Spieltheorie - Wiederholte Spiele

Janina Heetjans

12.06.2012

Inhaltsverzeichnis

8 Wiederholte Spiele

8.1 Einführung und Motivation

Vom wirtschaftswissenschaftlichen Standpunkt aus gesehen, ist die Idee sehr einleuchtend Spiele wiederholt zu betrachten, da sich gerade hier Entscheidungsprozesse oft wiederholen. Unternehmen entscheiden über Preise, Angebote oder das Werbeetat und zwar jedes Jahr, jede Woche oder sogar jeden Tag.
Mathematiker wollen die Logik langer Spielinteraktionen untersuchen. Es soll herausgefunden werden, ob ein Spieler das derzeitige Verhalten eines Gegenspielers beobachten und im Hinblick auf sein zukünftiges Spielverhalten interpretieren kann. Außerdem können durch Beobachtung der wiederholten Spiele die sozialen Phänomene *Bestrafung, Kooperation* und *Drohung* erklärt werden.
Als Einführungsbeispiel zur Verdeutlichung soll das Spiel ***Gefangenendilemma*** dienen.

	G	nG
G	-3,-3	0,-4
nG	-4,0	-1,-1

In vorherigen Vorträgen wurde bereits festegestellt, dass sich hier ein Nashgleichgewicht (NGG) in der Aktionenkombination (G,G) befindet. Wiederholt man das Gefangenendilemma nun zum Beispiel zweimal, würden beide Spieler eine bessere Auszahlung erzielen, wenn sie in der ersten Periode (nG,nG) und in der zweiten das NGG (G,G) spielten. Daraus ensteht die Basisidee der wiederholten Spiele, den Ausgang (nG,nG) in so vielen Perioden wie möglich zu realisieren. Es muss also eine kooperative Strategie entwickelt werden, in der ein beidseitig erstrebenswertes Ergebnis erlangt wird. Deshalb muss jeder Spieler glauben, dass das Abschweifen von dieser Strategie die Kooperation beendet und einen fortwährenden Verlust einleitet. Dieser Verlust löscht den kurzzeitigen Gewinn aus, der durch das Abweichen von, in diesem Beispie,l (nG,nG) erzielt wurde. Es stellt sich heraus, das folgende Maßnahmen nötig sind:

- Bestrafung bei Abweichungen von der kooperativen Strategie
- Bildung von Anreizen seine Mitspieler zu bestrafen
- Einbau glaubwürdiger Drohungen

Die ***Folk - Theoreme*** zeigen schließlich, dass die Gleichgewichtslösung nicht immer die beste im Hinblick auf die Auszahlung ist. Sie besagen, dass es möglich ist, eine höhere Auszahlung zu erzielen, falls die Spieler:

- geduldig genug sind
- vorauschauend handeln

Es werden zwei Arten von wiederholten Spielen unterschieden, die endlich und die unendlich oft wiederholten Spiele. Die Ergebnisse differenzieren je nach Modellierung stark. Wie sich später zeigen wird, setzt sich im endlich oft wiederholten (e.o.w.) Spiel die Nashgleichgewichtslösung durch, während bei unendlich oft wiederholten (u.o.w.) Spielen eine teilspielperfekte Gleichgewichtslösung oft die bessere Wahl ist. Hier werden zunächst die u.o.w. Spiele behandelt.

8.2 Unendlich oft wiederholte Spiele: Definitionen

Allgemein wird für den gesamten Vortrag vorausgesetzt, dass die Aktionsmenge A kompakt und die Präferenzrelation $\succsim$ stetig ist.

2.1 Defintion: Das strategische Spiel $G = \langle N, (A_i)_{i\in N}, (\succsim_i)_{i\in N}\rangle$, das u.o.w. wird heißt ***Basisspiel***.

