

## Wider Nash-Gleichgewichte\*

WOLFGANG SPOHN

Die Spieltheorie, im gleichen Jahr geboren wie Georg Meggle, auch wenn ihr Status nascendi länger währte, hat eine atemberaubende Karriere hingelegt. Sie hat sich zu *der* Grundlagentheorie der Wirtschaftswissenschaften entwickelt; zahllose ökonomische Phänomene lassen sich spieltheoretisch erklären; zahllose ökonomische Probleme haben eine spieltheoretische Lösung. Diese Leistung ist zweifelsohne dem Bourbaki-Programm in der Mathematik vergleichbar. In der Tat greift ihr Anspruch weit in die Sozialwissenschaften aus – auch hier kann sie auf große Erfolge verweisen, auch wenn die Reichweite des Anspruchs umstritten ist. Das Bild des rational entscheidenden Individuums, das in ihr theoretisch entwickelt wird, prägt weite Teile unserer kulturellen und politischen Ideologie.

Der Begriff des Nash-Gleichgewichts bildet dabei *die* Grundlage der Spieltheorie. Fast alle theoretischen Bemühungen drehen sich um ihn oder bauen auf ihm auf. Es gibt zwar mittlerweile viele Gleichgewichtsbegriffe, aber fast alle liegen zwischen dem engsten, dem des strikten Nash-Gleichgewichts, und dem weitesten, dem des Nash-Gleichgewichts (vgl. die Überblicksdiagramme in van Damme 1991, S. 335f.). Die Überzeugungskraft des Begriffs war und ist enorm, auch für mich.

Doch zweifle ich mittlerweile. Der Begriff beruht auf einer allen bewussten, völlig selbstverständlichen Annahme, der Annahme der kausalen Unabhängigkeit der Entscheidungen und Handlungen der Spieler. Diese ist, wie ich nicht mit exotischen Szenarien, sondern mit einer ganz geradlinigen Argumentation zeigen will, unbegründet. Lässt man sie fallen, so gelangt man zwangsläufig zu einem weiteren Gleichgewichtsbegriff, dem der Abhängigkeitsgleichgewichte, wie ich sie nenne. Diese verhalten sich signifikant anders; z. B. stellt, wie wir sehen werden, im Gefangenen-Dilemma auch die wechselseitige Kooperation ein Abhängigkeitsgleichgewicht dar. Das deutet schon an, dass sich, nimmt man diesen Begriff ernst, dramatische Änderungen an den Grundlagen der

Spieltheorie ergeben – Änderungen, die das Bild vom rational entscheidenden Individuum nicht verneinen, aber völlig umdrehen und daher gravierende theoretische und ideologische Konsequenzen haben.

Ich will versuchen, das in den nächsten vier Abschnitten glaubhaft zu machen. Im Abschnitt 1 erläutere ich kurz den Begriff des Nash-Gleichgewichts und wieso besagte Annahme in ihn eingebaut ist. Im Abschnitt 2 erkläre ich, zu welchem Gleichgewichtsbegriff man geführt wird, wenn man diese Annahme aufgibt. Im Abschnitt 3 begründe ich, wieso die Leugnung dieser Annahme keineswegs absurd ist, sondern natürlich und geboten sein kann. Abschnitt 4 malt die Konsequenzen dessen noch etwas weiter aus.

### 1. Nash-Gleichgewichte

Konzentrieren wir uns auf Zwei-Personen-Spiele in Normalform. Die Verallgemeinerung unserer Betrachtungen auf den  $n$ -Personen-Fall wird offensichtlich sein; die Übertragung auf andere Formen der Repräsentation von Spielen wäre zu durchdenken. Nennen wir die zwei Spieler Anna und Bodo. Anna hat eine Menge  $A = \{a_1, \dots, a_m\}$  von Optionen; das können wenige einfache Handlungen wie beim Knobeln sein oder auch viele komplexe Strategien wie beim Schach, die auf jeden möglichen Spielverlauf eine Antwort vorschreiben. Entsprechend hat Bodo die Menge  $B = \{b_1, \dots, b_n\}$  von Optionen. Der tatsächliche Spiel- oder, pompöser, Weltverlauf hängt von beider Entscheidung ab und möglicherweise von weiteren Kontingenzen, die die Spieler nicht im Griff haben, den so genannten Zufallszügen der Natur. Die möglichen Verläufe werden von den Spielern in der einen oder anderen Weise, ähnlich oder verschieden, bewertet. In der Normalform werden die Bewertungen von Anna und Bodo freilich auf ihre Bewertungen ihrer möglichen Strategiekombinationen reduziert, die ihre Erwartungen bezüglich der mehr oder weniger günstigen, sich aus der jeweiligen Strategiekombination ergebenden Entwicklung schon enthalten. Sei also  $u$  die Bewertungs- oder Nutzenfunktion von Anna und  $v$  die von Bodo; beide sind Funktionen von  $A \times B$ , der Menge der Strategiekombinationen, in  $\mathbb{R}$ , der Menge der reellen Zahlen.

Nach der Standardlehre verfügen Anna und Bodo nicht nur über reine, sondern auch über gemischte Strategien. Sei  $S$  die Menge von Annas gemischten Strategien, d. h. die Menge aller Wahrscheinlichkeits- oder kurz  $W$ -Verteilungen  $s$  über  $A$ ; entsprechend sei  $T$  die Menge aller  $W$ -

Verteilungen  $t$  über Bodos Menge  $B$  reiner Strategien. Bei der gemischten Strategiekombination  $\langle s, t \rangle$  hat Anna also einen erwarteten Nutzen von  $\sum_{i=1}^m \sum_{j=1}^n s(a_i) \cdot t(b_j) \cdot u(a_i, b_j)$ ; entsprechend für Bodo.

Eine Strategiekombination  $\langle s, t \rangle$  bildet nun genau dann ein *Nash-Gleichgewicht*, wenn sich kein Spieler durch eine individuelle Abweichung vom Gleichgewicht verbessern kann, wenn also für Anna für alle  $s' \in S$

$$\sum_{i,j} s(a_i) \cdot t(b_j) \cdot u(a_i, b_j) \geq \sum_{i,j} s'(a_i) \cdot t(b_j) \cdot u(a_i, b_j)$$

oder äquivalenterweise für alle  $a_k \in A$

$$\sum_{i,j} s(a_i) \cdot t(b_j) \cdot u(a_i, b_j) \geq \sum_j t(b_j) \cdot u(a_k, b_j)$$

und für Bodo das Entsprechende gilt. In gemischten Strategien gibt es in jedem Spiel mindestens ein solches Nash-Gleichgewicht.  $\langle s, t \rangle$  bildet ein *striktes Nash-Gleichgewicht*, wenn sich jeder Spieler durch individuelle Abweichung nur verschlechtern kann, wenn also in den obigen Ungleichungen  $\gg\ll$  durch  $\gg>\ll$  ersetzt wird. Ein solches muss es aber nicht geben; und wenn es eines gibt, so nur in reinen Strategien (wo also  $s(a_i) = 1$  für ein  $i$  und  $t(b_j) = 1$  für ein  $j$ ).

