

Perspektiven der Analytischen Philosophie  
Perspectives in Analytical Philosophy

Herausgegeben von  
Georg Meggle und Julian Nida-Rümelin

Band 2



Walter de Gruyter · Berlin · New York

1994

# Praktische Rationalität

Grundlagenprobleme und ethische  
Anwendungen des *rational choice*-Paradigmas

Herausgegeben von  
Julian Nida-Rümelin



Walter de Gruyter · Berlin · New York

1994

08270578X

Redaktion: Ulla Wessels

Das Buch enthält 31 Abbildungen und 3 Tabellen

Universitäts-  
Bibliothek  
München

© Gedruckt auf säurefreiem Papier,  
das die US-ANSI-Norm über Haltbarkeit erfüllt.

*Die Deutsche Bibliothek – CIP-Einheitsaufnahme*

**Praktische Rationalität** : Grundlagenprobleme und ethische  
Anwendungen des rational choice-Paradigmas / hrsg. von  
Julian Nida-Rümelin. — Berlin , New York : de Gruyter, 1993  
(Perspektiven der analytischen Philosophie ; Bd. 2)  
ISBN 3-11-013656-2  
NE: Nida-Rümelin, Julian [Hrsg.]; GT

© Copyright 1993 by Walter de Gruyter & Co., D-10785 Berlin  
Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung  
außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages  
unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikro-  
verfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Printed in Germany

Satz und Druck: Arthur Collignon GmbH, Berlin  
Buchbinderische Verarbeitung: Lüderitz & Bauer, Berlin  
Einbandentwurf: Rudolf Hüpler, Berlin

K 93 / 22 249

# Inhalt

Vorwort . . . . . V

## Erster Teil

### Grundlagenprobleme der Theorie praktischer Rationalität und des *rational choice*-Paradigmas

#### Kapitel I

~~X~~ Das *rational choice*-Paradigma: Extensionen und Revisionen  
JULIAN NIDA-RÜMELIN . . . . . 3

#### Kapitel II

Handlungserklärung – Handlungsrationalität  
JEAN-LOUIS ARNI . . . . . 31

#### Kapitel III

~~?~~ Entschluß und Beschluß: Zum Begriff der rationalen Entscheidung  
GEORG KOHLER . . . . . 109

#### Kapitel IV

Der besondere Status von Personen:  
Eine Anomalie für die Theorie praktischer Rationalität  
MARTINE NIDA-RÜMELIN . . . . . 143

#### Kapitel V

Der Nutzen von Befriedigung:  
Ein Zirkel in der Entscheidungstheorie  
ANNA KUSSER UND WOLFGANG SPOHN . . . . . 169

#### Kapitel VI

Wie läßt sich die Spieltheorie verstehen?  
WOLFGANG SPOHN . . . . . 197

#### Kapitel VII

~~X~~ Kooperations kollektiver Akteure:  
Zum Informationsgehalt angewandter Spieltheorie  
REINHARD ZINTL . . . . . 239

Zweiter Teil  
Moraltheoretische Anwendungen  
des *rational-choice*-Paradigmas

## Kapitel VIII

Zur rationalen Fundierung einer neo-utilitaristischen  
Wohlfahrtsökonomie

JOHANNES SCHMIDT . . . . . 261

## Kapitel IX

Rationales Entscheiden und moralisches Handeln

PETER KOLLER . . . . . 281

## Kapitel X

Die Verwendbarkeit von Verhandlungsmodellen im Rahmen  
der Ethik

RAINER TRAPP . . . . . 313

## Kapitel XI

Solidarität in einer egoistischen Welt: Eine Simulation

RAINER HEGSELMANN . . . . . 349

## Kapitel XII

Reputation und Unsichtbare Hand

RUDOLF SCHÜBLER . . . . . 391

## Kapitel XIII

Rationale Akteure und ‚natürliche Rechte‘

CHRISTINE CHWASZCZA . . . . . 411

## Kapitel XIV

Eine positionenorientierte Lösung des Liberalen Paradoxes

LUCIAN KERN UND JULIAN NIDA-RÜMELIN . . . . . 437

Zu den Autoren . . . . . 449

Register . . . . . 455

Erster Teil  
Grundlagenprobleme der Theorie praktischer  
Rationalität und des *rational choice*-Paradigmas

## Kapitel I

# Das *rational choice*-Paradigma: Extensionen und Revisionen

JULIAN NIDA-RÜMELIN

1. Das Grundmodell
2. Interaktionen
3. Kollektive Akteure
4. Strukturelle Rationalität
5. *rational choice* und Ethik

### 1. Das Grundmodell

Eine Person handelt rational, wenn ihre Handlungen im Hinblick auf die Ziele dieser Person sinnvoll erscheinen. Handlungen sind im Hinblick auf die Ziele einer Person sinnvoll, wenn sie als ein gutes Mittel gelten können, diese Ziele zu erreichen.<sup>1</sup>

Diese beiden Sätze konstituieren den nucleus einer wirkungsmächtigen Konzeption praktischer Rationalität, die eine Reihe wissenschaftlicher Disziplinen in unterschiedlichem Ausmaße prägt:<sup>2</sup> die theoretische Ökonomie, die politische Theorie, die theoretische Geschichtswissenschaft, die Soziologie u. a. Die Anwendung dieser Rationalitätskonzeption in den empirischen Einzelwissenschaften beruht auf der Annahme, daß das menschliche Verhalten in hinreichendem Maße rational sei, um soziale Phänomene erklären zu können. Die Faszination dieser Rationalitätskonzeption verdankt sich jedoch auch ihrer formalen Präzisierung, wie sie zunächst Ramsey (1931) und später insbesondere Savage (1954) vorgenommen haben.<sup>3</sup>

---

<sup>1</sup> Der philosophischen Analyse des Begriffs der Entscheidung widmet sich der Beitrag von Kohler in diesem Band, Kap. III.