2.2 Definition: Sei $G = \langle N, (A_i)_{i\in N}, (\succsim_i)_{i\in N}\rangle$ ein Basisspiel mit $A = \times_{i\in N} A_i$. Dann ist das ***u.o.w. Spiel*** $G(\infty)$ ein extensives Spiel mit perfekter Information und simultanen Spielzügen $\langle N, H, P, (\succsim_i^*)\rangle$ für das gilt:

- $H = \{\emptyset\} \cup \{\cup_{t=1}^{\infty} A^t\} \cup \{A^\infty\}$, wobei mit $\{\emptyset\}$ der Initialverlauf und mit $\{A^\infty\}$ die Menge der unendlichen Sequenzen $(a^t)_{t=1}^{\infty}$ von Aktionsprofilen von G gemeint ist.
- $P(h) = N$ für jeden offenen Verlauf $h \in H$, im Unterschied zum Basisspiel ist die Menge der Spieler hier einelementig. In jeder Periode wählen alle Spieler gleichzeitig eine Aktionen. Die Spielerfunktion bildet nun in die Potenzmenge von N ab.
- Die Präferenzrelation (PR) $\succsim_i^*$, die die PR $\succsim_i$ folgendermaßen erweitert:

  Sei $(a^t) \in A^\infty, a \in A, a' \in A$ dann gilt:

  $$a \succsim_i a' \quad \Leftrightarrow \quad (a^1, ..., a^{t-1}, a, a^{t+1}, ...) \succsim_i^* (a^1, ..., a^{t-1}, a', a^{t+1}, ...)$$

  (Diese Eigenschaft wird *schwache Seperabilität* genannt.)

Wie auch in vorherigen Vorträgen ist die Bewertung der Aktionen von der Auszahlungsfunktion u abhängig. Eine mögliche Form, die unendlichen Sequenzen zu bewerten ist die Folgende:

2.3 Definition: Mit ***Diskontierung*** wird diese Relation zwischen zwei Auszahlungen bezeichnet: Sei $\delta \in (0,1)$ der ***Diskontfaktor***, dann gilt für die Sequenzen $(a^t)_{t\geq 1}$ und $(b^t)_{t\geq 1}$ von Aktionen mit $u_i((a^t)_{t\geq 1}), u_i((b^t)_{t\geq 1}) \in \mathbb{R}^N$:

$$(a^t)_{t\geq 1} \succsim_i^* (b^t)_{t\geq 1} \quad :\Leftrightarrow \quad \sum_{t=1}^{\infty} \delta^{t-1}(u_i(a^t) - u_i(b^t)) \geq 0$$

Eine Auszahlungen $u_i((a^t)_{t\geq 1})$ wird demnach identifiziert mit:

$$(1-\delta)\sum_{t=1}^{\infty} \delta^{t-1} u_i(a^t),$$

wobei $(1-\delta)$ zur Normierung der Auszahlung dient.

Diskontierung gewichtet die Perioden unterschiedlich. Liegt δ nahe bei Eins, so sind die zukünftigen Auszahlungen stark gewichtet, wird hingegen stark abdiskontiert (δ klein) ist die Zukunft weniger wichtig. Im Buch werden noch zwei weitere Bewertungssysteme vorgestellt, die die Perioden symmetrisch bewerten. Diese werden hier aus Zeitgründen aber nicht behandelt.

Ein u.o.w. Spiel $G(\infty)$ mit PR $\succsim_i^*$, das die Diskontierung zur Bewertung der Auszahlungen nutzt, wird mit $G(\infty, \delta)$ bezeichnet.

Als nächstes wird der vollständige Verhaltensplan eingeführt, der die Entwicklung einer Strategie eines jeden Spielers abstrahiert.

2.4 Definition: Ein ***vollständiger Verhaltensplan*** von $G(\infty)$ für Spieler i hat diese Komponenten:

ⓐ Die ***Zustandsmenge*** Q_i.