Ein Nash-Gleichgewicht  $\langle s, t \rangle$  kann man auch als ein *Gleichgewicht der Meinungen* verstehen. Diese Interpretation ziehe ich in der Tat vor. Denn wieso sollte Anna die gemischte Strategie  $s$  spielen? Das ist doch nur vernünftig, wenn es ihr egal ist, welche reine Strategie  $a_i$  mit  $s(a_i) > 0$  bei ihrem Auswürfeln von  $s$  herauskommt. Wie aber kann ihr das egal sein? Doch nur dann, wenn alle  $a_i$  mit  $s(a_i) > 0$  gleich gut für sie sind, d. h. den gleichen erwarteten Nutzen  $\sum_j t(b_j) \cdot u(a_i, b_j)$  haben – worin eben  $t$  Annas Meinung über Bodos reine Strategien repräsentiert. Diese Egalität ist im Gleichgewicht  $\langle s, t \rangle$  gesichert. Entsprechendes gilt für Bodos reine Strategien  $b_j$  mit  $t(b_j) > 0$ , wenn  $s$  seine Meinung über Annas mögliche Handlungen bildet. Nur in einem solchen Meinungs-gleichgewicht können die Meinungen der Spieler übereinander wechselseitig bekannt oder gemeinsames Wissen sein (wie es bzgl. der Spielstruktur und der Nutzenfunktionen der Spieler schon immer von den Spieltheoretikern angenommen wurde). Anna kann nicht an ihrer Meinung  $t$  über Bodo festhalten und gleichzeitig vermuten, dass Bodo gar nicht die Meinung  $s$  über sie hat (solange sie ihn für einen Erwarteten-Nutzen-Maximierer hält).

Man kann sich nun weit in die Begründung solcher Gleichgewich-

te vertiefen. Man kann die Rationalität des Meinungsgleichgewichts in der Rationalität des Gleichgewichts in gemischten Strategien gründen; wenn es für Anna rational ist, die gemischte Gleichgewichtsstrategie  $s$  zu spielen und Bodo sie für rational hält, so hat Bodo offenbar rationalerweise die Meinung  $s$  über Anna. So kann man reden, sofern man die Rationalität eines Gleichgewichts in gemischten Strategien schon anderweitig begründet hat. Wenn man ohne diese Voraussetzung auskommen will – was ich, wie gesagt, vorziehe –, kann man versuchen, das Meinungsgleichgewicht direkt als rational zu erweisen oder es unter die Annahmen gemeinsamen Wissens zu subsumieren. Das habe ich alles in Spohn 1982 durchdekliniert – mit der etwas skeptischen Schlussfolgerung, dass die Rechtfertigung zunächst nur bis zu den rationalisierbaren Strategien trägt, wie sie dann bei Bernheim 1984 und Pearce 1984 hießen und tiefer untersucht wurden.

Wie auch immer, akzeptieren wir fürs Weitere die übliche Annahme, dass, was rational ist, auch unter Öffentlichkeit Bestand haben muss, d. h. wechselseitiges oder gemeinsames Wissen sein darf. Dann bleiben offenbar nur Nash-Gleichgewichte als rational übrig.<sup>1</sup> Doch ist – darauf kommt es mir an, und das möchte ich nun erläutern – in dieser abstrakten Repräsentation der sozialen Situation von Anna und Bodo die in der Einleitung angedeutete Annahme schon enthalten.

Bei einem Gleichgewicht  $\langle s, t \rangle$  in gemischten Strategien würfeln Anna und Bodo ihre Strategien jeweils für sich aus; es geht nicht um das Auswürfeln einer gemeinsamen  $W$ -Verteilung  $p$  auf  $A \times B$ , wie es etwas bei den korrelierten Gleichgewichten von Aumann der Fall ist (Aumann 1974 und 1987). Das Auswürfeln des einen hat keinen Einfluss auf das Auswürfeln des andern; darin liegt die kausale Unabhängigkeit von  $s$  und  $t$ . Diese gilt dann natürlich auch für die erwürfelten Handlungen selbst. Was Anna tut, hat keinen Einfluss darauf, was Bodo tut; und vice versa.<sup>2</sup> Wie könnte es auch? Der normale Weg wäre ja, dass Bodo sieht oder sonstwie darüber informiert wird, was Anna tut, und sich entsprechend verhält. Aber das ist explizit ausgeschlossen; das wäre anders zu modellieren. Und natürlich sollen auch unbewusste Einflussnahmen oder noch exotischere Szenarien ausgeschlossen sein.

Dasselbe gilt auch in der bevorzugten, weil schwächeren Interpretation von  $\langle s, t \rangle$  als Meinungsgleichgewicht. Annas Meinung besteht aus der absoluten oder nicht-bedingten  $W$ -Verteilung  $t$  für Bodos mögliche Handlungen, die sie damit als von ihr nicht beeinflussbar hinnimmt, wie etwa das morgige Wettergeschehen. Wenn sie meinte, wenigstens probabilistischen Einfluss auf Bodos Handlungen zu haben, dann müssten ih-

re Wahrscheinlichkeiten für Bodos Handlungen von ihren Handlungen abhängen, d. h. je nach ihrer Handlung eine andere und nicht konstant wie in  $t$  sein. Vice versa für Bodo. Wenn wir also das Meinungsgleichgewicht als Nash-Gleichgewicht konzipieren, so ist darin die Meinung der Spieler über die wechselseitige kausale Unabhängigkeit ihrer Handlungen eingebaut.

Diese Schlussfolgerung ist zugegebenermaßen nicht ganz zwingend. Sie setzt einen Zusammenhang zwischen bedingten subjektiven Wahrscheinlichkeiten und kausalen Überzeugungen voraus, der begründungsbedürftig ist; schließlich ist deterministische oder probabilistische Kausalität ein schwieriger Begriff. Freilich, so wie ich diesen Zusammenhang gerade vorausgesetzt habe, erscheint er höchst plausibel und hat er auch starke Begründungen;<sup>3</sup> ich habe selbst schon früh, in Spohn 1976/78, Abschn. 3.3, für ihn argumentiert; Abschnitt 3 kommt darauf zurück.

Dennoch liegt hier der Hund begraben. Die kausale Unabhängigkeit von Annas und Bodos Handlungen will ich gar nicht bezweifeln; das erschiene mir abwegig. Die kausale Unabhängigkeit von Annas und Bodos Entscheidungen oder Intentionen ist aber etwas subtil Anderes. Sie will ich in Frage stellen; und das wird dann spezifische Auswirkungen auf die Form von Annas und Bodos subjektiven Wahrscheinlichkeiten haben.

In Abschnitt 3 will ich erklären, was ich damit meine. Im Moment will ich lediglich die Konsequenz, dass Annas und Bodos Meinungen übereinander jeweils in absoluten subjektiven Wahrscheinlichkeiten bestehen, als aufgebbar betrachten. Dann ist der Begriff des Nash-Gleichgewichts nicht mehr anwendbar. Könnte etwas anderes an seine Stelle treten? Das ist Gegenstand des nächsten Abschnitts.

