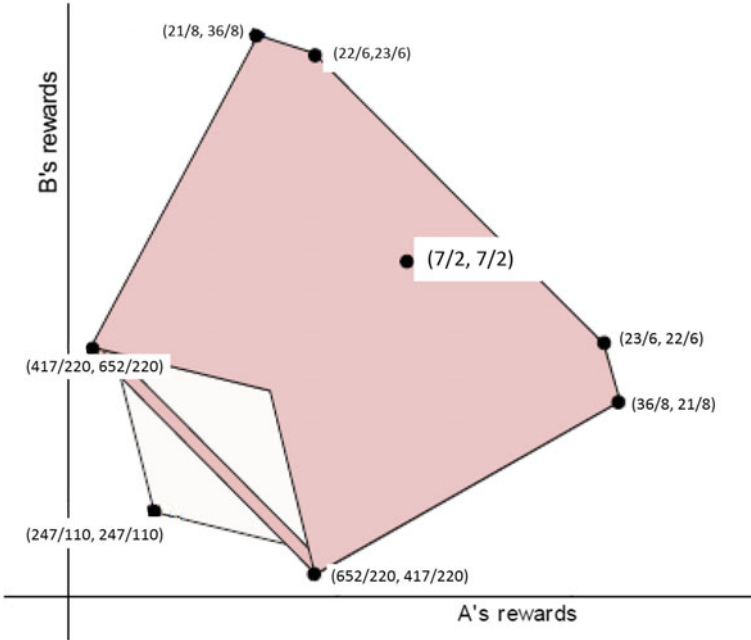
Fig. 12.3 The set $P^{\mathcal{JC}}$: on the lower left-hand side the hexagon of Fig. 12.2, for other rewards both states are visited infinitely often

Theorem 1 *Each pair of rewards in E can be supported by an equilibrium.*

Proof Let $(x, y) \in E$, then a pure-strategy combination $(\pi, \sigma) \in \mathcal{JC}$ exists such that $\gamma(\pi, \sigma) = (x, y)$. Let $\varepsilon = \frac{1}{2} \min(x - v^A, y - v^B)$ and let $\pi^p (\sigma^p)$ be a punishment-strategy of $A (B)$, i.e., a strategy holding his opponent to at most $v^B + \varepsilon$ $(v^A + \varepsilon)$. Let

$$\pi_t^* \equiv \begin{cases} \pi_t & \text{if } j_k = \sigma_k^* \text{ for all } k < t, \\ \pi_t^p & \text{otherwise.} \end{cases}$$

$$\sigma_t^* \equiv \begin{cases} \sigma_t & \text{if } i_k = \pi_k^* \text{ for all } k < t, \\ \sigma_t^p & \text{otherwise.} \end{cases}$$

Here, $i_t (j_t)$ denotes the action taken by player $A (B)$ at stage t of the play. Clearly, $\gamma(\pi^*, \sigma^*) = \gamma(\pi, \sigma) = (x, y)$. Suppose player A were to play π' such that $\pi_k' \neq \pi_k^*$ for some k, then player B would play according to σ^p from then on. Since, $\gamma^A(\pi', \sigma^p) \leq v^A + \varepsilon < x$, it follows immediately that player A cannot improve against σ^*. A similar statement holds in case player B deviates unilaterally. Hence, (π^*, σ^*) is an equilibrium.

Such a pair of strategies (π^*, σ^*) is called an equilibrium involving threats, e.g., Hart [28], Van Damme [74], Thuijsman and Vrieze [71].

Joosten et al. [42] prove by construction that each reward in the convex hull of E can be supported by an equilibrium, too. Equilibrium rewards in the convex hull of E *not* in E can be obtained by history-dependent strategies with threats, which are *neither* jointly-convergent, *nor* pure. The construction of Joosten et al. [42] involves a randomization phase which obviously violates the pure-strategy part. The randomization phase serves to identify and communicate to both players which equilibrium pair of jointly convergent pure strategies is to be played afterwards. So, this also violates the very notion of jointly convergent strategies. This construction need not work for every stochastic game, but for the present class of games it does as no state is absorbing (permanently).

Whether equilibria exist yielding rewards that are not in the convex hull of E, is an open question. Such equilibria then must be associated with strategies which are not jointly convergent. For instance, in the example here, it can be shown by construction that rewards in the convex hull of $\left(\frac{417}{220}, \frac{417}{220}\right)$ and $P^{\mathcal{JC}}$ can be obtained for the average reward criterion using the limes inferior. Similarly, although this is out of the scope of this chapter, one can obtain the convex hull of $\left(\frac{7}{4}, \frac{7}{4}\right)$ and $P^{\mathcal{JC}}$ as feasible rewards for the average reward criterion using the limes superior. For the latter criterion all additional rewards Pareto dominate all equilibrium rewards in $P^{\mathcal{JC}}$. Therefore, these rewards can be supported by equilibria as well for this alternative evaluation criterion.

Theorem 1 hinges on the possibility of punishing unilateral deviations, as in e.g., Hämäläinen et al. [25]. So, we cannot restrict ourselves to Markov or stationary strategies as these types of strategies do not offer the strategic richness to allow punishing. History-dependent strategies do offer the required flexibility, but it is an open question whether less general classes of strategies might suffice. What is clear though, is that action independent strategies do not.