<sup>2</sup> Eine detaillierte Darstellung bietet der nachfolgende Beitrag von Arni in diesem Band, Kap. II.

<sup>3</sup> Zur Einführung in die rationale Entscheidungstheorie eignet sich Stegmüller (1973). Zu den unterschiedlichen entscheidungstheoretischen Modellen vgl. Spohn (1978). Wichtige Beiträge sind in Munier (1987) und Abell (1991) zusammengestellt. Vgl. auch Trapp (1985); Anand (1987); Person (1988).

Der naheliegende Gedanke, daß rationale Personen ihre Handlungen so wählen, daß sie ihre Ziele optimieren, den wir im folgenden als *rational choice*-Paradigma umschreiben, gewinnt seine Attraktivität als Analyseinstrumentarium aber auch aufgrund der unbefriedigenden Erklärungskraft makro-soziologischer Theorien wie des historischen Materialismus und des Funktionalismus.<sup>4</sup> Der methodologische Individualismus in der Sozialwissenschaft ist zwar logisch und systematisch unabhängig vom *rational choice*-Paradigma als Erklärungsansatz, aber theoriegeschichtlich besteht eine enge Verbindung zwischen methodologischem Individualismus und *rational choice*-Paradigma. Die vordergründige Politisierung des Gegensatzes von ‚kritischen‘ makro-soziologischen Theorien historisch-materialistischer Provenienz einerseits, und ‚rechter‘ bzw. markt-orientierter *rational choice*-Konzeptionen andererseits ist dabei schon lange obsolet. Insbesondere Jon Elster (1979, 1983) hat mit seinen Schriften dazu beigetragen, daß wissenschaftsideologische Auseinandersetzungen diese Diskussion kaum noch belasten. Das *rational choice*-Paradigma hat ohne Zweifel Konjunktur.

Das *rational choice*-Paradigma hat seine Wurzeln in der theoretischen Ökonomie und der Statistik – Wurzeln, die es bis heute nicht verleugnen kann und die bei einigen Autoren (z. B. G. Becker) zu einem Rückgriff auf eine individualistisch-egoistische Anthropologie führen, wie es für die ökonomische Theorie bis in die 30er Jahre dieses Jahrhunderts hinein prägend war. Für die egoistische Anthropologie der traditionellen ökonomischen Theorie optimieren rationale Personen nicht irgendwelche subjektiven Ziele, sondern ihr persönliches Wohlergehen oder eben ihren ‚Nutzen‘.<sup>5</sup> Vor dem metatheoretischen Hintergrund des seit Anfang dieses Jahrhunderts erstarkenden logischen Empirismus erschien aber die Fundierung einer empirischen Theorie durch einen Begriff wie dem ‚Nutzen‘ einer Person als problematisch. Den strengen empiristischen Sinnkriterien genügte dieser Begriff nicht und mußte daher als ‚metaphysisch‘ gelten. Die Eliminierung des klassischen Nutzenbegriffes und seine Ersetzung durch die moderne Nutzentheorie wird daher vielfach als bloße begriffliche Präzisierung interpretiert – eine genauere Sichtung zeigt jedoch, daß dies eine irreführende Betrachtungsweise ist: Der Übergang zur modernen Nutzentheorie, wie sie bei Ramsey (1931) initiiert wurde und bei Savage

<sup>4</sup> Vgl. dazu Wiesenthal in Elster (1987).

<sup>5</sup> Zur Problematik des Verhältnisses von Nutzen, Wünschen und Präferenzenerfüllung (Befriedigung) vgl. den Beitrag von Kusser/Spohn in diesem Band, Kap. V. Ein spezielles Problem der Theorie rationaler Entscheidung behandelt M. Nida-Rümelin in diesem Band, Kap. IV.



(1954) in ausgereifter Form vorliegt, ist zugleich der Übergang von einer instrumentell-egoistischen Theorie praktischer Rationalität zu einer Kohärenztheorie praktischer Rationalität. Das Bemühen, die mikro-ökonomische Theorie so weit als möglich unverändert zu lassen, hat zur Verschleierung dieses Sachverhaltes beigetragen. Die Kohärenzaxiome der modernen ökonomischen Theorie machen nicht nur keinen Gebrauch von *egoistischen* Annahmen (die rationale Person optimiert ihre persönlichen Interessen), sondern sind auch unabhängig von einer *instrumentellen* Konzeption praktischer Rationalität.