## 2. Abhängigkeitsgleichgewichte

Jetzt will ich also zulassen, dass Annas durch ihre Handlungen oder reinen Strategien bedingte Wahrscheinlichkeiten für Bodos Handlungen oder reine Strategien variieren können; ihre Überzeugungen haben also die Form  $q(b_j|a_i)$ , wobei für jedes  $a_i \in A$   $q(\cdot|a_i)$  eine  $W$ -Verteilung über  $B$  ist. Umgekehrt haben Bodos Überzeugungen die Form  $r(a_i|b_j)$ , wobei für jedes  $b_j$   $r(\cdot|b_j)$  eine  $W$ -Verteilung über  $A$  ist.

Was heißt es für Anna, unter diesen Annahmen *rational* zu sein? Es heißt, den *bedingt erwarteten Nutzen zu maximieren*, d. h. ein  $a_i$  zu wäh-

len, für das  $\sum_i q(b_j|a_i) \cdot u(a_i, b_j)$  maximal ist. Das war der entscheidende Fortschritt von Fishburn 1964 gegenüber Savage 1954. Savage kannte nur absolute Wahrscheinlichkeiten des Subjekts für handlungsunabhängige Umstände; Fishburn fand Savages Trennung von Umständen und Konsequenzen problematisch – und benötigte sie auch nicht mit seiner Annahme, dass das Subjekt irgendwelche durch seine möglichen Handlungen bedingte Wahrscheinlichkeiten für den Rest der Welt hat (die ja nicht für alle Propositionen mit den Handlungen variieren müssen) und relativ dazu den bedingt erwarteten Nutzen maximiert. Diese Auffassung darf als allgemein akzeptiert gelten.

So weit unterlagen Annas und Bodos subjektive Wahrscheinlichkeiten  $q$  und  $r$  keinen weiteren Bedingungen. Nun wollen wir aber wiederum annehmen – wie das auch schon bei den Nash-Gleichgewichten der Leitgedanke war –, dass diese Wahrscheinlichkeiten kein privates Geheimnis sein dürfen, sondern Öffentlichkeit aushalten müssen, d. h. gemeinsames Wissen sein können oder gar sind. Daraus ergeben sich zwei gravierende Beschränkungen.

Die erste Beschränkung – die bei Nash-Gleichgewichten keine Rolle spielte, weil sie dort von vornherein erfüllt war – ist, dass sich Annas und Bodos bedingte Wahrscheinlichkeiten aus einer gemeinsamen  $W$ -Verteilung  $p$  über  $A \times B$  ergeben müssen, d. h. es muss ein solches  $p$  geben, so dass für alle  $i$  und  $j$   $p(b_j|a_i) = q(b_j|a_i)$  und  $p(a_i, |b_j) = r(a_i, |b_j)$ . Das ist nicht garantiert, wie man sich vielleicht durch die folgende Überlegung klar machen kann: Wenn etwa  $A$  und  $B$  jeweils nur zwei Optionen enthalten, so hat  $q$  zwei freie Parameter, etwa  $q(b_1|a_1)$  und  $q(b_1|a_2)$ ; die bedingten Wahrscheinlichkeiten für  $b_2$  sind dann festgelegt. Ebenso hat  $r$  zwei freie Parameter.  $q$  und  $r$  haben also zusammen vier freie Parameter.  $p$  hat hingegen nur drei freie Parameter, etwa  $p(a_1, b_1)$ ,  $p(a_1, b_2)$  und  $p(a_2, b_1)$ ; der vierte Wert  $p(a_2, b_2)$  muss sich dann mit den anderen zu 1 addieren. Diese Betrachtung zeigt, dass es viele  $q$  und  $r$  geben muss, die sich nicht zu einem  $p$  zusammenfügen.

Die Herleitbarkeit von Annas  $q$  und Bodos  $r$  zu einer gemeinsamen Verteilung  $p$  folgt jedoch aus der Möglichkeit gemeinsamen Wissens. Wenn es eine solche gemeinsame Verteilung  $p$  nicht gibt, dann kann Anna nicht Bodos Wahrscheinlichkeiten  $r$  kennen, wissen, dass Bodo ihre Wahrscheinlichkeiten kennt und gleichzeitig an ihrer Einschätzung  $q$  festhalten. Das ist soweit nur eine reine Konsistenzbedingung, die, wie gesagt, im Fall der Nash-Gleichgewichte garantiert war, und die zur Folge hat, dass wir nur noch diese gemeinsame Verteilung  $p$  zu betrachten brauchen.

Es tritt nun zweitens die Bedingung des gemeinsamen Wissens der Rationalität hinzu. Gemäß fraglichem  $p$  gilt ja  $p(a_i) > 0$  für einige von Annas Handlungen  $a_i \in A$ ; jedes solche  $a_i$  muss daher in Bodos Meinung gleichfalls unter manchem seiner Handlungen positive Wahrscheinlichkeit haben. Wie kann das sein? Bodo weiß, dass Anna rational ist, d. h., dass Anna, wie gerade ausgeführt, ihren bedingt erwarteten Nutzen maximiert. Wenn das nur die Handlung  $a_i$  tut, so müsste  $p(a_i) = 1$  gelten und Bodo sich sicher sein, dass Anna  $a_i$  tut. Wenn hingegen mehrere Handlungen aus Annas Sicht in Frage kommen, Bodo ihnen also zu Recht positive Wahrscheinlichkeiten zuweist, so müssen alle den gleichen maximalen bedingt erwarteten Nutzen für Anna haben. Mutatis mutandis gilt das gleiche für Bodo.

Die Annahme gemeinsamer Wissbarkeit führt uns mithin zu dem folgenden Gleichgewichtsbegriff. Die  $W$ -Verteilung  $p$  auf  $A \times B$  ist ein *Abhängigkeitsgleichgewicht* gdw. für alle  $i$  mit  $p(a_i) > 0$  und für alle  $k = 1, \dots, m$  überhaupt

$$\sum_j p(b_j|a_i) \cdot u(a_i, b_j) \geq \sum_j p(b_j|a_k) \cdot u(a_k, b_j)$$

und umgekehrt für alle  $j$  mit  $p(b_j) > 0$  und alle  $l = 1, \dots, n$  überhaupt

$$\sum_i p(a_i|b_j) \cdot v(a_i, b_j) \geq \sum_i p(a_i|b_l) v(a_i, b_l)$$

gilt, wenn also alle Handlungen, die für Anna bzw. für Bodo in Frage kommen, d. h. gemäß  $p$  positive Wahrscheinlichkeiten haben, maximalen (und daher den gleichen) bedingt erwarteten Nutzen für Anna bzw. Bodo haben.

Ich kann hier diesen Begriff nicht sehr vertiefen und muss mich auf ein paar Bemerkungen beschränken. Zunächst: Für manche  $a_k \in A$  oder  $b_l \in B$  mag  $p(a_k) = 0$  oder  $p(b_l) = 0$  gelten, so dass relativ dazu keine bedingten Wahrscheinlichkeiten definiert sind und die gerade gegebene Definition daher sinnlos ist. Dieser Mangel lässt sich freilich in präziser und adäquater Weise beheben; vgl. dazu Spohn 2007, Abschn. 2.