There is no contradiction between strategy pairs being both jointly-convergent and history-dependent, or for that matter cooperative, e.g., Tołwinski [72], Tołwinski et al. [73], Krawczyk and To łwinski [48], or incentive strategies, or combinations, e.g., Ehtamo and Hämäläinen [15–18].

12.4.3 On Computing Threat Points

We illustrate Theorem 1 and the notions introduced. Moreover, we use the examples to show the scope of the problem of computing threat points. The next example shows that linear programs may not suffice.

Example 4 Assume that player B uses his second action at all stages of the play. Now, consider the (nonlinear) program

$$\min_{q_2,q_4,q_6,q_8} 6q_2 + \tfrac{11}{2}q_4 + 3q_6 + \tfrac{11}{4}q_8$$
$$\text{s.t. } 1 = q_2 + q_4 + q_6 + q_8$$
$$0 = (1 - p_2)q_2 + (1 - p_4)q_4 - p_6q_6 - p_8q_8$$
$$p_2 = \left[\tfrac{6}{10} - \tfrac{11}{20}q_4 - \tfrac{11}{10}q_8\right]_+$$
$$p_4 = \left[\tfrac{3}{10} - \tfrac{11}{16}q_4 - \tfrac{11}{8}q_8\right]_+$$
$$p_6 = \left[\tfrac{4}{10} - \tfrac{11}{8}q_4 - \tfrac{11}{4}q_8\right]_+$$
$$p_8 = \left[\tfrac{1}{10} - \tfrac{11}{4}q_4 - \tfrac{11}{2}q_8\right]_+$$
$$0 \le q_2, q_4, q_6, q_8.$$

Clearly, $q_8 = 1$ yields rewards equal to $\tfrac{11}{4}$; all other feasible rewards involve $q_8 < 1$ yielding a reward strictly higher than $\tfrac{11}{4}$. Evidently, player B can guarantee himself at least 2.75. This implies $v^B \ge 2.75$.

Next, we aim to show that player A can hold his opponent to at most 2.75 by using his second action at all stages of the play. First, we argue that the best reply of player B resulting in a pair of jointly convergent strategies yields at most 2.75. Then, we argue that if B uses a strategy resulting in a pair of strategies which is not jointly convergent, then this cannot yield more than 2.75. We do not provide the lengthy computations underlying our findings,[6] only intuitions.

For the first part, since we assume that the pair of strategies is jointly convergent, we may consider the (nonlinear) program

$$\max_{q_3,q_4,q_7,q_8} \tfrac{7}{2}q_3 + \tfrac{11}{2}q_4 + \tfrac{7}{4}q_7 + \tfrac{11}{4}q_8$$
$$\text{s.t. } 1 = q_3 + q_4 + q_7 + q_8$$
$$0 = (1 - p_3)q_3 + (1 - p_4)q_4 - p_7q_7 - p_8q_8$$
$$p_3 = \left[\tfrac{6}{10} - \tfrac{11}{20}q_4 - \tfrac{11}{10}q_8\right]_+$$
$$p_4 = \left[\tfrac{3}{10} - \tfrac{11}{16}q_4 - \tfrac{11}{8}q_8\right]_+$$
$$p_7 = \left[\tfrac{4}{10} - \tfrac{11}{8}q_4 - \tfrac{11}{4}q_8\right]_+$$
$$p_8 = \left[\tfrac{1}{10} - \tfrac{11}{4}q_4 - \tfrac{11}{2}q_8\right]_+$$
$$0 \le q_3, q_4, q_7, q_8.$$

Observe that if $p_7 = 0$, then $p_8 = 0$ as well, hence $q_3 = q_4 = 0$. Then, the maximization program implies $q_8 = 1$ and the value of the objective function is $\tfrac{11}{4}$. Let us define $e_k = (q_3, q_4, q_7, q_8)$ by $q_k = 1$, $q_j = 0$ for $j \ne k$. Now, $p_7 = 0$ if the relative frequency vector (q_3, q_4, q_7, q_8) is in

$$S^0 = conv\left\{\{e_4, e_8, \left(\tfrac{78}{110}, \tfrac{32}{110}, 0, 0\right), \left(0, \tfrac{32}{110}, \tfrac{78}{110}, 0\right)\right\}$$
$$\cup \left\{\left(0, 0, \tfrac{94}{110}, \tfrac{16}{110}\right), \left(\tfrac{94}{110}, 0, 0, \tfrac{16}{110}\right)\right\}\right\},$$

where $conv\, S$ denotes the convex hull of set S. Possible higher rewards are only to be found for $(q_3, q_4, q_7, q_8) \in \Delta^3 \setminus S^0$.