Sei  $X$  eine Menge von Alternativen, die von den Präferenzen einer rationalen Person in eine Rangfolge gebracht werden, dann verlangt die Kohärenztheorie praktischer Rationalität, daß diese Rangfolge eine *Ordnungsrelation*  $R$  auf  $X$  bildet, d. h., daß die (schwachen) Präferenzen der rationalen Person reflexiv, vollständig und transitiv sind. Während *Reflexivität* für (schwache) Präferenzen trivial ist (jede Alternative ist mindestens so gut wie sie selbst), ist Vollständigkeit und Transitivität, wie empirische Untersuchungen zeigen, in vielen Fällen nicht gegeben.<sup>6</sup> Die Bedingung der *Vollständigkeit* verlangt von einer rationalen Person, daß sie für beliebige Alternativen  $x$ ,  $y$  eine Präferenz hat, d. h., daß sie  $x$  gegenüber  $y$  (schwach) vorzieht oder (einschließend)  $y$  gegenüber  $x$  (schwach) vorzieht. Die Bedingung der *Transitivität* verlangt, daß eine rationale Person, die eine schwache Präferenz für  $x$  gegenüber  $y$  und  $y$  gegenüber  $z$  hat, dann auch eine schwache Präferenz für  $x$  gegenüber  $z$  hat. Während man die Bedingung der Reflexivität als Bedeutungspostulat des Begriffs ‚Präferenz‘ interpretieren kann, liegt es nahe, die Bedingungen der Vollständigkeit und der Transitivität als genuine Rationalitätspostulate zu interpretieren. Demnach wäre es begrifflich nicht ausgeschlossen, daß eine Person intransitive oder unvollständige Präferenzen hat.

Die genannten drei (Ordnungs-)Bedingungen reichen jedoch nicht hin, um vom *qualitativen* Begriff der Präferenz zum *quantitativen* Begriff des Nutzens überzugehen. Dazu ist es erforderlich, die Alternativenmenge  $X$  probabilistisch zu  $X^*$  zu erweitern. Die Präferenzen der Person beziehen sich jetzt nicht mehr nur auf die Alternativen aus  $X$ , sondern auch auf beliebige Wahrscheinlichkeitsverteilungen (Lotterien) über  $X$ . Zunächst muß nun auch bezüglich der probabilistisch erweiterten Alternativenmenge  $X^*$  gefordert werden, daß die Präferenzen einer rationalen Person reflexiv, vollständig und transitiv sind. Zusätzlich werden folgende vier Kohärenzaxiome postuliert:

<sup>6</sup> Vgl. dazu Tversky/Kahneman (1984).

*Reduktion:* Eine rationale Person ist indifferent zwischen zwei Wahrscheinlichkeitsverteilungen über  $X$ , wenn eine in die andere durch Umformungen gemäß dem wahrscheinlichkeitstheoretischen Kalkül überführbar ist.

*Substitution:* Wenn eine rationale Person zwischen einer Wahrscheinlichkeitsverteilung  $x^*$  und einer bestimmten Alternative  $x$  aus  $X$  indifferent ist, dann können  $x^*$  und  $x$  in beliebigen Kontexten ausgetauscht werden, ohne daß sich die Präferenzen ändern.

*Kontinuität:* Wenn  $x_a$  die beste und  $x_z$  die schlechteste Alternative in  $X$  ist, dann gibt es für jede Alternative  $x$  aus  $X$  eine Wahrscheinlichkeit  $p$  für  $x_a$ , so daß die rationale Person zwischen der Wahrscheinlichkeitsverteilung  $[p x_a \ \& \ (1-p) x_z]$  und  $x$  indifferent ist.

*Monotonie:* Von zwei Wahrscheinlichkeitsverteilungen zwischen  $x$  und  $y$  zieht die rationale Person diejenige vor, die eine höhere Wahrscheinlichkeit für die bessere Alternative vorsieht.

Keines dieser Axiome nimmt inhaltlich auf bestimmte Motivationen der handelnden Personen Bezug oder macht Voraussetzungen darüber, was für diese betreffende Person subjektiv wertvoll sei. *Die Plausibilität dieser Axiome ist nicht nur unabhängig von jeglicher inhaltlicher Festlegung, etwa auf ökonomische oder egoistische Ziele, sondern auch unabhängig davon, ob rationale Handlungen subjektive Ziele der handelnden Person optimieren. Zwischen der instrumentell-egoistischen Ausgangstheorie und der Kohärenztheorie praktischer Rationalität scheint keine logische Verbindung zu bestehen.*

*Nun läßt sich beweisen, daß eine Präferenzrelation  $R$ , die die oben aufgeführten Kohärenzaxiome erfüllt, durch eine reellwertige Funktion repräsentiert werden kann:* Es gibt eine reellwertige Funktion  $u$  über  $X^*$ , die bis auf lineare Transformationen eindeutig bestimmt ist und für die gilt:  $\forall x, y \in X^*: u(x) \geq u(y) \leftrightarrow x R y$ .<sup>7</sup> Dieses sog. *Nutzentheorem* ist zunächst nichts anderes als ein harmloses *Metrisierungstheorem*, das den qualitativen Begriff der Präferenz in einen quantitativen Begriff, eine reellwertige Funktion  $u$ , überführt. Die Repräsentierbarkeit von Präferenzen durch eine reellwertige Funktion ist notwendige und hinreichende Bedingung dafür, daß die Kohärenzaxiome von einer gegebenen Präferenzrelation  $R$  erfüllt werden. Wir haben zunächst keinerlei Anlaß zu der Vermutung, daß diese Funktion  $u$  etwas mit den persönlichen Interessen der Person, wie es die traditionelle ökonomische Theorie annahm, oder (allgemeiner) mit den subjektiven Zielen

<sup>7</sup> Vgl. Luce/Raiffa (1957), Kap. 2.

(konsequentialistische Interpretation) der Person zu tun hat bzw. diese repräsentiert.