Dann ist anzumerken, dass man Abhängigkeitsgleichgewichte nicht mit den korrelierten Gleichgewichten von Aumann verwechseln darf (Aumann 1974 und 1987). Ein korreliertes Gleichgewicht ist ebenfalls eine gemeinsame  $W$ -Verteilung  $p$  über  $A \times B$ . Aber bei ihm geht es, grob gesagt, darum, ob es für die Spieler je für sich vorteilhaft ist, die in  $p$  festgeschriebene Abhängigkeit zu brechen und stattdessen etwas zu tun, was unter der durch  $p$  gegebenen marginalen  $W$ -Verteilung über die Op-

tionen des anderen Spielers optimal ist. Wenn es einen solchen Vorteil nicht gibt, dann wird kein Spieler die Abhängigkeit brechen, und  $p$  ist ein korreliertes Gleichgewicht. Aber schon diese grobe Charakterisierung zeigt, dass hinter korrelierten und Abhängigkeitsgleichgewichten ganz verschiedene Gedanken stecken.

Der Begriff des Abhängigkeitsgleichgewichts ist weiter als der des Nash-Gleichgewichts. Diejenigen  $p$  über  $A \times B$ , die sich in voneinander unabhängige  $W$ -Verteilungen  $s$  über  $A$  und  $t$  über  $B$  zerlegen lassen – und nur für ein solches  $p$  ergibt der Begriff des Nash-Gleichgewichts Sinn –, sind gemäß den gegebenen Definitionen offenbar genau dann (allerdings degenerierte) Abhängigkeitsgleichgewichte, wenn sie Nash-Gleichgewichte sind. In welcher Weise Abhängigkeits- über Nash-Gleichgewichte hinausgehen (und sich auch von korrelierten Gleichgewichten unterscheiden), habe ich in Spohn 2007, Abschn. 3, an einigen markanten Zwei-Personen-Spielen exemplifiziert.

Das wichtigste Beispiel ist dabei das Gefangenendilemma. Seine Normalform wird z. B. durch die Doppelmatrix in Abbildung 1 gegeben.

	$v$		
		$c$	$d$
$u$		2	3
	$c$	2	0
	$d$	0	1
		3	1

Abbildung 1: Gefangenendilemma in seiner Normalform

Dabei ist Anna die Zeilenwählerin und Bodo der Spaltenwähler, » $c$ « steht für »kooperieren« und » $d$ « für »defektieren«. Es ist leicht zu erkennen, dass  $d$  für jeden Spieler besser ist als  $c$ , egal, was der andere macht; d. h.,  $c$  wird von  $d$  strikt dominiert. Also ist  $\langle d, d \rangle$  das einzige Nash-Gleichgewicht im Gefangenendilemma.  $\langle d, d \rangle$  erweist sich auch als das einzige korrelierte Gleichgewicht in diesem Spiel.

Hingegen gibt es darin überraschenderweise zwei ganze Familien von Abhängigkeitsgleichgewichten, eine, in der die beiden Spieler gewissermaßen asymmetrisch negativ, und eine, in der sie symmetrisch positiv verschränkt sind. Letztere wird durch die Matrix in Abbildung

2 für alle  $x$  mit  $0 \leq x \leq 1$  definiert.

$p$	$c$	$d$
$c$	$\frac{1}{2}x(1+x)$	$\frac{1}{2}x(1-x)$
$d$	$\frac{1}{2}x(1-x)$	$\frac{1}{2}(1-x)(2-x)$

Abbildung 2: Abhängigkeitsgleichgewichte im Gefangenendilemma

Dass gemäß diesem  $p$  Kooperation wie Defektion für alle  $x$  für beide Spieler in der Tat den gleichen bedingt erwarteten Nutzen haben, muss man nachrechnen. Was man aber unmittelbar sieht, ist, dass  $p(\langle c, c \rangle)$  gegen 1 geht, wenn  $x$  gegen 1 geht, und  $p(\langle c, c \rangle) = 1$  für  $x = 1$ . Das heißt, dass  $p(\langle c, c \rangle) = 1$  oder eben direkt die beiderseitige Kooperation ein Abhängigkeitsgleichgewicht ist, das eben darin liegt, dass jeder Spieler glaubt, dass der andere genau dann kooperiert, wenn er kooperiert. Es ist sogar das schwach Pareto-dominante Abhängigkeitsgleichgewicht; in keinem anderen können die Spieler einen höheren bedingt erwarteten Nutzen erzielen.<sup>4</sup> (Zu Details siehe wiederum Spohn 2007, Abschn. 3.) Ich werde auf das Beispiel noch einmal zurückkommen.

Ich räume freilich ein, dass die Theorie der Abhängigkeitsgleichgewichte so weit nur höchst rudimentär entwickelt ist. Das liegt einerseits an ihrer Neuigkeit (und ihrer vielleicht fragwürdigen Sinnhaftigkeit) und andererseits an mathematischen Unerfreulichkeiten (man verwickelt sich mit ihr in Bezug auf  $n$ -Personen-Spiele in polynomische Gleichungen  $n$ -ten Grades anstelle von linearen Gleichungen). Immerhin konnte ich zeigen, dass jedes Paar  $\langle a_i, b_j \rangle$  reiner Strategien bzw. jedes  $p$  mit  $p(a_i, b_j) = 1$ , das ein Nash-Gleichgewicht schwach Pareto-dominiert, auch ein Abhängigkeitsgleichgewicht ist (s. Spohn 2007, Abschn. 4); und ich vermute, dass genau solche Paare  $\langle a_i, b_j \rangle$  reiner Strategien Abhängigkeitsgleichgewichte sind, die mindestens so gut wie die Maximin-Strategien der Spieler sind. Das legt zumindest nahe, dass genau die nicht Pareto-dominierten Kombinationen reiner Strategien vom Standpunkt der Theorie der Abhängigkeitsgleichgewichte aus die interessantesten sind.

Da Nash- immer auch Abhängigkeitsgleichgewichte sind, ist die Existenz letzterer garantiert. Das Auswahlproblem verschärft sich allerdings mit letzteren. In Bezug auf Nash-Gleichgewichte bestand dieses

darin, ob man angesichts vieler Nash-Gleichgewichte noch eine rationale Auswahl treffen und so zu spezifischeren Empfehlungen gelangen kann; innerhalb der Standard-Spieltheorie ist das umstritten.<sup>5</sup> Umso weniger kann ich dazu in Bezug auf Abhängigkeitsgleichgewichte sagen. Es gibt da einfach noch sehr viel Arbeit.

Gleichwohl lassen diese Bemerkungen erahnen, dass die Spieltheorie, wenn man sie über dem Begriff des Abhängigkeitsgleichgewichts statt dem des Nash-Gleichgewichts aufbaute, ganz anders aussähe, als sie es jetzt tut. Das führt uns auf die Frage zurück: Warum sollte man sich auf diesen Begriff einlassen? So weit er bisher zu verstehen war, unterstellt er, dass beide Spieler wenigstens glauben, mit ihrer Handlung die des anderen kausal zu beeinflussen. Da es tatsächlich keine kausalen Zirkel gibt – ihre Existenz anzunehmen wäre Unfug –, muss sich mindestens einer der Spieler, und eigentlich beide, massiv täuschen. Ein Begriff, der nur auf einer solchen Täuschung aufbaut, ist jedoch uninteressant; sie kann ja nicht gemeinsames Wissen sein.

Am Ende von Abschnitt 1 habe ich behauptet, dass das nicht unser Problem sei. Diese Behauptung ist noch rätselhaft. Der nächste Abschnitt will das Rätsel auflösen.

