

Auszug aus: R. Axelrod „Die Evolution der Kooperation“, München 1991 (2. Aufl.) , S. 7-9

Das Gefangenendilemma ist ein Spiel mit zwei Spielern, von denen jeder zwei Entscheidungsmöglichkeiten hat, nämlich zu kooperieren¹ oder nicht zu kooperieren. Nichtkooperation nennen wir Defektion². Jeder muß seine Wahl treffen, ohne zu wissen, wie der andere sich verhalten wird. Unabhängig vom Verhalten des jeweils anderen führt Defektion zu einer höheren Auszahlung als Kooperation³. Das Dilemma⁴ liegt darin, daß es für jeden Spieler, unabhängig vom Verhalten des anderen, vorteilhafter ist, zu defektieren⁵, daß jedoch beiderseitige Defektion für jeden Spieler ungünstiger ist als wechselseitige Kooperation. Wie ist das möglich? Den Ablauf dieses einfachen Spiels – welches die Grundlage des vorliegenden Buches ist – zeigt Abbildung 1. Der Zeilenspieler wählt eine Zeile, entweder Kooperation oder Defektion. Gleichzeitig wählt der Spaltenspieler eine Spalte, ebenfalls entweder Kooperation oder Defektion. Beide Entscheidungen zusammen führen auf eines der vier möglichen Ergebnisse der Matrix⁶. Wenn beide Spieler kooperieren, stehen sie recht gut da. Beide erhalten R (für engl.: *Reward*), die *Belohnung für wechselseitige Kooperation*. In der Abbildung 1 beträgt sie 3 Punkte. Diese Zahl kann z. B. eine Auszahlung in Dollar sein, die jeder Spieler bei beiderseitiger Kooperation erhält. Wenn ein Spieler kooperiert und

der andere defektiert, erhält der defektierende T (für engl.: *Temptation*), die *Versuchung zu defektieren*, während der kooperierende Spieler die *Auszahlung S* (für engl.: *Sucker's payoff*⁷) *des gutgläubigen Opfers* erhält. Im Beispiel sind $T = 5$ und $S = 0$ Punkte. Wenn beide defektieren, erhalten beide $P = 1$ (für engl.: *Punishment*), die *Strafe für wechselseitige Defektion*.

		Spaltenspieler	
		Kooperation	Defektion
Zeilenspieler	Kooperation	R = 3, R = 3	S = 0, T = 5
	Defektion	T = 5, S = 0	P = 1, P = 1

Abbildung 1: Das Gefangenendilemma

Was sollten Sie in einem solchen Spiel tun? Nehmen Sie an, Sie sind der Zeilenspieler und Sie denken, der Spaltenspieler werde kooperieren. Das bedeutet, daß Sie eines der beiden Ergebnisse in der ersten Spalte von Abbildung 1 erhalten werden. Sie haben die Wahl, entweder ebenfalls zu kooperieren und 3 Punkte zu erhalten, oder zu defektieren und 5 Punkte zu erhalten. Also lohnt es sich zu defektieren, wenn sie annehmen, der andere Spieler werde kooperieren. Nehmen Sie demgegenüber an, der andere Spieler würde defektieren. Nun sind Sie in der zweiten Spalte von Abbildung 1 und Sie haben die Wahl zwischen Kooperation, wodurch Sie zum gutwilligen Opfer würden und 0 Punkte erhielten, und

Defektion, woraus sich für Sie 1 Punkt als Ergebnis wechselseitiger Defektion ergäbe. Es lohnt sich also zu defektieren, wenn Sie denken, der andere Spieler werde defektieren. Das bedeutet, daß es besser ist zu defektieren, wenn Sie denken, der andere Spieler werde kooperieren, *und* daß es besser ist zu defektieren, wenn Sie denken, der andere Spieler werde defektieren. Also lohnt es sich für Sie, zu defektieren, was immer der andere Spieler macht.

Dieselbe Logik gilt natürlich auch für den anderen Spieler. Darum sollte der andere Spieler defektieren, unabhängig davon, was er von Ihnen erwartet. Dann aber erhalten Sie beide $P = 1$, ein schlechteres Ergebnis als $R = 3$ bei beiderseitiger Kooperation. Individuelle Rationalität⁸ führt für beide zu einem schlechteren Ergebnis als nötig. Das ist das Grundproblem des Gefangenendilemmas.