Wir haben zu Beginn dieses Abschnittes völlig offengelassen, was die Präferenzen einer Person sind oder wie man diese empirisch feststellt. Das sog. *revealed preference*-Modell der rationalen Entscheidungstheorie identifiziert Handlungen mit der manifesten Präferenz für bestimmte Elemente aus  $X^*$  (angesichts der bestehenden Handlungsoptionen). Wenn man nun unter  $X$  eine Menge von Zuständen bzw. möglichen Welten versteht, dann liegt es nahe,  $u$  über  $X$  als Repräsentanz der subjektiven Bewertung von Weltzuständen und  $u$  über  $X^* \setminus X$  als Erwartungswert zu interpretieren, denn es läßt sich zeigen, daß  $u$  bezüglich der Wahrscheinlichkeitsverteilungen linear ist, d. h., daß etwa gilt:  $u(x^*) = p u(x) + (1-p) u(x')$ , wenn  $x^* = [p x \ \& \ (1-p) x']$

Da das Metrisierungstheorem aus den oben nur natursprachlich wiedergegebenen Axiomen logisch-deduktiv ableitbar ist, muß jeder, der die Kohärenzaxiome als adäquate Bedingungen für die Präferenzen einer rationalen Person ansieht, bei dieser Interpretation auch anerkennen, daß das Entscheidungsverhalten einer rationalen Person die so konstituierte reellwertige Funktion  $u$  (die sog. Nutzenfunktion) über  $X$  (bzw.  $X^*$ ) maximiert.

Die konsequentialistische Interpretation des Nutzentheorems besagt, daß eine rationale Person Zustände bewertet — diese Bewertung muß nicht explizit sein, sie kann auch implizit durch das Entscheidungsverhalten der Person zum Ausdruck kommen — und ihre Handlungen jeweils so wählt, daß der Erwartungswert dieser Bewertung über Zustände bei Zuordnung der durch die Handlung konstituierten Wahrscheinlichkeitsverteilung (über Zustände) maximiert wird. Wenn nun noch zusätzlich angenommen wird, daß diese Erwartungswert-Maximierung das *Handlungsmotiv* einer rationalen Person ist, dann ist der Übergang von der Kohärenztheorie praktischer Rationalität zur *konsequentialistischen Optimierungstheorie* perfekt. Aus der ursprünglichen Theorie, *rational sei eine Person, deren Präferenzen kohärent sind, ist eine andere Theorie entstanden, die behauptet, rational ist eine Person, die den Erwartungswert ihrer subjektiven Bewertung von Weltzuständen maximiert*. Daß dieser Übergang nicht logisch zwingend ist, kann man sich anhand folgender Überlegung klarmachen.

Man stelle sich einen idealen Kantischen Akteur vor, der seine Handlungspräferenzen — zumindest soweit es sich nicht um pragmatische, sondern um moralische Fragen handelt — an der Erfüllung des moralischen Gesetzes (des kategorischen Imperativs) orientiert. Man nehme weiter an, daß es sich um einen ‚vernünftigen‘ Kantianer handelt, der zumindest in

Fällen, in denen die Befolgung des kategorischen Imperativs katastrophale Folgen hätte, eine Handlung wählt, die nicht im Einklang mit dem kategorischen Imperativ ist. Für unser Argument ist die Annahme hinreichend, daß der kategorische Imperativ bisweilen Handlungen empfiehlt, deren Folgen nicht erwartungswertmaximierend bezüglich der subjektiven Zustandsbewertungen der handelnden Person sind. Der Kantische Akteur wird sich bisweilen für Handlungen entscheiden, deren Folgen (auch bei Berücksichtigung moralischer Weitungen) nicht optimal sind, bzw. (im probabilistischen Fall) deren Erwartungswert nicht maximal ist. Können die Präferenzen des so charakterisierten Kantischen Akteurs die Bedingungen der Kohärenztheorie praktischer Rationalität erfüllen?

*Wenn  $X$  die Menge der Weltzustände ist und die Präferenzen über Handlungen identifiziert werden mit Präferenzen über die den Handlungen zuzuordnenden Wahrscheinlichkeitsverteilungen, dann ist das Entscheidungsverhalten des Kantischen Akteurs nicht im Einklang mit den Kohärenzbedingungen.* Dies kann man sich folgendermaßen klar machen: Eine Handlung  $h$  und eine Handlung  $h'$  können die gleiche Wahrscheinlichkeitsverteilung über Weltzustände mit sich führen, und dennoch mag  $h$  durch den kategorischen Imperativ erlaubt und  $h'$  verboten sein. Die Einordnung einer Handlung in die Präferenzrelation des Kantischen Akteurs erfolgt eben nicht nur aufgrund der durch die Handlungen induzierten Wahrscheinlichkeitsverteilungen über Folgen (bzw. über Weltzustände), sondern eben auch aufgrund von Konformität bzw. Defektivität gegenüber einem Kriterium (dem kategorischen Imperativ).