### *3. Zur Verursachung von und Korrelation zwischen Handlungen*

Das Argument, das ich in diesem Abschnitt führen will, wird verwickelt klingen; dabei ist es im Kern ganz einfach. In Spohn 2003, Abschn. 3–5, habe ich es, so akkurat und lückenlos ich es konnte, formal ausgeführt; hier wage ich eine informelle Kurzdarstellung, die vielleicht durchsichtiger ist, auch wenn sie die formale Version nicht überflüssig macht.

Unser Problem ist, wie gesagt, eine Erklärung für die von Abhängigkeitsgleichgewichten unterstellten handlungsbedingten Wahrscheinlichkeiten für die Handlungen des jeweils anderen Spielers zu finden, ohne den Spielern damit unhaltbare Kausalüberzeugungen zu unterstellen.

Der erste Gedanke mag sein, dass das doch gar kein Problem ist; das ist einfach die alte Geschichte von Korrelation und Kausalität. Natürlich können zwei Variablen, hier die Handlungen der beiden Spieler, korreliert, d. h. probabilistisch abhängig sein, ohne dass die eine, die zeitlich frühere, auf die andere einen kausalen Einfluss hat. Allenfalls gilt Reichenbachs Prinzip der gemeinsamen Ursache – das sich meines Erachtens beweisen lässt (s. Spohn 1994) –, wonach zwei korrelierte Variablen eine gemeinsame Ursache haben müssen, wenn nicht die eine

die andere kausal beeinflusst.

Im Allgemeinen ist das richtig; und ich werde darauf zurückkommen. Doch muss der zweite Gedanke sein, dass dieser allgemeine Punkt aus der Sicht eines Spielers nicht zutrifft, wenn es speziell um seine eigenen Handlungen geht. Aus dieser Sicht, d. h. in seinem Modell seiner Entscheidungssituation sind diese Handlungen nämlich exogene Variablen, die darin Wirkungen, aber keine Ursachen haben. Das rationale Subjekt will die wahrscheinlichen Folgen seines Handelns optimieren; Ursachen seines Handelns – ob sie nun in der praktischen Überlegung selbst oder anderswo bestehen – spielen für die Folgenoptimierung keine Rolle.<sup>6</sup> Dann kann das Subjekt aber die Korrelation zwischen seinem Handeln und anderen Variablen nicht als Folge einer gemeinsamen Ursache sehen. In diesem speziellen Fall steht die Korrelation immer für eine kausale Abhängigkeit.

Diese Einsicht ist gerade der Witz der lange kochenden Diskussion über Newcombs Problem (s. etwa Campbell/Sowden 1985), in der die große Mehrheit den Standpunkt der so genannten kausalen Entscheidungstheorie eingenommen hat, der diese Einsicht zwingend erschien. Wenn, kontrafaktischerweise, meine Lust zu rauchen und mein erhöhtes Lungenkrebsrisiko beide ausschließlich genetisch verursacht und daher bloß korreliert sind, dann wäre es absurd, mir die Lust von dieser Korrelation verderben zu lassen; an dem Lungenkrebsrisiko kann ich durch Rauchverzicht auch nichts mehr ändern. Die Exogenität der Handlungen ist dann auch in den trunkierten Graphen von Pearl 2000, Abschn. 3.2, und den manipulierten Graphen von Spirtes et al. 1993, Abschn. 3.7 – s. aber auch Spohn 1978, Abschn. 3.3 und 5.2 – theoretisch explizit entwickelt worden. Intuitiv war diese Einsicht aber schon vor all diesen theoretischen Diskussionen klar. Darum blieb nur die Möglichkeit, Gleichgewichte als Nash-Gleichgewichte zu konzipieren.

Im Allgemeinen ist auch dieser zweite Gedanke richtig. Doch hat er selbst wieder eine Lücke: in dem und *nur* in dem Fall, dass das Entscheiden des Subjekts – d. h. seine Entscheidungssituation, die nichts anderes als seine subjektive Sicht derselben ist – aus seiner Sicht nicht nur sein Handeln, sondern auch eine andere Variable  $X$  kausal beeinflusst, nimmt sein Handeln auf  $X$  zwar keinen Einfluss, ist aber trotzdem in seiner Sicht mit  $X$  korreliert – und zwar in einer für seine praktische Überlegung, d. h. für seinen bedingt erwarteten Nutzen relevanten Weise. Das ist des Pudels Kern, den ich jetzt erklären will.

Wenn es nun also doch irgendwie wichtig werden soll, dass das Handeln aufgrund gemeinsamer Ursachenverhältnisse mit anderen Va-

riablen korreliert ist, so müssen wir uns zunächst fragen, wie Handlungen überhaupt verursacht sind. Offenbar auf höchst komplexe Weise; die vielfältigsten Umstände haben Einfluss auf unser Handeln. Aber da wir nur rationales Handeln rationaler Subjekte betrachten, müssen all diese Einflüsse über die Wünsche und Überzeugungen des Subjekts, über seine Repräsentation der Entscheidungssituation inklusive seiner subjektiven Wahrscheinlichkeiten und Bewertungen vermittelt sein, die sich in einer Intention oder Entscheidung zu einer bestimmten Handlung bündeln. Das ist die direkte Ursache der Handlung gemäß der kausalen Handlungstheorie von Hempel 1961/62 und Davidson 1963, die heute die herrschende ist (natürlich ist einzuräumen, dass trotzdem noch eine opake Vermittlung zwischen mentalem Zustand und Körperbewegung notwendig ist).

Jede Entscheidungssituation verursacht mindestens eine Handlung, nämlich eine darin optimale; es können aber auch mehrere sein, insofern ja ein ganzer Handlungsverlauf zur Entscheidung kommen kann. Umgekehrt kann jede Handlung aber aus der Sicht des Subjekts nur durch genau eine Entscheidungssituation verursacht sein: mindestens eine, weil sie sonst gar keine intentionale Handlung wäre, und höchstens eine, weil man über eine Handlung nicht zweimal entscheiden kann; wird später noch einmal entschieden, so ist, jedenfalls aus der Sicht des Subjekts, vorher eigentlich noch gar nicht entschieden worden.

Ferner ist eine Entscheidungssituation eine vollständige Ursache der in ihr entschiedenen Handlung(en); Handlungen haben nicht nebenher noch andere Ursachen; das hatten wir schon festgestellt. Schließlich liegt die Entscheidungsvariable nur kausal, nicht unbedingt auch zeitlich unmittelbar vor der zugehörigen Handlungsvariablen; es wird nicht immer erst im letzten Moment entschieden.

Natürlich ist das ein recht holzschnittartiges, idealisiertes Bild vom Zusammenhang zwischen Entscheidungssituationen, d. h. komplexen Mengen von (quantitativ abgestuften) Überzeugungen und Wünschen, und Handlungen. Aber was wir über dieses idealisierte Bild sagen können, gilt a fortiori für realistischer aufgeweichte Bilder. Insbesondere ist das, was ich nun eine Entscheidungsvariable genannt habe, die die möglichen Entscheidungssituationen enthält, aus denen heraus über eine Menge alternativer Handlungen oder Optionen entschieden wird, nichts sonderlich scharf Umrissenes. Das bewusste Sich-vor-Augen-Halten der eigenen Wünsche und Überzeugungen und die explizite praktische Überlegung, die dann in ein zeitlich eng lokalisierbares Fassen einer Intention oder Entscheidung mündet, ist ja eher selten; diese expli-

zite Form der Willensbildung wäre auf Dauer viel zu anstrengend. Was die relevanten Wünsche und Überzeugungen dann sind, ist oft nicht ganz klar, und ab wann eine Intention dann richtig feststeht, ist auch oft nicht klar.