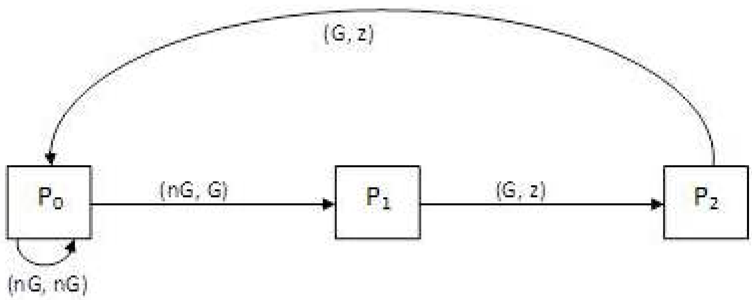
ⓑ Den ***Anfangszustand*** $q_i^0 \in Q_i$.

ⓒ Die ***Ausgabefunktion*** $f_i : Q_i \to A_i$, die jedem Zustand eine Aktion zuordnet.

ⓓ Die ***Übergangsfunktion*** $\tau_i : Q_i \times A \to Q_i$, die dem Paar aus Zustand und AKP einen Zustand zuordnet.

2.5 Beispiele: Zwei Verhaltenspläne für das u.o.w. Gefangenendilemma

1. *Bestrafung und Vergebung*



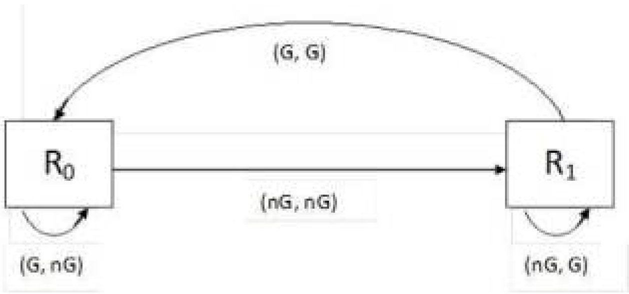
ⓐ $Q_1 = \{P_0, P_1, P_2\}$

ⓑ $q_1^0 = P_0$

ⓒ $f_1(x) = \begin{cases} nG, & \text{für } x = P_0 \\ G, & \text{für } x = P_k, k \in \{1,2\} \end{cases}$

ⓓ $\tau_1(x,y) = \begin{cases} P_0, & \text{für } (x,y) = (P_0,(nG,nG)) \text{ oder } (x,y) = (P_2,(G,z)) \text{ für } z \in \{nG, G\} \\ P_1, & \text{für } x = P_0 \text{ und } y = (nG,G) \\ P_2, & \text{für } x = P_1 \text{ und } y = (G,z), z \in \{nG, G\} \end{cases}$

2. *Gegenpol*



3. Vergleich: Spieler 1 nutzt *Bestrafung und Vergebung* (M1) und Spieler 2 nutzt *Gegenpol* (M2):

Periode	Zustand M1	Zustand M2	Ausgang	Auszahlung
1	P_0	R_0	(nG, nG)	(-1,-1)
2	P_0	R_1	(nG, G)	(-4, 0)
3	P_1	R_1	(G, G)	(-3,-3)
4	P_2	R_0	(G, nG)	(0,-4)
5	P_0	R_0	(nG, nG)	(-1,-1)

Es ist auffällig, dass in Periode 5 der Anfangszustand wieder erreicht ist. Es entwickelt sich eine Schleife der Länge $\gamma = 5$. Dies ist üblich für Verhaltenspläne mit endlicher Zustandsmenge.

2.6 Definition: Die ***Minmax - Auszahlung*** ist folgendermaßen definiert:

$$v_i = \min_{a_{-i} \in A_{-i}} \max_{a_i \in A_i} u_i(a_{-i}, a_i)$$

Die maximale Auszahlung, die ein Spieler i noch erlangen kann, wenn seine Gegenspieler die geringste Auszahlung für ihn erreichen wollen.

An dieser Stelle wird eine neue Notation eingeführt: p_{-i} ist die Aktion, die alle Spieler außer Spieler i wählen, um Spieler i zu bestrafen. $b_i(p_{-i})$ ist die bestmögliche Antwort von Spieler i auf die Bestrafung im Hinblick auf die Auszahlung.