Diese Inkompatibilität von Kohärenztheorie und Kantischer Rationalität ist aber aufgehoben, wenn man eine Uminterpretation der Alternativenmenge vornimmt. Da für die Kohärenztheorie praktischer Rationalität die Kohärenz von Handlungspräferenzen grundlegend ist, (die die simultane Zuordnung von Glaubens- und Wünschbarkeitsgraden erlaubt), liegt es nahe, zunächst die Kohärenz der Handlungspräferenzen des Kantischen Akteurs zu prüfen. Die zu ordnenden Alternativen betreffen also Handlungsoptionen. *Die Präferenzen des Kantischen Akteurs über Handlungsoptionen erfüllen aber idealiter die Bedingungen der Kohärenztheorie praktischer Rationalität.* In einem ersten Schritt kann man  $X$  mit einer Menge von Handlungsoptionen identifizieren und  $X^*$  mit Lotterien über solche Optionen: Je nach Umstand vollziehe ich mit einer Handlung  $h$  eine Handlung  $h'$  oder, falls ein anderer Umstand eintritt, eine Handlung  $h''$  etc. Die Präferenzen des Kantischen Akteurs über eine so interpretierte Alternativenmenge  $X^*$  sind kohärent qua konsequentialistischer Abschwächung – auch ohne Rückgriff auf Stufenfunktionen oder transfinite Werte – durch eine simultane

Zuordnung entsprechender Glaubens- und Wünschbarkeitsfunktionen repräsentierbar.<sup>8</sup> Allerdings ist der Übergang vom Kohärenz- zum Optimierungsmodell hier ausgeschlossen. Die beiden entscheidenden Schritte dieses Übergangs sind nicht realisierbar: *Erstens lassen sich den Handlungen nicht Wahrscheinlichkeitsverteilungen über Weltzustände im üblichen Sinne zuordnen, und zweitens läßt sich u als quantitative Darstellung der Handlungspräferenzen nicht als Repräsentation der Handlungsmotive interpretieren. Die Zuordnung quantitativer Werte zu den Elementen aus  $X^*$  repräsentiert nur die kohärente Präferenzrelation über Handlungen und repräsentiert nicht die subjektive Wünschbarkeit von Weltzuständen im üblichen Sinne.* Ich habe von „Weltzuständen im üblichen Sinne“ gesprochen, da es natürlich möglich ist, den Handlungen ‚formal‘ Wahrscheinlichkeitsverteilungen über Weltzustände zuzuordnen, sofern die Beschreibungen dieser Weltzustände Merkmale enthalten wie „Die Handlung der Person erfüllte den kategorischen Imperativ“. Mit dieser Modifikation läßt sich dann allerdings das Kohärenzmodell nicht mehr in das Optimierungsmodell im konsequentialistischen Sinne überführen, denn dann sind es nicht Zustände, die optimiert werden, vielmehr erhalten bestimmte Merkmale der Handlung selbst einen intrinsischen Status.

Die Erweiterung des Anwendungsbereiches des *rational choice*-Modells vom ursprünglichen egoistischen Paradigma der ökonomischen Klassiker über den subjektivistischen Konsequentialismus zur Kohärenztheorie beinhaltet eine grundlegende Revision der Konzeption praktischer Rationalität. Die konsequentialistische und die kohärentistische Konzeption praktischer Rationalität sind logisch voneinander unabhängig, auch wenn in der ökonomischen Literatur das Nutzentheorem fälschlicherweise bisweilen als Beleg ihrer Äquivalenz ausgegeben wird.

## 2. Interaktionen

Die Theorie rationaler Entscheidung befaßt sich neben Risiko- und Unsicherheitssituationen auch mit *Interaktionssituationen*. Da dieser Zweig der Theorie rationaler Entscheidung historisch aus der mathematischen Analyse von Gesellschaftsspielen hervorgegangen ist, wird er bis heute – irreführend – als ‚Spieltheorie‘ bezeichnet.<sup>9</sup> Spieltheorie und Entscheidungstheorie i. e. S. befassen sich mit der *individuellen rationalen Entscheidung*.

<sup>8</sup> Dies wird näher ausgeführt in Nida-Rümelin (1993), § 51.

<sup>9</sup> Eine kanonische Einführung bietet nach wie vor Luce/Raiffa (1957). Vgl. auch Holler/ Illing (1991).

Entscheidungstheoretisch i. e. S. wird die rationale Entscheidung relativ zu der gegebenen subjektiven Bewertungsfunktion der Konsequenzen und der subjektiven bedingten Wahrscheinlichkeitsfunktion der Umstände bestimmt. Von welcher Art die Umstände sind, ist dabei zunächst nicht wesentlich. In einer Interaktionssituation sind die Umstände die Entscheidungen anderer Personen (und evtl. zusätzlich Zufallsereignisse). Die Ausweitung des *rational choice*-Paradigmas von entscheidungstheoretischen Optimierungsproblemen („Spiele gegen die Natur“) zu spieltheoretischen Optimierungsproblemen scheint daher zumindest für solche Interaktionssituationen, in denen die Entscheidungen der beteiligten Personen unabhängig voneinander erfolgen (nicht-kooperative Spiele) unproblematisch zu sein.<sup>10</sup>

Tatsächlich tauchen jedoch bei dieser Ausweitung des *rational choice*-Paradigmas dadurch Probleme auf, daß die handelnde Person davon ausgehen muß, daß die Interaktionspartner ebenfalls rational entscheiden, so daß die wechselseitigen Erwartungen darüber, was der anderen Person als rational erscheint, über die Rationalität einer Entscheidung mit bestimmen. Im Gegensatz zur Entscheidungstheorie i. e. S., für die die auf Umstände gerichteten Erwartungen nicht Gegenstand der Theorie sind, werden nun diese Erwartungen, soweit sie sich auf die Entscheidungen der Interaktionspartner beziehen, insofern Gegenstand der Theorie, als diese auf wechselseitiger Rationalitätserwartung beruhen – was aber in einer spezifischen Interaktionssituation jeweils rational ist, soll wiederum im Rahmen der Spieltheorie beantwortet werden. Die Spieltheorie versucht, dieser Problematik dadurch gerecht zu werden, daß sie nur solche Entscheidungen als rational auszeichnet, die Bestandteile von *Gleichgewichtspunkten der Interaktionsstruktur* sind.