Furthermore, $1 - p_4 > 1 - p_3 \ge \tfrac{4}{10} \ge p_7 > p_8$, hence $q_3 + q_4 \le \tfrac{1}{2} \le q_7 + q_8$. So, only tuples (q_3, q_4, q_7, q_8) in

---

[6]They are available on request, of course.

$$S^1 = conv \left\{ \left\{ \left(\tfrac{1}{2}, 0, \tfrac{1}{2}, 0 \right), \left(\tfrac{1}{2}, 0, \tfrac{39}{110}, \tfrac{16}{110} \right), \left(\tfrac{23}{110}, \tfrac{32}{110}, \tfrac{1}{2}, 0 \right) \right\} \right.$$
$$\left. \cup \left\{ e_7, \left(0, 0, \tfrac{94}{110}, \tfrac{16}{110} \right), \left(0, \tfrac{32}{110}, \tfrac{78}{110}, 0 \right) \right\} \right\}.$$

may yield higher rewards than $\tfrac{11}{4}$. This follows from the observation that the sum of the probabilities to move to (from) *Low* is always above (below) $\tfrac{4}{10}$, hence the (long term) proportion of the play spent in *Low* is at least $\tfrac{1}{2}$.

The points in S^1 satisfying the restriction

$$0 = (1 - p_3)q_3 + (1 - p_4)q_4 - p_7q_7 - p_8q_8$$

form a two-dimensional manifold, say M, and the restriction is clearly violated in a neighborhood of the plane

$$P = conv \left\{ \left(\tfrac{1}{2}, 0, \tfrac{39}{110}, \tfrac{16}{110} \right), \left(\tfrac{23}{110}, \tfrac{32}{110}, \tfrac{1}{2}, 0 \right), \left(0, 0, \tfrac{94}{110}, \tfrac{16}{110} \right), \left(0, \tfrac{16}{55}, \tfrac{39}{55}, 0 \right) \right\}$$

which is the facet of S^1 opposite the line segment $\left(\tfrac{1}{2} - x, 0, \tfrac{1}{2} + x, 0 \right)$, $x \in [0, \tfrac{1}{2}]$. Hence, M does not intersect P. The following defines for $\alpha \in [0, \tfrac{16}{55}]$ a family of two-dimensional planes in S^1:

$$S(\alpha) = \left\{ (q_3, q_4, q_7, q_8) \in S^1 | q_4 + q_8 = \alpha \right\}.$$

For increasing α, we establish whether $S(\alpha) \cap M \neq \varnothing$, and in that case the intersection is either a point, a line segment or a two-dimensional subset of $S(\alpha)$. Any unique point in this intersection with the highest weight on q_4 clearly maximizes the objective function for $S(\alpha)$; otherwise a one-dimensional set of points exist with highest weights on q_4, then the point with the highest weight on q_3 is the solution with respect to $S(\alpha)$. So, for fixed α one observes immediately that $q_4 = \alpha$ and $q_8 = 0$ for any solution with respect to $S(\alpha)$.

Take $q_3 + q_7 = 1$, then $1 - p_3 = p_7 = \tfrac{4}{10}$ which in turn implies $q_3 = q_7 = \tfrac{1}{2}$. In this case, $\tfrac{7}{2}q_3 + \tfrac{11}{2}q_4 + \tfrac{7}{4}q_7 + \tfrac{11}{4}q_8 = \tfrac{21}{8}$. To obtain higher values of the objective function q_4 should be increased from zero while keeping $q_8 = 0$. The final point is that the one-dimensional set of solutions restricted to such $S(\alpha)$ for $\alpha \in [0, \tfrac{16}{55}]$ "beginning at" $\left(\tfrac{1}{2}, 0, \tfrac{1}{2}, 0 \right)$ does not lead to higher values of the objective function than $\tfrac{21}{8}$.

As no solution satisfying the restrictions of the maximization problem, yields more than $\tfrac{11}{4}$ in $\Delta^3 \backslash S^0$, the solution is located in S^0, so the global solution is $q_8 = 1$; the connected reward to player B is 2.75. As player A can hold B to this amount, we have $v^B \leq 2.75$. Hence, under the assumption that the outcome of the maximization problem of player B against his opponent using his second action in any state and at any stage, is a jointly convergent pair of strategies, we find $v^B = 2.75$.

Now, we continue our reasoning with the **assumption that the maximization problem does not** result in a pair of jointly-convergent strategies. First, note that the latter expression in the present framework means that B uses a strategy σ against

player A playing $\pi^* = (2, 2, 2, \ldots)$, such that $q^t = \left(q_3^t, q_4^t, q_7^t, q_8^t\right)$ never converges, i.e., q^t must move around in the three-dimensional unit simplex forever.

Note that if for some (non-jointly convergent) pair of strategies (π^*, σ) and some T, it holds that[7] $\{q^t\}_{t \geq T} \subset S^0$, then $\lim_{t \to \infty} q_3^t = \lim_{t \to \infty} q_4^t = 0$. This follows from the circumstance that $p_7\left(q^t\right) = p_8\left(q^t\right) = 0$ for all $t \geq T$. So, the long-term average payoffs at point t in time for t sufficiently large satisfy

$$\frac{7}{4}q_7^t + \frac{11}{4}q_8^t = \frac{7}{4}q_7^t + \frac{11}{4}\left(1 - q_7^t\right) = \frac{11}{4} - q_7^t < \frac{11}{4}.$$

This means that $\gamma^B(\pi^*, \sigma) < \frac{11}{4}$.