Das Gefangenendilemma ist einfach eine abstrakte Formulierung einiger sehr verbreiteter und sehr interessanter Situationen, in denen Defektion für jede Person individuell am vorteilhaftesten ist, während andererseits jeder durch wechselseitige Kooperation besser gestellt wäre. Die Definition des Gefangenendilemmas erfordert, daß zwischen den vier verschiedenen möglichen Ergebnissen bestimmte Beziehungen bestehen. Erstens gibt es eine Ordnung der vier Auszahlungen: Im besten Fall kann ein Spieler T erhalten, also die Versuchung, den kooperierenden anderen Spieler durch Defektion auszunutzen. Das schlechteste Ergebnis eines Spielers ist S , die Auszahlung des gutgläubigen Opfers, während der andere Spieler defektiert. Für die Ordnung der beiden ande-

¹ kooperieren – zusammenarbeiten
² Defektion – „Abfall“, „Treuebruch“
³ Kooperation – Zusammenarbeit
⁴ Dilemma – „Wahl zwischen zwei (gleich unangenehmen) Dingen, Zwangslage“ (Duden 5)
⁵ defektieren – „abfallen“, „die Treue brechen“
⁶ Matrix – Zahlenschema

⁷ sucker – „Dummkopf“; payoff – „Auszahlung“, „Resultat“

⁸ Rationalität – Vernünftigkeit

ren Ergebnisse wird angenommen, daß die Belohnung R für wechselseitige Kooperation größer ist als P , die Strafe für wechselseitige Defektion. Das führt zu einer
85 Rangfolge $T > R > P > S$ hinsichtlich der Präferenzen⁹ für die vier Auszahlungen.

Zweitens können die Spieler nicht dadurch aus ihrem Dilemma entkommen, daß sie sich abwechselnd ausbeuten. Diese Annahme besagt, daß eine gleiche
90 Chance, auszubeuten und ausgebeutet zu werden, kein so gutes Ergebnis für einen Spieler darstellt wie wechselseitige Kooperation. Es wird daher angenommen, daß die Belohnung R für wechselseitige Kooperation größer ist als der Durchschnitt aus der Versu-
95 chung T und der Auszahlung S des gutgläubigen Opfers. Gemeinsam mit der Rangordnung der Auszahlungen definiert diese Annahme das Gefangenendilemma.

Zwei Egoisten, die das Spiel *einmal* spielen, werden also beide ihre dominante¹⁰ Strategie, nämlich Defektion wählen und jeder wird als Ergebnis weniger bekommen als das, was beide bei wechselseitiger Ko-
100 operation hätten erhalten können. Wenn das Spiel über eine bekannte endliche Anzahl von Durchgängen gespielt wird, haben die Spieler immer noch kei-
105 nen Anreiz zu kooperieren. Das gilt sicherlich für den letzten Zug, weil es dann keine Zukunft gibt, die durch ihn beeinflusst wird. Beim vorletzten Zug wird keiner der Spieler einen Anreiz zur Kooperation besit-
110 zen, weil beide die Defektion des anderen Spielers im letzten antizipieren¹¹ können. Aus diesem Gedankengang ergibt sich für jede Folge von Spieldurchgängen, die von bekannter endlicher Länge ist, daß sich das Spiel über alle Züge zurück zu beiderseitiger Defek-

115 tion ab dem ersten Schritt auflösen wird ... Nicht anwendbar ist diese Überlegung, falls die Spieler über eine unbestimmte Anzahl von Durchgängen interagieren¹² werden. Unter realistischen Umständen können die Spieler aber tatsächlich meistens nicht si-
120 cher sein, wann die letzte Interaktion¹³ zwischen ihnen stattfinden wird. Wie später gezeigt wird, kann bei einer unbestimmten Anzahl Interaktionen Kooperation entstehen. Die Aufgabe besteht dann darin, die notwendigen und hinreichenden Bedingungen für die
125 Entstehung von Kooperation ausfindig zu machen.

⁹ *Präferenz* – „Vorrang, Vorzug“ (Duden 5)

¹⁰ *dominant* – vorherrschend

¹¹ *antizipieren* – vorwegnehmen

¹² *interagieren* – wechselseitig, aufeinander bezogen handeln

¹³ *Interaktion* – wechselseitige, aufeinander bezogene Handlung