2.7 Definition: $v \in \mathbb{R}^N$ ist ein ***erreichbares Auszahlungsprofil*** von G wenn es für v eine Linearkombination von Auszahlungsprofilen in $\mathbb{R}$ gibt mit $v = \sum_{a \in A} \alpha_a u(a)$, $\alpha_a \in \mathbb{Q}^+$ und $\sum_{a \in A} \alpha_a = 1$.

Die Menge der erreichbaren AP'e ist eine konvexe Menge, die durch Auszahlungsvektoren der reinen Strategien beschränkt ist.

2.8 Definition: Ein AP w für das gilt $w_i \geq v_i \quad \forall i \in N$, heißt ***individuell rational***. Gilt $w_i > v_i \quad \forall i \in N$, so heißt w_i ***strikt individuell rational.***

D. h. ein AP ist individuell rational, wenn die Auszahlung eines jeden Spielers $i \in N$ mind. genauso hoch wie die Minmax - Auszahlung und strikt individuell rational, wenn sie für jeden Spieler höher ist als die Minmax - Auszahlung.

8.3 Folk - Theoreme für unendlich oft wiederholte Spiele

In diesem Kapitel werden NGG'e der u.o.w. Spiele untersucht. Diese sind nicht gleichzusetzen mit der Wiederholung der NGG'e aus dem Basisspiel. Es zeigt sich, dass die AP'e der NGG von $G(\infty, \delta)$ beliebig nahe bei den erreichbaren, strikt individuell rationalen AP'en liegen.
Außerdem wird deutlich, dass zur Verwirklichung der NGG'lösung eine Strategie entwickelt werden muss, die die Abweichung von der Kooperation bestraft. Eine dieser Strategien heißt zum Beispiel Trigger[1] *- Strategie.*

3.1. Proposition: Für alle $\delta \in (0,1)$ ist jedes AP eines NGG's von $G(\infty, \delta)$ ist ein individuell rationales AP.
(*Es soll also gezeigt werden, dass die Auszahlung eines NGG's von $G(\infty, \delta)$ für alle $\delta \in (0,1)$ für alle Spieler mindestens so hoch ist wie die Minmax - Auszahlung.*)

Beweis: Sei w ein AP von $G(\infty, \delta)$, das nicht individuell rational ist, d.h. $v_i > w_i$ für einen Spieler $i \in N$. Dann kann w nicht die Auszahlung eines NGG's des wiederholten Spiels sein. Sei s die Strategie, die das AP w generiert. In einem NGG wäre die Strategie von Spieler i definiert als $s_i'(h) = b_i(s_{-i}(h)) \quad \forall h \in H$, was die beste Antwort auf die Strategie s_{-i} der Gegenspieler ist und zumindest die Auszahlung v_i in jeder Periode generiert. ↯ □

3.2 Proposition: Nash Folk - Theorem
Sei w ein erreichbares, strikt individuell rationales AP von G. Dann existiert für alle $\epsilon > 0$ ein $\bar{\delta} < 1$, so dass $G(\infty, \delta)$ für alle $\delta > \bar{\delta}$ ein NGG mit AP w' hat, für das gilt:

$$|w - w'| < \epsilon.$$

Beweis: Sei $w = \sum_{a \in A} \frac{\beta_a}{\gamma} u(a)$ ein erreichbares, individuell rationales AP, wobei β_a die Häufigkeit der Aktion $a \in A$ in einer Schleife der Länge γ ist. Demnach gilt $\gamma = \sum_{a \in A} \beta_a$. Dann entspricht $\frac{\beta_a}{\gamma}$ α_a aus der Defintion der individuell rationalen AP'e, denn es gilt

$$\sum_{a \in A} \frac{\beta_a}{\gamma} = \frac{\sum_{a \in A} \beta_a}{\gamma} = \frac{\gamma}{\gamma} = 1.$$