Furthermore, let $S^2 = conv\{e_7, e_8, \left(\frac{4}{7}, \frac{3}{7}, 0, 0\right), \left(0, \frac{11}{15}, \frac{4}{15}, 0\right)\}$. Then it is easily confirmed that $\frac{7}{2}q_3 + \frac{11}{2}q_4 + \frac{7}{4}q_7 + \frac{11}{4}q_8 \leq \frac{11}{4}$ for all $q \in S^2$. Hence, if for some (π^*, σ) it holds that

$$\limsup_{T \to \infty} \left[\Pr_{\pi^*, \sigma} \left[\frac{\#\left\{q^t \subset S^2 | t \leq T\right\}}{T} \right] \geq \varepsilon \right] > 0 \text{ for all } \varepsilon > 0,$$

then $\gamma^B(\pi^*, \sigma) \leq \frac{11}{4}$.

Let $S^3 = \Delta^3 \setminus \left(S^0 \cup S^2\right)$ and note that $\frac{7}{2}q_3 + \frac{11}{2}q_4 + \frac{7}{4}q_7 + \frac{11}{4}q_8 \geq \frac{11}{4}$ for $q \in S^3$. By choosing a set of convenient (but not even tight) upper and lower bounds it takes quite some effort to confirm that if for some (π^*, σ)

$$\limsup_{T \to \infty} \left[\Pr_{\pi^*, \sigma} \left[\frac{\#\left\{q^t \subset S^3 | t \leq T\right\}}{T} \right] \geq \varepsilon \right] = 0 \text{ for all } \varepsilon > 0,$$

then $\gamma^B(\pi^*, \sigma) < \frac{11}{4}$. This contradiction implies that it is impossible to guarantee play such that the resulting relative frequencies vectors stay in S^3 (hence out of $S^0 \cup S^2$) almost forever.

So, candidates to yield a limiting average reward higher than $\frac{11}{4}$ must induce play such that relative frequency vectors stay forever in $\left(S^0 \setminus S^2\right) \cup S^3$. However, in $\left(S^0 \setminus S^2\right) \cup S^3$ there is persistent drift away from $conv\{e_3, e_4\}$ because the transition probabilities from Low to High are small and the transition probabilities from High to Low are large. Away from $conv\{e_3, e_4\}$ means towards $conv\{e_7, e_8\}$ which implies that the play will induce relative frequency vectors in S^2. Note that due to the assumption that (π^*, σ) is not jointly convergent means e_8 can only be approached infinitely often by relative frequency vectors from S^2 or returning to S^2, yielding limiting average rewards below $\frac{11}{4}$.

The negative results above imply that the maximization problem can be solved in jointly convergent strategies in this example. Hence, $v^B = 2.75$ (Fig. 12.4).

---

[7]Hordijk et al. [34] show that a stationary strategy suffices as a best reply against a fixed stationary strategy, and we may write the next sequence as a deterministic one.
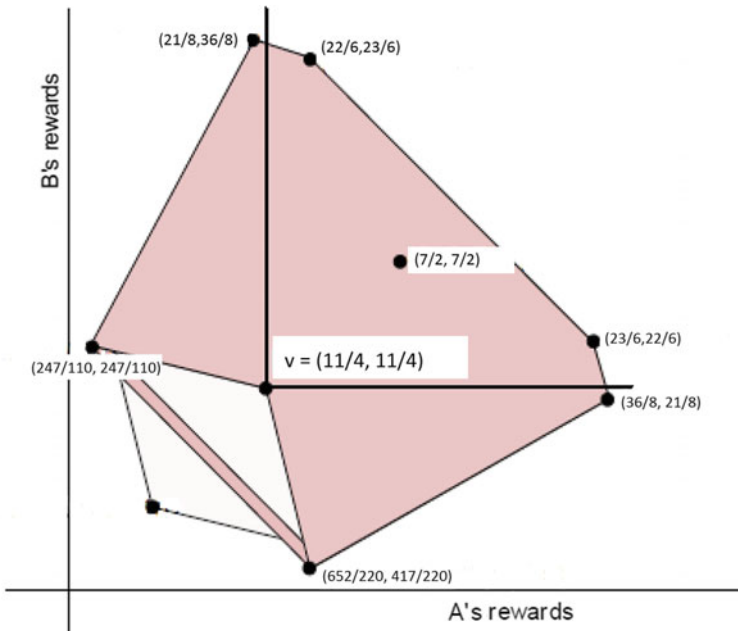
Fig. 12.4 Each pair of jointly convergent pure-strategy rewards to the "north-east" of $v = (2.75, 2.75)$ can be supported by an equilibrium involving threats

Example 4 illustrates that finding threat points may be cumbersome as it requires at least a nonlinear program. Our approach was to alternate a minimization and a maximization program against sequences of stationary strategies to obtain lower and upper bounds for the threat point. If solutions coincide, as in the example above after two steps, we are done. Otherwise, all rewards yielding more than the lowest upper bound established can be associated to equilibria involving threats.