Ein Gleichgewichtspunkt ist eine Kombination von Entscheidungen (bzw. Strategien), für die gilt, daß kein Beteiligter ein Interesse daran hat, von seiner Entscheidung abzurücken, wenn alle übrigen bei ihrer Entscheidung bleiben. Man kann diese Voraussetzung auch als Konsequenz der *potentiellen Publizität spieltheoretischer Rationalitätsempfehlungen* interpretieren:<sup>11</sup> Nur solche Rationalitätsempfehlungen werden zugelassen, deren allgemeine Bekanntheit kein Grund sein kann, von ihnen abzuweichen. Mit Hilfe des Begriffs des gemeinsamen Wissens<sup>12</sup> kann man diese Fassung der Entscheidungsrationalität in Interaktionssituationen auch folgender-

<sup>10</sup> Zu den Anwendungsproblemen der Spieltheorie vgl. Zintl in diesem Band, Kap. VII.

<sup>11</sup> Vgl. den Beitrag von Spohn in diesem Band, Kap. VI.

<sup>12</sup> Die Konzeption gemeinsamen Wissens geht auf Lewis (1969) zurück; vgl. auch Heal (1978) und Bicchieri (1989).

maßen charakterisieren: *Eine Entscheidung ist in einer Interaktionssituation rational, wenn sie unter der Annahme gemeinsamen Wissens erwartungswertmaximierend ist.* Das gemeinsame Wissen bezieht sich auf die Struktur der Interaktionssituation (1), repräsentiert durch die mathematische Beschreibung des Spiels (in Normalform oder in extensiver Form), die Annahme der Rationalität der Beteiligten im Sinne der Nutzenerwartungswertmaximierung (2) und die (spieltheoretischen) Rationalitätsempfehlungen für die betreffende Interaktionssituation (3). Gemeinsames Wissen bezüglich dieser drei Propositionen heißt: Jede beteiligte Person weiß (1), (2), (3), jede beteiligte Person weiß, daß jede beteiligte Person (1), (2), (3) weiß, jede beteiligte Person weiß, daß jede beteiligte Person weiß, daß jede beteiligte Person (1), (2), (3) weiß, usw. Lewis spezifiziert den Begriff des gemeinsamen Wissens auf einen Sachverhalt S, von dem gilt, daß jede beteiligte Person Grund hat zu glauben, daß S jedem Beteiligten der Interaktionssituation anzeigt, daß jede beteiligte Person Grund hat zu glauben, daß S, usw. S wird damit zur Basis des gemeinsamen Wissens. Bei Koordination durch Abmachungen wäre dieser Sachverhalt etwa die getroffene Abmachung. In anderen Fällen erscheint es aber als künstlich, einen Sachverhalt zu konstruieren, der Grund des gemeinsamen Wissens ist. Gemeinsames Wissen beruht häufig auf Regularitäten, Normen und Institutionen, die die Isolierung eines Sachverhaltes als Grund gemeinsamen Wissens als abwegig erscheinen lassen.

Die Kombination des *rational choice*-Paradigmas mit der Annahme gemeinsamen Wissens bezüglich der drei genannten Propositionen bereitet jedoch Probleme. Man betrachte etwa den Fall einer Interaktionssituation mit einem einzigen Gleichgewichtspunkt in gemischten Strategien. Es handele sich etwa um ein  $2 \times 2$ -Spiel, A habe die Alternativen  $a_1$  und  $a_2$ , B habe die Alternativen  $b_1$  und  $b_2$ , der Gleichgewichtspunkt sei: A wähle mit 30% Wahrscheinlichkeit  $a_1$  und mit 70% Wahrscheinlichkeit  $a_2$  und B wähle mit 60% Wahrscheinlichkeit  $b_1$  und mit 40% Wahrscheinlichkeit  $b_2$ . Die Rationalitätsempfehlung fordert also A resp. B auf, sich jeweils für eine bestimmte Wahrscheinlichkeitsverteilung über  $a_1$  und  $a_2$  bzw.  $b_1$  und  $b_2$  zu entscheiden. Weder A noch B dürfen sich für  $a_1$  oder  $a_2$  bzw.  $b_1$  oder  $b_2$  entscheiden. Die geforderte Entscheidung könnte etwa so realisiert werden, daß A und B jeweils weiße und rote Kugeln im Mischungsverhältnis 3 zu 7 bzw. 6 zu 4 in eine Urne geben und dann nach guter Durchmischung je nach zufällig herausgegriffener Kugelfarbe  $a_1$  oder  $a_2$  bzw.  $b_1$  oder  $b_2$  realisieren. *Die eigentliche Entscheidung besteht in der Wahl eines bestimmten Mischungsverhältnisses der Kugeln in der Urne.* Das im ersten Abschnitt dargestellte Kontinuitätsaxiom der Kohärenztheorie praktischer