Das heißt aber nicht, dass sie gar nie richtig oder erst im letzten Moment feststeht, wenn die Handlung zur Ausführung kommt und nicht mehr aufgehalten werden kann. Dass Entscheidung und Handlung im Prinzip zeitlich entkoppelt sind, ist mir wichtig. Immer wieder gehe ich abends mit einem Vorsatz ins Bett, um ihn am nächsten Morgen einfach auszuführen. Wo der nächste Urlaub hingeht, muss man mittlerweile ein halbes Jahr vorher entscheiden. Und ich verfasse meinen letzten Willen in der Hoffnung, dass es noch 30 Jahre oder länger dauern möge, bis er zur Ausführung kommt.

Der springende Punkt ist nun dieser: Wenn wir auf diese Weise explizit Handlungen als durch Entscheidungssituationen verursacht betrachten, dann müssen wir auch zulassen, dass solche Komplexe von Wünschen, Überzeugungen und Intentionen auch andere Wirkungen haben als die jeweilige Handlung, also gemeinsame Ursache von Handlung und Anderem sind. Wir erläutern gelegentlich unsere Intention (auch wenn man das vielleicht als weitere Handlung modellieren müsste). Oft sieht man uns unsere Wünsche und Absichten an unserer Mimik und Gestik an. Das scheint mir für den zwischenmenschlichen Verkehr auch wichtig zu sein; extrem kontrollierte Menschen, die uns nur über ihre Handlungen einen Einblick in ihr Innenleben gestatten, sind uns unheimlich. Das viel diskutierte Toxin Puzzle (s. Kavka 1983) setzt genau an dieser Stelle an, indem es fiktiverweise das (durch ein Zerebroskop feststellbare) Fassen einer Intention und nicht erst die Handlung selbst zum Gegenstand einer Belohnung macht.

Wie soll das entscheidende Subjekt mit einer solchen Möglichkeit umgehen? Sind solche Nebenwirkungen der Entscheidung zu berücksichtigen? Ja, unbedingt. Wenn im Toxin-Fall das Fassen einer Intention belohnt wird, so fasse ich die Absicht, sofern die Belohnung die Unannehmlichkeiten der Absichtsausführung überwiegt. Wenn im Raucher-Fall weder das fragliche Gen noch das Rauchen selbst, sondern absurderweise allein der Wunsch zu rauchen den Lungenkrebs begünstigt, so habe ich den Wunsch besser nicht (und rauche mithin auch nicht).

Doch wie kann dieser Umstand in der Modellierung einer Entscheidungssituation Berücksichtigung finden? Diese enthält ja zunächst nur die möglichen Handlungen und all die anderen Dinge oder Propositionen, auf die sich die Wünsche und Überzeugungen des Subjekts be-

ziehen, und eben diese Wünsche und Überzeugungen; sie enthält aber nicht, gleichsam reflexiv, die mögliche Entscheidungssituation selbst als gesonderte Variable. Wie auch? Wie gesagt, die Verursachtheit der eigenen Handlungen ist nicht Bestandteil der praktischen Überlegung.

An dieser Stelle wird die Sache etwas verwickelt. Man kann auch solche reflexiven Entscheidungsmodelle formulieren und ihr Verhältnis zu den bisher betrachteten »normalen« nicht-reflexiven Modellen studieren; dieses Verhältnis ergibt sich daraus, dass reflexives und nicht-reflexives Modell in gewisser Weise dieselbe Entscheidungssituation modellieren (vgl. Spohn 2003, Abschn. 4.3). Die Konsequenz dieser verwickelten Modellierung sollte freilich einleuchten: Im nicht-reflexiven Modell schlagen sich solche Nebenwirkungen der Entscheidungssituation gerade in einer Korrelation dieser Nebenwirkungen mit den Handlungen nieder. Im nicht-reflexiven Modell ist diese Korrelation unerklärlich und darum konnte, wie geschildert, niemand sie in Betracht ziehen. In der reflexiven Erweiterung zeigt sich aber eine gemeinsame Ursache für diese Korrelation, nämlich die (nicht-reflexive) Entscheidungssituation selbst. *In diesem Fall und nur in diesem Fall können aus der Sicht des entscheidenden Subjekts die eigenen Handlungen mit anderen Dingen korreliert sein, ohne für sie (probabilistisch) ursächlich zu sein.*

So biegt sich schließlich der Kreis meiner Überlegungen zu den Abhängigkeitsgleichgewichten zurück. Denn die in ihnen formulierte Korrelation zwischen den Handlungen der beiden Spieler müssen wir jetzt nicht als die falsche Überzeugung interpretieren, mit der eigenen Handlung auf die des anderen Einfluss zu nehmen. Sie kann auch in der geschilderten Weise Ausdruck einer angenommenen gemeinsamen Ursache sein.

Was ist in diesem Fall die gemeinsame Ursache? Wenn die Handlung eines jeden Spielers allein durch seine Entscheidungssituation verursacht ist, diese Entscheidungssituation aber nebenher die Handlungen des anderen beeinflussen soll, so kann die fragliche gemeinsame Ursache nur in der gemeinsamen Herausbildung der Entscheidungssituation der Spieler liegen. Ist das plausibel? Ich denke, ja. Ich sagte ja schon, dass so eine Entscheidungssituation nichts scharf Umrissenes ist. Sie ist etwas zeitlich Ausgedehntes, in dieser Ausdehnung sich Entfaltendes, was genügend Raum für Interaktionen und wechselseitige Abhängigkeiten lässt, die aus jedweder Form von Kommunikation zwischen den Spielern erwachsen. Im ursprünglichen Gefangenen-Dilemma etwa ist es vollkommen plausibel, dass die beiden Gangster nicht erst in den getrennten Verhörtzellen ihre Entscheidungen treffen, wie es von der Ge-

schichte bzw. von der Polizei perfiderweise suggeriert wird, sondern als verschworene Gemeinschaft, die sich im Laufe ihrer gemeinsamen Raubzüge herausgebildet hat, schon längst, im kooperativen Abhängigkeitsgleichgewicht verschränkt, entschieden sind.

Ich will das Bild, das sich auf diese Weise ergibt, in Bezug auf das Gefangenendilemma noch einmal pauschal zusammenfassen: Die Entscheidungssituationen der beiden Spieler können sich über einen gewissen Zeitraum hinweg entwickeln und dabei kausal ineinander verschränken. Das führt dazu, dass die Entscheidungssituation des einen Spielers sowohl für die eigene Handlung ursächlich ist wie auch auf die Entscheidungssituation des anderen Spielers (und damit indirekt auf dessen Handlung) kausalen Einfluss nimmt. Wenn dieser Prozess unter der Bedingung gemeinsamen Wissens steht, so muss er in ein Abhängigkeitsgleichgewicht münden, und zwar rationalerweise in das Pareto-optimale kooperative Gleichgewicht, in dem jeder Spieler glaubt, dass der andere kooperiert, sofern er selber kooperiert, und indem das Kooperieren den bedingt erwarteten Nutzen maximiert.