We can interpret every minimization and maximization program as a single controller stochastic game (cf., e.g., Parthasarathy and Raghavan [60]). However, the circumstance that the number of states captured in the relative frequency vectors (please recall our remarks on this issue in Sect. 12.3) is not finite takes our problem out of the scope of the algorithms implied to compute the associated values (e.g., Filar and Raghavan [19], Vrieze [75], see Raghavan and Filar [62] for a survey). Hordijk et al. [34] show that a stationary strategy suffices as a best reply against a fixed stationary strategy, and the optimization problems mentioned reduce to Markov decision problems (cf., e.g., Filar and Vrieze [20]). We used these results partially above,[8] but found not much help in them otherwise.

---

[8]In earlier versions of our paper we were too quick to conclude that the associated optimization problems yield jointly convergent strategies. A referee pointed out a flaw in our reasoning, which by the way, makes to problem of finding an optimal strategy against a fixed strategy even much harder to solve. If jointly convergent strategies do not yield a solution, play never settles down measured in

The *general* problem is equivalent to finding the value of a zero-sum stochastic game. Well-known techniques from standard stochastic game theory, e.g., Bewley and Kohlberg [6, 7] and Mertens and Neyman [56], offer insufficient solace because of the state space which is not finite but denumerable.

12.5   Conclusions

We added an innovation to the framework of Small Fish Wars (e.g., Joosten [37, 38, 41]) by allowing endogeneity in the transition structure: transition probabilities depend on the actions taken by the agents currently in the current state and on the history of the play. In this new setting states may become absorbing *temporarily*. Here, this feature is used to model the phenomenon that, even if the agents turn to ecologically sound exploitation policies, it may take a long time before the first transition to a state yielding higher outcomes occurs if the state *Low* turns out to have become temporarily absorbing. Thus, we capture hysteresis, called a poaching pit in the management of natural resources literature (cf., e.g., Bulte [11]). Hysteresis is an empirical phenomenon and may be observed in the slow recovery of coastal cod stocks in Canada after a moratorium on cod fishing since 1992 (cf., Rose et al. [63]). More recent estimates of stocks show a less bleak picture due to recent developments unrelated to resource management, but the stocks are still far removed from high historical levels.

Our approach generalizes standard stochastic games,[9] too. We propose methods of analysis originally introduced in Joosten et al. [42] inspired by Folk Theorems for stochastic games e.g., Thuijsman and Vrieze [71], Joosten [35, 36] and Schoenmakers [67], and developed further in for instance Joosten [37, 38, 41]. Crucial notion is that of jointly-convergent strategies which justify the necessary steps in creating analogies to the Folk Theorem. In our view, it is convenient that the complex model arising from endogenous transition probabilities may be solved quite analogously to repeated games.[10]

---

the space of the relative frequency vectors and the sequence of relative frequency vectors induced is essentially stochastic.

[9]At several presentations the question was raised whether our games should not be presented as stochastic games with infinitely many states. We agree that our games fall into this class, as they can be rewritten as such. We prefer our presentation because of its simplicity and the circumstance that we were able to generate a number of results. Moreover, we are very sceptic about which known results from the analysis of stochastic games with infinitely many states would be helpful to obtain results for ours.

[10]We like our rather complex model to resemble repeated games for psychological reasons and for reasons of ease of communication for instance with less mathematically inclined people (politicians, civil servants). Many people have learned about the repeated prisoners' dilemma in educational programs, so offering our model in a simple fashion may offer windows of opportunity for communication with the general public. To present our model as a stochastic game with infinitely many states might scare researchers but more likely less mathematically inclined people away.

Our analysis of a special example with hysteresis shows that a "tragedy of the commons" can be averted by sufficiently patient rational agents[11] maximizing their utilities non-cooperatively. All equilibrium rewards yield more than the amounts associated to the permanent ruthless exploitation of the resource. Pareto optimal equilibrium rewards correspond to strategy pairs involving a considerable amount of restraint on the part of the agents, and are considerably higher than no-restraint rewards and slightly higher than perfect-restraint rewards.

To present a tractable model and to economize on notations, we kept the fish stock fixed yet stochastic, i.e., the variation in stock size and catches is only due to random effects; we imposed symmetry and used the three "twos": two states, two players and two actions. Two distinct states allow to model the kind of transitions we had in mind; two agents are minimally required to model strategic interaction; two stage-game actions leave something to choose. In order to capture additional real-life phenomena observed, such as seasonalities or other types of correlations, a larger number of states may be required. Furthermore, more levels or dimensions of restraining measures may be necessary. Adding states, (asymmetric) players or actions changes nothing to our approach conceptually.

By keeping the model and its analysis relatively simple, hence presumably more tractable, further links to and comparisons with contributions in the social dilemma literature, cf., e.g., Komorita and Parks [1994], Heckathorn [30], Marwell and Oliver [53] where dyadic choice is predominant, may be facilitated. Our resource game is to be associated primarily with a social trap, see e.g., Hamburger [26], Platt [61], Cross and Guyer [14] of which the 'tragedy of the commons' cf., e.g., Hardin [27], Messick et al. [55], Messick and Brewer [54]) is a special notorious example.