In einem Durchlauf der Schleife wird nun β_a mal jedes AKP $a \in A$ gespielt. Sei also (a^t) das sich hieraus ergebene AKP dieser Schleife bei unendlicher Wiederholung.
Sei s_i die Strategie von Spieler i, die sich in jeder Periode t für a_i^t entscheidet, solange alle anderen Spieler a_{-i}^t wählen. Weicht ein Spieler j in Periode t' von $a_j^{t'}$ ab, so wählt Spieler i von nun an $(p_{-j})_i$ und bleibt dabei. Dann ist s ein NGG, denn es lohnt sich für keinen Spieler abzuweichen.
Sei w' das AP der Strategie s. Noch zu zeigen:

$$|w - w'| = |\sum_{a \in A} \frac{\beta_a}{\gamma} u(a) - (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} u(a^t)| < \epsilon$$

[1]= Auslöser

Berechne w': Seien $a_1, ..., a_n$ die möglichen Aktionen von Spieler i, die mit der Häufigkeit β_{a_j} vorkommen für $j = 1, ..., n$. Setze außerdem $\beta_{a_0} = 0$ und $m = \beta_{a_{j-1}} - \beta_{a_j} - 1$ für $j = 1, ..., n$.

$$\begin{aligned}
w_i' &= (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} \; u_i(a^t) \\
&\overset{\text{Def. } a^t}{=} (1-\delta) \sum_{t=1}^{\infty} \sum_{j=1}^{n} (\sum_{l=0}^{m} \delta^{\sum_{k=0}^{j-1} \beta_{a_k} + l + \gamma(t-1)}) \; u_i(a_j) \\
&= \sum_{t=1}^{\infty} (\delta^\gamma)^{t-1} \; (1-\delta) \sum_{j=1}^{n} (\sum_{l=0}^{m} \delta^{\sum_{k=0}^{j-1} \beta_{a_k} + l}) \; u_i(a_j) \\
&= \sum_{t=1}^{\infty} (\delta^\gamma)^{t-1} \sum_{j=1}^{n} (\delta^{\sum_{k=0}^{j-1} \beta_{a_k}} - \delta^{\sum_{k=0}^{j} \beta_{a_k}}) \; u_i(a_j) \\
&= \sum_{t=1}^{\infty} (\delta^\gamma)^{t-1} \sum_{j=1}^{n} (1 - \delta^{\beta_{a_j}}) \; \delta^{\sum_{k=0}^{j-1} \beta_{a_k}} \; u_i(a_j) \\
&\overset{\text{geom. Reihe}}{=} \frac{1}{1-\delta^\gamma} \sum_{j=1}^{n} (1 - \delta^{\beta_{a_j}}) \; \delta^{\sum_{k=0}^{j-1} \beta_{a_k}} \; u_i(a_j)
\end{aligned}$$

Betrachte nun w_i' für $\delta \to 1$:

$$\begin{aligned}
&\lim_{\delta \to 1} w_i' \\
&\overset{w_i' \text{einf.}}{=} \lim_{\delta \to 1} \frac{1}{1-\delta^\gamma} \sum_{j=1}^{n} (1 - \delta^{\beta_{a_j}}) \; \delta^{\sum_{k=0}^{j-1} \beta_{a_k}} \; u_i(a_j) \\
&= \sum_{j=1}^{n} \lim_{\delta \to 1} \frac{1 - \delta^{\beta_{a_j}}}{1 - \delta^\gamma} \; \underbrace{\delta^{\sum_{k=0}^{j-1} \beta_{a_k}}}_{\to 1} \; u_i(a_j) \\
&= \sum_{j=1}^{n} \lim_{\delta \to 1} \frac{1 - \delta^{\beta_{a_j}}}{1 - \delta^\gamma} \; u_i(a_j) \\
&\overset{\text{l'hospital}}{=} \sum_{j=1}^{n} \lim_{\delta \to 1} \frac{\beta_{a_j} \overbrace{\delta^{\beta_{a_j - 1}}}^{\to 1}}{\gamma \underbrace{\delta^{\gamma - 1}}_{\to 1}} \; u_i(a_j) \\
&\overset{\text{Def. } a^t}{=} \sum_{a \in A} \frac{\beta_a}{\gamma} \; u_i(a)
\end{aligned}$$