Natürlich steht es den Spielern frei, die wechselseitige Abhängigkeit zu brechen; auch Nash-Gleichgewichte sind (degenerierte) Abhängigkeitsgleichgewichte. Der Zeitpunkt der eigentlichen Entscheidung kann so spät gelegt sein, dass die kausale Interaktion nicht nur auf der Handlungs-, sondern auch auf der Entscheidungsebene ausgeschlossen ist. Aber manchmal ist es eben vernünftiger, in der vorgegebenen Abhängigkeit zu verharren anstatt aus ihr auszubrechen. Im Gefangenendilemma ist das jedenfalls so. Der Clou der hier nur informell angedeuteten Theoriebildung wäre gerade, für die Wahl der Entscheidungszeitpunkte, für frühe Selbstbindung oder späte Entscheidung, für Abhängigkeit oder Unabhängigkeit noch Rationalitätskriterien zu liefern; in diesem Umkreis liegt auch die Integration von »resolute choice« und »sophisticated choice«, welche McClennen 1990 noch als konkurrierende Entscheidungsregeln erörtert, in einer einheitlichen Rationalitätstheorie.

#### *4. Nachbetrachtung*

Mein Nahziel habe ich damit erreicht; nämlich den Denkfehler, der uns auf Nash-Gleichgewichte festzulegen schien, zu identifizieren, und so der weiteren Begriffsbildung von Abhängigkeitsgleichgewichten Sinn und damit auch Interesse zu verschaffen.

Aber, so wird man sich fragen, ist denn in den über 50 Jahren

intensivsten spieltheoretischen Nachdenkens nicht Ähnliches überlegt worden? Ja und nein. Natürlich gibt es Versuche, die im vorigen Abschnitt beschriebene wechselseitige Abhängigkeit explizit im Rahmen der Standard-Spieltheorie zu modellieren. Insbesondere haben etwa Harsanyi/Selten 1988, S. 4–7 und 18–23, dem Gefangenen-Dilemma ein aus so genannten Selbstbindungszügen bestehendes Spiel vorgeschaltet und gezeigt, dass das Nash-Gleichgewicht des so erweiterten Spiels die Kooperation im Gefangenen-Dilemma einschließt. Und Myerson 1991, S. 249–57, führt vor, wie sich durch eine Theorie der »preplay communication« Aumanns korrelierte Gleichgewichte und die darin liegende Abhängigkeit auf die normalen Nash-Gleichgewichte reduzieren. Das sind höchst aufschlussreiche Überlegungen; und es wäre zu überprüfen, ob sie sich direkt auf die Abhängigkeitsgleichgewichte anwenden lassen.

Die Absicht hinter solchen Überlegungen scheint mir aber verkehrt zu sein. Natürlich kann man versuchen, solche Abhängigkeiten als Ergebnisse spezieller Spiele im Rahmen der Theorie der Nash-Gleichgewichte zu beschreiben. Aber man kann auch umgekehrt die Existenz solcher Abhängigkeiten als gegeben betrachten und angesichts dessen, wie ich es begonnen habe, eine Theorie darüber entwickeln, welches Verhalten in solchen Abhängigkeiten rational ist, und so auch darüber, wie solche Abhängigkeiten rationalerweise zu gestalten sind.

Der Punkt ist dann weniger, welche Theorie die allgemeinere ist; das kann man so oder so herum sehen. Abhängigkeitsgleichgewichte sind offenkundig allgemeiner als Nash-Gleichgewichte; aber man kann auch, wenn das erfolgreich ist, Abhängigkeitsgleichgewichte als Nash-Gleichgewichte in speziellen Spielen darstellen. Der Punkt ist vielmehr die grundlegende Veränderung unserer Rationalitätsvorstellungen für Spielsituationen, die mit der direkten Betrachtung der Abhängigkeitsgleichgewichte einhergeht. Man denkt, im einmaligen Gefangenen-Dilemma wäre die Defektion vollkommen rational und im endlich iterierten Gefangenen-Dilemma lasse sich die Kooperation nur als ein Produkt beschränkter Rationalität verstehen. Dabei ist es genau umgekehrt. Die Kooperation ist im einmaligen und im iterierten Gefangenen-Dilemma vollkommen rational, und die abweichende Defektion ist nur durch mangelndes Vertrauen in die Rationalität des jeweils anderen oder durch mangelndes gemeinsames Wissen um die Rationalität zu erklären. Man muss sich davon freimachen, dass die Standard-Entscheidungs- und -Spieltheorie *definieren*, was rational ist; sie machen dazu nur einen Explikationsvorschlag, der seine Schwächen hat – die gerade im endlich

iterierten Gefangenen-Dilemma besonders drastisch hervortreten.

Falsch ist so auch das insbesondere durch die ökonomische Theoriebildung transportierte ideologische Bild vom frei und unabhängig entscheidenden Individuum, dem aus seinem Eigeninteresse heraus die Rationalität von Kooperation nur schwer und nur unter besonderen Umständen zu vermitteln ist. Dieses Bild wird schon von der Beobachtung unterhöhlt, dass unsere handlungsleitenden Bewertungen sich in komplexer Weise aus egozentrischen und altruistischen Interessen zusammensetzen – man kann sich nur darüber streiten, wie sich die Gewichte verteilen.<sup>7</sup> Unabhängig davon, wie sich die Nutzenfunktionen zusammensetzen, ist das Bild aber auch aufgrund der hier angestellten Überlegungen falsch. Wir stehen immer in zwischenmenschlichen Abhängigkeiten; und individuelle Rationalität kann uns diktieren, diese Abhängigkeiten zur beider- und allseitigen Nutzenmaximierung anzuerkennen.

Ich bin davon überzeugt, dass diese Überlegungen zu den Grundlagen der Spieltheorie ein großes Potential für die politische Philosophie haben. Die immer betonte Möglichkeit einer vielleicht sogar dramatischen Divergenz von individueller und kollektiver Rationalität stellt sich dadurch anders dar. Die Auseinandersetzung zwischen liberalen und kommunitaristischen Idealen erhält dadurch neue Nahrung. Die implizit oder explizit spieltheoretischen Staatsbegründungen von Hobbes bis Rawls und darüber hinaus rücken dadurch in ein neues Licht. Dieses Raunen seriös auszuführen, ist freilich eine andere Sache und keine, die hier noch Platz hätte.