Ongoing related research focusses on designing algorithms improving computational efficiency of existing ones to generate large sets of jointly-convergent pure-strategy rewards. The algorithms used to find the rewards visualized in consecutive figures in this chapter are unacceptably slow. This was an unpleasant surprise as they were in fact modifications of algorithms working extremely rapidly in models within the same and related frameworks (e.g., Joosten [37, 39–41]). The new algorithms not only generate the desired sets within acceptable computing times here, but also seem much more efficient than our algorithms used before when applied to certain repeated games, stochastic games and games with frequency dependent stage payoffs (cf., Joosten and Samuel [43, 44]).

Related ongoing research is devoted to computing threat points with spin-offs of the algorithms of Joosten and Samuel [43, 44] for the same models as mentioned in the previous paragraph. This is a solution born out of necessity because very little

---

[11]Our agent is not the individual fisherman, but rather countries, regions, villages or cooperatives. Whether or not the latter care for the future sufficiently to induce sustainability (see e.g., Ostrom [58], Ostrom et al. [59] for optimistic views), individual fisherman's preferences seem too myopic (cf., e.g., Hillis and Wheelan [32]). Next to impatience of the agents, their number, communication, punishment possibilities and the observability of actions taken influence the likelihood that the tragedy of the commons can be averted (cf., e.g., Komorita and Parks [47], Ostrom [58, 59], Steg [70]).

is known on finding threat points in this new framework. Future research should address this knowledge gap.

Future research should combine the various modifications and extensions of the original Small Fish Wars [37] with the innovation presented here. Joosten [41] adds various price-scarcity feedbacks to the model, as well as another low-density phenomenon called the Allee effect. For the majority of results and our methods of analysis we anticipate to need no more than the notion of jointly convergent strategies and continuity of stage payoff functions and transition probability functions involved.

We envision applications of stochastic games with endogenous transitions where hysteresis-like phenomena occur, for instance shallow lakes (e.g., Scheffer [65], Carpenter et al. [12], Mäler et al. [52]), labor markets (e.g., Blanchard and Summers [9]), climate change (e.g., Lenton et al. [49]), or more general, where tipping or regime shifts may occur [2, 66]. We also see possible extensions of earlier models on (un)learning by (not) doing, cf., Joosten et al. [42, 45], and related work, e.g., Schoenmakers et al. [68], Schoenmakers [67], Flesch et al. [22].

References

1. Amir, R.: Stochastic games in economics and related fields: an overview. In: Neyman, A., Sorin, S. (eds.) Stochastic Games and Applications. NATO Advanced Study Institute, Series D, pp. 455–470. Kluwer, Dordrecht (2003)
2. Anderson, T., Carstensen, J., Hernández-Garcia, E., Duarte, C.M.: Ecological thresholds and regime shifts: approaches to identification. Trends Ecol. Evol. **24**, 49–57 (2008)
3. Armstrong, M.J., Connolly, P., Nash, R.D.M., Pawson, M.G., Alesworth, E., Coulahan, P.J., Dickey-Collas, M., Milligan, S.P., O'Neill, M., Witthames, P.R., Woolner, L.: An application of the annual egg production method to estimate spawning biomass of cod (Gadus morhua L.), plaice (Pleuronectes platessa L.) and sole (Solea solea L.) in the Irish Sea. ICES J. Mar. Sci. **58**, 183–203 (2001)
4. Aumann, R.: Game engineering. In: Neogy, S.K., Bapat, R.B., Das, A.K., Parthasarathy, T. (eds.) Mathematical Programming and Game Theory for Decision Making, pp. 279–286. World Scientific, Singapore (2008)
5. BenDor, T., Scheffran, J., Hannon, B.: Ecological and economic sustainability in fishery management: a multi-agent model for understanding competition and cooperation. Ecol. Econ. **68**, 1061–1073 (2009)
6. Bewley, T., Kohlberg, E.: The asymptotic theory of stochastic games. Math Oper Res. **1**, 197–208 (1976)
7. Bewley, T., Kohlberg, E.: The asymptotic solution of a recursive equation occuring in stochastic games. Math. Oper. Res. **1**, 321–336 (1976)
8. Billingsley, P.: Probability and Measure. Wiley, New York (1986)
9. Blanchard, O., Summers, L.: Hysteresis and the European unemployment problem. In: Fisher, S. (ed.) NBER Macroecon. Annu., pp. 15–78. MIT Press, Cambridge (1986)
10. Brooks, S.E., Reynolds, J.D., Allison, A.E.: Sustained by snakes? seasonal livelihood strategies and resource conservation by Tonle Sap fishers in Cambodia. Hum. Ecol. **36**, 835–851 (2008)
11. Bulte, E.H.: Open access harvesting of wildlife:the poaching pit and conservation of endangered species. Agric. Econ. **28**, 27–37 (2003)
12. Carpenter, S.R., Ludwig, D., Brock, W.A.: Management of eutrophication for lakes subject to potentially irreversible change. Ecol. Appl. **9**, 751–771 (1999)