Also haben wir jetzt gezeigt, dass für $\delta \to 1$ gilt $|w_i - w_i'| = 0$. Das heißt für alle $i \in N$ gibt es ein $\bar{\delta}_i < 1$, so dass für alle $\delta > \bar{\delta}_i$ gilt $|w_i - w_i'| < \epsilon$. Wähle nun $\bar{\delta} = \max_{i \in N} \; \bar{\delta}_i$ dann folgt die Behauptung. $\square$

Im letzten Beweis wurde die Trigger - Strategie angewandt. Das Abweichen der Gegenspieler von (a^t) war der Auslöser, der eine für immer andauernde Bestrafung einleitet.

3.3 Bemerkung: Es ist wichtig, dass die Drohung glaubwürdig ist. Hier ein Beispiel für eine unglaubwürdige Drohung.

	A	D
A	2,3	1,5
D	0,2	0,2

Es ergibt sich $p_{-1} = p_{-2} = D$ als Bestrafungsaktion. A dominiert die Aktion D strikt für Spieler 1. Er würde also selbst unter der Bestrafung leiden, was sie auf Dauer unglaubwürdig macht. Dies führt zu der Betrachtung von teilspielperfekten Gleichgewichten (TSGG), die eine optimale Lösung nach jedem Verlauf erfordern.

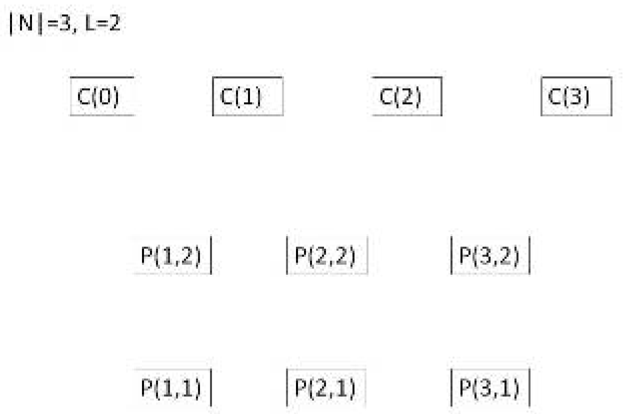
Für das teilspielperfekte Folk - Theorem ist die Trigger - Strategie zu hart. Eine Bestrafung muss nur solang andauern, bis der Gewinn des abgewichenen Spielers wieder neutralisiert ist. Hinzu kommt allerdings, dass auch die Spieler bestraft werden müssen, die es versäumen zu bestrafen, denn wie wir gerade gesehen haben, kann es für einen Spieler ebenfalls unvorteilhaft sein zu bestrafen. Wenn der Diskontfaktor hingegen sehr klein ist, müssen Bestrafungen eine unbeschränkt lange Zeit andauern. Deshalb wird eine neue Strategie eingeführt, in der die Spieler belohnt werden, die vorschriftsmäßig bestrafen.

3.4 Proposition: Teilspielperfektes Folk - Theorem
Sei a^* ein AKP, das eine strikt individuell rationale Auszahlung einbringt. Außerdem seien $(a(i))_{i \in N}$ Aktionen von G, die eine strikt individuell rationale Auszahlung generieren, so dass für jeden Spieler $i \in N$ gilt:

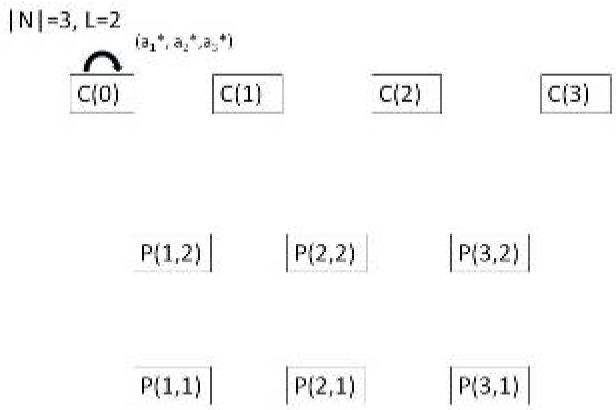
$$a^* \succ_i a(i) \text{ und } a(j) \succ_i a(i) \quad \forall j \in N \setminus \{i\}.$$

Dann existiert ein $\bar{\delta} < 1$, so dass es für alle $\delta > \bar{\delta}$ ein TSGG von $G(\infty, \delta)$ gibt, das einen Pfad $(a^t)_{t \geq 0}$ generiert, für den $a^* = a^t$ für alle t gilt.
(*Es wird gezeigt, dass die TSGG (a^t) den AKP'en mit einer strikt individuell rationalen Auszahlung entsprechen.*)

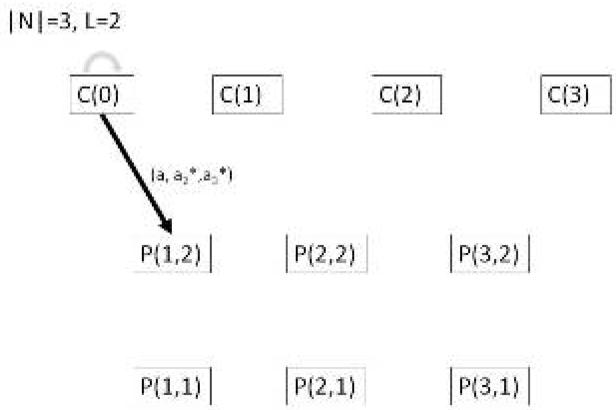
Zum Teilspielperfekten Folk - Theorem: Exemplarische Darstellung des Verhaltensplan für den Fall $|N| = 3, L = 2$:



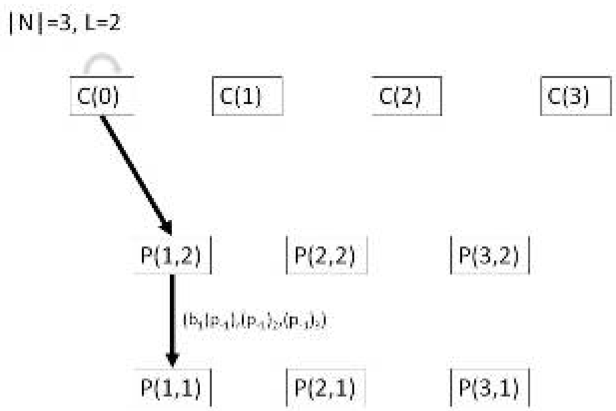
Die Zustandsmenge



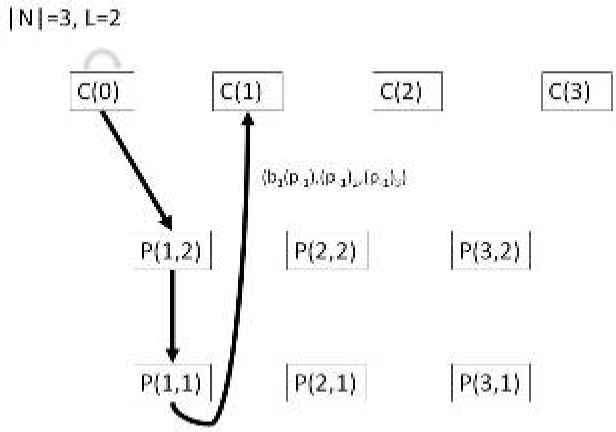
Alle drei Spieler halten die Kooperation aufrecht und spielen a^*