### *Anmerkungen*

- \* Diese Schrift ist Georg Meggle zu seinem 65. Geburtstag gewidmet. Seine immer aufrechte und manchmal kämpferische Art stünde auch ihr gut an; für ihre weitreichenden, hier aber nur angedeuteten sozialphilosophischen Konsequenzen hoffe ich auf seinen Beifall. Drum habe ich sie für diese Festschrift geschrieben. Sehr herzlich bedanke ich mich bei Christoph Fehige für seine hilfreichen Kommentare.
- 1 Dieser Begriff des gemeinsamen Wissens liegt ja auch der handlungstheoretischen Semantik und Kommunikationstheorie von Meggle 1981 zugrunde. Das Meinungsgleichgewicht wird allerdings schon durch zweifache Hochstufung des wechselseitigen Wissens erzwungen, nicht erst durch unendlichfache Hochstufung, durch welche Meggle und andere übereinstimmend das gemeinsame Wissen charakterisieren; vgl. Spohn 1982, S. 253.
- 2 An dieser Stelle ist es doch ratsam, sich unter den reinen Strategien einzelne

Handlungen vorzustellen. Letztere stehen klarerweise in kausalen Beziehungen. Bei komplexen Handlungsplänen ist das zumindest fragwürdig, da sie zeitlich und modal ausgedehnt sind, d. h. sich auf mögliche Situationen einstellen, die zumeist gar nicht akut werden. Freilich hat »Plan« schon die Ambiguität, auf der ich später herumreiten will: das Fassen eines Plans kann ein lokales, in kausalen Beziehungen stehendes Ereignis sein, seine Ausführung, wie gesagt, eher nicht.

- 3 Cf. Spirtes et al. 1993, Abschn. 3.7, Meek/Glymour 1994 und Pearl 2000, Kap. 3 und 4.
- 4 Diese Überlegung scheint der Begründung beiderseitiger Kooperation durch das so genannte »mirror principle« zu ähneln (vgl. Davis 1977). Doch schien mir diese Begründung immer kurzschlüssig zu sein (s. Spohn 2003, S. 250) und erst durch die Abhängigkeitsgleichgewichte hinreichenden Rückhalt zu gewinnen. Außerdem funktioniert das Spiegelprinzip nur, wenn die Lage der beiden Spieler vollkommen symmetrisch ist – eine Beschränkung, der die Theorie der Abhängigkeitsgleichgewichte natürlich nicht unterliegt.
- 5 Harsanyi/Selten 1988 ist ein heroischer Versuch, dieses Auswahlproblem allgemein zu lösen.
- 6 Diese Unverursachtheit der eigenen Handlungen aus der Sicht der praktischen Überlegung liefert meines Erachtens einen basalen Sinn von Handlungs- oder Willensfreiheit; vgl. dazu Spohn i. E..
- 7 Vgl. dazu Fehige 2004 und die darin meines Erachtens überzeugend ausgeführte Begründung einer Apriori-Empathie. Nur seine starken Behauptungen über besagte Gewichtungen wollen mir nicht einleuchten.

### Literatur

- AUMANN 1974. Robert J. Aumann, »Subjectivity and Correlation in Randomized Strategies«, *Journal of Mathematical Economics* 1 (1974), S. 67–96.
- 1987. Robert J. Aumann, »Correlated Equilibrium as an Expression of Bayesian Rationality«, *Econometrica* 55 (1987), S. 1–18.
- BERNHEIM 1984. B. Douglas Bernheim, »Rationalizable Strategic Behavior«, *Econometrica* 52 (1984), S. 1007–28.
- CAMPBELL/SOWDEN 1985. Richmond Campbell und Lanning Sowden (Hrsg.), *Paradoxes of Rationality and Cooperation*, Vancouver, University of British Columbia Press, 1985.
- VAN DAMME 1991. Eric van Damme, *Stability and Perfection of Nash Equilibria*, 2. Aufl., Berlin, Springer, 1991.
- DAVIDSON 1963. Donald Davidson, »Actions, Reasons, and Causes«, *Journal of Philosophy* 60 (1963), S. 685–700.
- DAVIS 1977. Lawrence Davis, »Prisoners, Paradox, and Rationality«, *American Philosophical Quarterly* 114 (1977), S. 319–27.
- FEHIGE 2004. Christoph Fehige, *Soll ich?*, Stuttgart, Reclam, 2004.
- FISHBURN 1964. Peter C. Fishburn, *Decision and Value Theory*, New York, Wiley, 1964.
- HARSANYI/SELTEN 1988. John C. Harsanyi und Reinhard Selten, *A General Theory of Equilibrium Selection in Games*, Cambridge, Mass., MIT Press, 1988.

- HEMPEL 1961/62. Carl Gustav Hempel, »Rational Action«, *Proceedings and Addresses of the APA* 35 (1961/62), S. 5–23.
- KAVKA 1983. Gregory S. Kavka, »The Toxin Puzzle«, *Analysis* 43 (1983), S. 33–6.
- MCCLENNEN 1990. Edward F. McClennen, *Rationality and Dynamic Choice*, Cambridge, Cambridge U. P., 1990.
- MEEK/GLYMOUR 1994. Christopher Meek und Clark Glymour, »Conditioning and Intervening«, *British Journal for the Philosophy of Science* 45 (1994), S. 1001–21.
- MEGGLE 1977. Georg Meggle (Hrsg.), *Analytische Handlungstheorie, Band 1, Handlungsbeschreibungen*, Frankfurt a. M., Suhrkamp, 1977.
- MYERSON 1991. Roger B. Myerson, *Game Theory. Analysis of Conflict*, Cambridge, Mass., Harvard U. P., 1991.
- PEARCE 1984. D. G. Pearce, »Rationalizable Strategic Behavior and the Problem of Perfection«, *Econometrica* 52 (1984), S. 1029–50.
- PEARL 2000. Judea Pearl, *Causality. Models, Reasoning, and Inference*, Cambridge, Cambridge U. P., 2000.
- SAVAGE 1954. Leonard J. Savage, *The Foundations of Statistics*, New York, Wiley, 1954.
- SPIRITES ET AL. 1993. Peter Spirtes, Clark Glymour und Richard Scheines, *Causation, Prediction, and Search*, Berlin, Springer, 1993.
- SPOHN 1976/78. Wolfgang Spohn, *Grundlagen der Entscheidungstheorie*, Dissertation, Universität München 1976, veröff. Kronberg/Ts., Scriptor 1978, vergriffen; pdf-Version unter <http://www.uni-konstanz.de/FuF/Philo/Philosophie/philosophie/files/ge.buch.gesamt.pdf>.
- 1982. Wolfgang Spohn, »How to Make Sense of Game Theory«, in Stegmüller et al. 1982, S. 239–70.
- 1994. Wolfgang Spohn, »On Reichenbach's Principle of the Common Cause«, in Salmon/Wolters 1994, S. 215–39.
- 2003. Wolfgang Spohn, »Dependency Equilibria and the Causal Structure of Decision and Game Situations«, *Homo Oeconomicus* 20 (2003), S. 195–255.
- 2007. Wolfgang Spohn, »Dependency Equilibria«, *Philosophy of Science* 74 (2007), S. ■.
- I. E. Wolfgang Spohn, »Der Kern der Willensfreiheit«, in Sturma i. E.
- STURMA I. E. Dieter Sturma (Hrsg.), *Julian Nida-Rümelin über Vernunft und Freiheit*, Berlin, de Gruyter, im Erscheinen. dito
- STEGMÜLLER ET AL. 1982. Wolfgang Stegmüller, Wolfgang Balzer und Wolfgang Spohn (Hrsg.), *Philosophy of Economics*, Berlin, Springer, 1982.
- SALMON/WOLTERS 1994. Wesley C. Salmon und Gereon Wolters (Hrsg.), *Logic, Language, and the Structure of Scientific Theories*, Pittsburgh, Pittsburgh U. P., 1994.