13. Courchamp, F., Angulo, E., Rivalan, P., Hall, R.J., Signoret, L., Meinard, Y.: Rarity value and species extinction: the anthropogenic Allee effect. PLoS Biol. **4**, 2405–2410 (2006)
14. Cross, J.G., Guyer, M.J.: Social Traps. University of Michigan Press, Ann Arbor (1980)
15. Ehtamo, H., Hämäläinen, R.P.: On affine incentives for dynamic decision problems. In: Başar, T. (ed.) Dynamic Games and Applications in Economics, pp. 47–63. Springer, Berlin (1986)
16. Ehtamo, H., Hämäläinen, R.P.: Incentive strategies and equilibria for dynamic games with delayed information. JOTA **63**, 355–369 (1989)
17. Ehtamo, H., Hämäläinen, R.P.: A cooperative incentive equilibrium for a resource management problem. J. Econ. Dyn. Control. **17**, 659–678 (1993)
18. Ehtamo, H., Hämäläinen, R.P.: Credibility of linear equilibrium strategies in a discrete-time fishery management game. Group Decis. Negot. **4**, 27–37 (1995)
19. Filar, J., Raghavan, T.E.S.: A matrix game solution to a single-controller stochastic game. Math. Oper. Res. **9**, 356–362 (1984)
20. Filar, J., Vrieze, O.J.: Competitive Markov Decision Processes. Springer, Berlin (1996)
21. Flesch, J.: Stochastic games with the average reward. Ph.D. thesis, Maastricht University, ISBN 90-9012162-5 (1998)
22. Flesch, J., Schoenmakers, G., Vrieze, O.J.: Loss of skills in coordination games. Int. J. Game Theory **40**, 769–789 (2011)
23. Forges, F.: An approach to communication equilibria. Econometrica **54**, 1375–1385 (1986)
24. Hall, R.J., Milner-Gulland, E.J., Courchamp, F.: Endangering the endangered: the effects of perceived rarity on species exploitation. Conserv. Lett. **1**, 75–81 (2008)
25. Hämäläinen, R.P., Haurie, A., Kaitala, V.: Equilibria and threats in a fishery management game. Optim. Control. Appl. Methods **6**, 315–333 (1985)
26. Hamburger, H.: N-person prisoner's dilemma. J. Math. Psychol. **3**, 27–48 (1973)
27. Hardin, G.: The tragedy of the commons. Science **162**, 1243–1248 (1968)
28. Hart, S.: Nonzero-sum two-person repeated games with incomplete information. Math. Oper. Res. **10**, 117–153 (1985)
29. Haurie, A., Krawczyk, J.B., Zaccour, G.: Games and Dynamic Games. World Scientific, Singapore (2012)
30. Heckathorn, D.D.: The dynamics and dilemmas of collective action. Am. Sociol. Rev. **61**, 250–277 (1996)
31. Herings P.J.J., Predtetchinski, A.: Voting in collective stopping games, working paper Maastricht University (2012)
32. Hillis, J.F., Wheelan, J.: Fisherman's time discounting rates and other factors to be taken into account in planning rehabilitation of depleted fisheries. In: Antona, M., et al. (eds.) Proceedings of the 6th Conference of the International Institute of Fisheries Economics Trade, pp. 657–670. IIFET-Secretariat, Paris (1994)
33. Holden, M.: The Common Fisheries Policy: Origin, Evaluation and Future. Fishing News Books, Blackwell (1994)
34. Hordijk, A., Vrieze, O.J., Wanrooij, L.: Semi-Markov strategies in stochastic games. Int. J. Game Theory **12**, 81–89 (1983)
35. Joosten, R.: Dynamics, Equilibria, and Values. Ph.D. thesis, Faculty of Economics and Business Administration, Maastricht University (1996)
36. Joosten, R.: A note on repeated games with vanishing actions. Int. Game Theory Rev. **7**, 107–115 (2005)
37. Joosten, R.: Small Fish Wars: a new class of dynamic fishery-management games. ICFAI J. Manag. Econ. **5**, 17–30 (2007a)
38. Joosten, R.: Small Fish Wars and an authority. In: Prinz, A. (ed.) The Rules of the Game: Institutions, Law, and Economics, pp. 131–162. LIT, Berlin (2007)
39. Joosten, R.: Strategic advertisement with externalities: a new dynamic approach. In: Neogy, S.K., Das, A.K., Bapat, R.B. (eds.) Modeling, Computation and Optimization. ISI Platinum Jubilee Series, vol. 6, pp. 21–43. World Scientific Publishing Company, Singapore (2009)
40. Joosten, R.: Long-run strategic advertisement and short-run Bertrand competition. Int. Game Theory Rev. **17**, 1540014 (2015). https://doi.org/10.1142/S0219198915400149