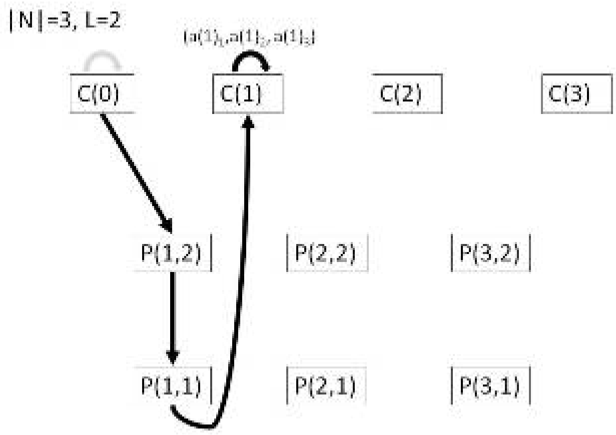
Spieler 1 weicht ab von a^*, Übergang zu $P(1,2)$



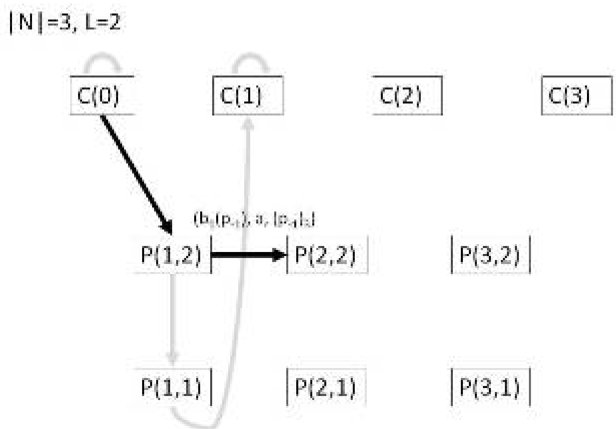
Spieler 2 und Spieler 3 bestrafen vorschriftsmäßig, Übergang zu $P(1,1)$



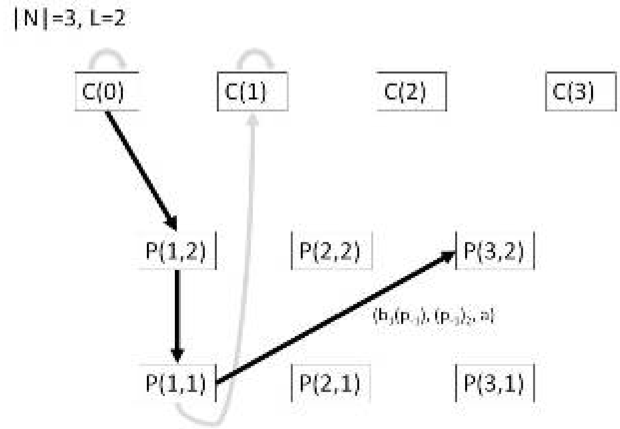
Spieler 2 und Spieler 3 bestrafen vorschriftsmäßig, Übergang zu $C(1)$



Alle drei Spieler halten die Kooperation aufrecht und spielen $a(1)$



Spieler 2 ist von p_{-1} *abgewichen, Übergang zu* $P(2,2)$



Spieler 3 ist von p_{-1} abgewichen, Übergang zu $P(3,2)$

Bisher war die Folge von AP'en abhängig von dem Abweicher und nicht von dem Verlauf, der zum Abweichen geführt hat. Außerdem war eine Bedingung, dass δ nahe bei 1 liegt. Im Buch von Osborne und Rubinstein wird an dieser Stelle noch gezeigt, dass es eine Strategie für alle $\delta \in (0,1)$ gibt, die ein TSGG von $G(\infty, \delta)$ ist.

8.4 Quellen

Martin J. Osborne, Ariel Rubinstein: *A Course in Game Theory*, 1994