41. Joosten, R.: Strong and weak rarity value: resource games with complex price-scarcity relationships. Dyn. Games Appl. **16**, 97–111 (2016)
42. Joosten, R., Brenner, T., Witt, U.: Games with frequency-dependent stage payoffs. Int. J. Game Theory **31**, 609–620 (2003)
43. Joosten, R., Samuel, L.: On stochastic fishery games with endogenous stage payoffs and transition probabilities. In: Proceedings of 3rd Joint Chinese-Dutch Workshop on Game Theory and Applications and 7th China Meeting on Game Theory and Applications. CCIS-series. Springer, Berlin (2017)
44. Joosten, R., Samuel, L.: On the computation of large sets of rewards in ETP-ESP-games with communicating states. Research memorandum, Twente University, The Netherlands (2017)
45. Joosten, R., Thuijsman, F., Peters, H.: Unlearning by not doing: repeated games with vanishing actions. Games Econ. Behav. **9**, 1–7 (1993)
46. Kelly, C.J., Codling, E.A., Rogan, E.: The Irish Sea cod recovery plan: some lessons learned. ICES J. Mar. Sci. **63**, 600–610 (2006)
47. Komorita, S.S., Parks, C.D.: Social Dilemmas. Westview Press, Boulder (1996)
48. Krawczyk, J.B., Tołwinski, B.: A cooperative solution for the three nation problem of exploitation of the southern bluefin tuna. IMA J. Math. Appl. Med. Biol. **10**, 135–147 (1993)
49. Lenton, T.M., Livina, V.N., Dakos, V., Scheffer, M.: Climate bifurcation during the last deglaciation? Clim. Past **8**, 1127–1139 (2012)
50. Levhari, D., Mirman, L.: The great fish war: an example using a dynamic Cournot-Nash solution. Bell J. Econ. **11**, 322–334 (1980)
51. Long, N.V.: A Survey of Dynamic Games in Economics. World Scientific, Singapore (2010)
52. Mäler, K.-G., Xepapadeas, A., de Zeeuw, A.: The economics of shallow lakes. Environ. Resour. Econ. **26**, 603–624 (2003)
53. Marwell, G., Oliver, P.: The Critical Mass in Collective Action: A Micro-Social Theory. Cambridge University Press, Cambridge (1993)
54. Messick, D.M., Brewer, M.B.: Solving social dilemmas: a review. Annu. Rev Pers. Soc. Psychol. **4**, 11–43 (1983)
55. Messick, D.M., Wilke, H., Brewer, M.B., Kramer, P.M., Zemke, P.E., Lui, L.: Individual adaptation and structural change as solutions to social dilemmas. J Pers. Soc. Psychol. **44**, 294–309 (1983)
56. Mertens, J.F., Neyman, A.: Stochastic games. Int. J. Game Theory. **10**, 53–66 (1981)
57. Oosthuizen, E., Daan, N.: Egg fecundity and maturity of North Sea cod, gadus morhua. Neth. J. Sea Res. **8**, 378–397 (1974)
58. Ostrom, E.: Governing the Commons. Cambridge University Press, Cambridge (1990)
59. Ostrom, E., Gardner, R., Walker, J.: Rules, Games, and Common-Pool Resources. Michigan University Press, Ann Arbor (1994)
60. Parthasarathy, T., Raghavan, T.E.S.: An orderfield property for stochastic games when one player controls the transition probabilities. J. Optim. Theory Appl. **33**, 375–392 (1981)
61. Platt, J.: Social traps. Am. Psychol. **28**, 641–651 (1973)
62. Raghavan, T.E.S., Filar, J.: Algorithms for stochastic games, a survey. Z. Oper. Res. **33**, 437–472 (1991)
63. Rose, G.A., Bradbury, I.R., de Young, B., Fudge, S.B., Lawson, G.L., Mello, L.G.S., Robichaud, D., Sherwood, G., Snelgrove, P.V.R., Windle, M.J.S.: Rebuilding Atlantic Cod: Lessons from a Spawning Ground in Coastal Newfoundland. In: Kruse, G.H., et al. (eds.) 24th Lowell Wakefield Fisheries Symposium on Resiliency of gadid stocks to fishing and climate change, pp. 197–219 (2008)
64. Sanchirico, J.N., Smith, M.D., Lipton, D.W.: An empirical approach to ecosystem-based fishery management. Ecol. Econ. **64**, 586–596 (2008)
65. Scheffer, M.: The Ecology of Shallow Lakes. Chapman & Hall, London (1998)
66. Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B.: Catastrophic shifts in ecosystems. Nature **413**, 591–596 (2001)
67. Schoenmakers, G.M.: The profit of skills in repeated and stochastic games. Ph.D. thesis Maastricht University (2004)

68. Schoenmakers, G.M., Flesch, J., Thuijsman, F.: Coordination games with vanishing actions. Int. Game Theory Rev. **4**, 119–126 (2002)
69. Shapley, L.: Stochastic games. Proc. Natl. Acad. Sci. USA **39**, 1095–1100 (1953)
70. Steg, L.: Motives and behavior in social dilemmas relevant to the environment. In: Hendrickx, L., Jager, W., Steg, L. (eds.) Human Decision Making and Environmental Perception. Understanding and Assisting Human Decision Making in Real-Life Settings, pp. 83–102 (2003)
71. Thuijsman, F., Vrieze, O.J.: The power of threats in stochastic games. In: Bardi, M., et al. (eds.) Stochastic and Differential Games, Theory and Numerical Solutions, pp. 343–358. Birkhauser, Boston (1998)
72. Tołwinski, B.: A concept of cooperative equilibrium for dynamic games. Automatica **18**, 431–441 (1982)
73. Tołwinski, B., Haurie, A., Leitmann, G.: Cooperative equilibria in differential games. JOTA **119**, 182–202 (1986)
74. Van Damme, E.E.C.: Stability and Perfection of Nash Equilibria. Springer, Berlin (1992)
75. Vrieze, O.J.: Linear programming and undiscounted games in which one player controls transitions. OR Spektrum **3**, 29–35 (1981)