Rationalität schließt jedoch aus, daß eine rationale Person eine Wahrscheinlichkeitsmischung von zwei Konsequenzen ( $a_1$  und  $a_2$  bzw.  $b_1$  und  $b_2$ ) jeweils beiden Konsequenzen zugleich (strikt) vorzieht. *Unabhängig von den Erwartungen, die die Interaktionsbeteiligten A und B bezüglich der Entscheidungen der jeweils anderen Person haben, kann daher die Entscheidung für die als Gleichgewichtsstrategie empfohlene Wahrscheinlichkeitsmischung nicht als rational (i. S. der Nutzenerwartungswertoptimierung) gelten.* Es gibt nur eine Erwartung, die die Wahl der als Gleichgewichtspunkt empfohlenen gemischten Strategie *nicht irrational* macht: Dies ist die Erwartung, daß die jeweilige andere Person sich für die vom Gleichgewichtspunkt bestimmte Wahrscheinlichkeitsmischung entscheidet. Aber für diesen besonderen Fall gilt, daß die Wahl *jeder* Entscheidung, sei es die Entscheidung für eine der reinen Strategien ( $a_1$  und  $a_2$  bzw.  $b_1$  und  $b_2$ ) oder für eine der gemischten Strategien (Wahrscheinlichkeitsverteilungen über  $a_1$  und  $a_2$  bzw.  $b_1$  und  $b_2$ ) gleichermaßen rational sind. *Das gemeinsame Wissen der Rationalität der Interaktionsbeteiligten legitimiert also in keiner Weise die Erwartung, daß sich die Interaktionsbeteiligten an die Handlungsempfehlung des Gleichgewichtspunktes halten.* Die Empfehlung „Wähle bei Vorliegen eines einzigen Gleichgewichtspunktes die Gleichgewichtsstrategie“ läßt sich also aus der Optimierungskonzeption praktischer Rationalität nicht herleiten.

Dieser Befund ergibt sich aber nicht nur als ein allgemeines Defizit von Gleichgewichtspunkten aus gemischten Strategien, sondern verschärft sich angesichts spezifischer Interaktionsstrukturen.\*

Man betrachte etwa folgende Interaktionsstruktur:

		B	
		$b_1$	$b_2$
A	$a_1$	2/2	3/3
	$a_2$	3/3	1/1

Aufgrund der Annahme gemeinsamen Wissens der Interaktionsbeteiligten A und B wissen bei dieser Konstellation beide, daß ihre jeweiligen Rationalitätserwägungen wegen der symmetrischen Struktur des Spiels bezüglich  $a_1$  und  $a_2$  bzw.  $b_1$  und  $b_2$  zum gleichen Ergebnis führen müssen. D. h., wenn für A  $a_1$  rational ist, muß auch für B  $b_1$  rational sein und umgekehrt, gleiches gilt für  $a_2$  und  $b_2$ . Sollte eine gemischte Strategie, also eine Wahrscheinlichkeitsverteilung über  $a_1$  und  $a_2$  für A rational sein, mit der Wahrscheinlichkeit  $p$  für  $a_1$ , dann müßte die gemischte Strategie mit  $p$  für



$b_1$  (und  $1-p$  für  $b_2$ ) für B rational sein (und umgekehrt). Üblicherweise wird die Gleichgewichtsorientierung der Spieltheorie als Folge der Annahme gemeinsamen Wissens dargestellt. In diesem Beispiel zeigt sich jedoch, daß die Annahme gemeinsamen Wissens bezüglich der Struktur des Spieles in Konflikt mit der Gleichgewichtstheorie gerät. Man kann sich dies folgendermaßen klarmachen. Die beiden Gleichgewichtspunkte sind  $a_2, b_1$  und  $a_1, b_2$ . Beide Gleichgewichtspunkte sind aber aufgrund der Symmetrie der Interaktionssituation nicht erreichbar, ohne mit der Prämisse gemeinsamen Wissens bezüglich Interaktionsstruktur, Rationalität der Beteiligten und Kriterien in Konflikt zu geraten. Gleiches gilt für gemischte Strategien; denn wäre eine gemischte Strategie mit der Wahrscheinlichkeit  $p$  für  $a_1$  eine Gleichgewichtsstrategie für A, so wäre aufgrund der Symmetrie der Interaktionsstruktur auch die gemischte Strategie mit  $p$  für  $b_1$  eine Gleichgewichtsstrategie für B. Alle Strategien mit gleicher Wahrscheinlichkeit für  $a_1$  bzw.  $b_1$  sind jedoch, wie sofort ersichtlich, keine Gleichgewichtsstrategien dieser Interaktionsstruktur.

Bei allen Interaktionsstrukturen, in denen die Rationalität der individuellen Entscheidungen einer Person davon abhängt, welche Entscheidungen von den anderen Interaktionsbeteiligten erwartet werden können, von diesen aber wiederum mindestens eine Entscheidung nur in Abhängigkeit von der Erwartung der eigenen Handlungswahl getroffen werden kann, ergeben sich Probleme für die Verbindung von optimierender Rationalität und der Annahme gemeinsamen Wissens. Als Beispiel betrachte man etwa folgende Interaktionsstruktur:

		B	
		$b_1$	$b_2$
A	$a_1$	2/2	4/4
	$a_2$	3/3	1/1

Wenn A erwartet, daß B sich für  $b_1$  entscheidet, dann wird sich A für  $a_2$  entscheiden. Wenn A hingegen erwartet, daß B sich für  $b_2$  entscheidet, dann wird A sich für  $a_1$  entscheiden. Ebenso gilt für B, daß  $b_2$  nur unter der Bedingung optimierend ist, wenn B erwarten kann, daß sich A für  $a_1$  entscheidet, und ebenso ist  $b_1$  nur optimierend, wenn B erwarten kann, daß sich A für  $a_2$  entscheidet. Unabhängig von den Erwartungen über die Entscheidungen der anderen Personen läßt sich jedoch über die Rationalität einer dieser Strategiewahlen nichts aussagen. Man kann dies auch so

formulieren: *Unter der Annahme gemeinsamen Wissens ist die Rationalität in Interaktionsstrukturen dieser Art unterbestimmt.* Das gilt in diesem Falle auch bei Berücksichtigung gemischter Strategien. Erwartet A, daß sich B für die gemischte Strategie „75% Wahrscheinlichkeit für  $b_1$  und 25% Wahrscheinlichkeit für  $b_2$  (0,75  $b_1$  & 0,25  $b_2$ )“ entscheidet, dann ist A zwischen  $a_1$  und  $a_2$  indifferent, beide Entscheidungen liefern den gleichen Nutzenerwartungswert, ebenso alle Wahrscheinlichkeitsverteilungen über  $a_1$  und  $a_2$ . Erwartet A, daß sich B für eine gemischte Strategie mit einer für  $b$ -geringeren Wahrscheinlichkeit als 75% entscheidet, dann ist  $a_1$  für A optimierend, während  $a_2$  für den Fall einer größeren Wahrscheinlichkeit für  $b_1$  optimierend wäre. Auch unter Einschluß gemischter Strategien gilt also, daß eine Rationalitätsempfehlung unter der Annahme gemeinsamer Wissens nicht gegeben werden kann: *Rationale Nutzenmaximierer können hier diejenigen Erwartungen, die die Rationalität ihrer Strategien bestimmen würden, nicht haben.*

Die in dem letzten Beispiel deutlich gewordene Unterbestimmtheit praktischer Rationalität ist deshalb irritierend, weil das ursprüngliche, hier auf Interaktionssituationen ausgeweitete *rational choice-Paradigma durch seine charakteristische Subjektivierung praktischer Rationalität auf subjektive Wahrscheinlichkeiten und subjektive Wünschbarkeiten eine derartige Unterbestimmtheit ausschließt.*

In einer Interaktionssituation von der Art des vieldiskutierten Gefangenen-Dilemmas ist dagegen Rationalität im Sinne des ursprünglichen *rational choice-Paradigmas* eindeutig bestimmt; aber sie führt zu einem Ergebnis, das an der Adäquatheit dieser Konzeption zweifeln läßt:

		B	
		$b_1$	$b_2$
A	$a_1$	3/3	1/4
	$a_2$	4/1	2/2

In dieser Interaktionsstruktur ist für beide Personen jeweils die zweite Entscheidung ( $a_2$  resp.  $b_2$ ) dominant: Wie auch immer sich die andere Person entscheidet, für beide denkbaren Umstände der je individuellen Handlung ist  $a_2$  bzw.  $b_2$  optimierend.

Wenn man von der Annahme ausgeht, daß eine Theorie praktischer Rationalität die Aufgabe hat, Handlungsempfehlungen zu formulieren, wird die Fragwürdigkeit des *rational choice-Paradigmas* für Interaktionssi-

tuationen von der Art des Gefangenen-Dilemmas besonders deutlich. Man stelle sich vor, A und B wendeten sich ratsuchend an zwei Entscheidungstheoretiker  $T_1$  und  $T_2$ . Aufgrund der vollständigen Symmetrie der Situation können A und B erwarten, daß  $T_1$  und  $T_2$  jeweils beiden Interaktionsbeteiligten dieselbe Empfehlung geben, also entweder A und B  $a_1$  bzw.  $b_1$  empfehlen oder A und B  $a_2$  bzw.  $b_2$  empfehlen. Wenn nun  $T_1$   $a_1$  und  $b_1$  empfiehlt, während  $T_2$   $a_2$  und  $b_2$  empfiehlt, steht zweifelsfrei fest, welchem Rat zu folgen für A und B günstiger ist: Es ist für beide Interaktionsbeteiligten jeweils günstiger, dem Rat von  $T_1$  zu folgen. Der dem optimierenden Grundmodell des *rational choice*-Paradigmas verpflichtete Entscheidungstheoretiker  $T_2$  empfiehlt  $a_2$  und  $b_2$  als die dominanten Strategien (und Gleichgewichtspunkt) der Interaktionssituation. Als Ratgeber ist  $T_2$  in Interaktionssituationen dieser Art nicht erfolgreich.

Nun kann natürlich eingewandt werden, daß diese vermeintliche Erfolglosigkeit nur darauf beruht, daß zwei Funktionen praxisbezogener, ratgebender Kriterien vermengt würden: Ein Kriterium kann seine ratgebende Funktion im Hinblick auf die Präferenzen einer spezifischen Person wahrnehmen (individuelle Rationalität) oder es kann Empfehlungen für das Verhalten einer Gruppe von Personen geben (kollektive Rationalität). Die spieltheoretische Empfehlung, in Gefangenen-Dilemma-Situationen die dominante und damit nicht-kooperative Entscheidung zu wählen, sei individuell rational, wenn auch nicht kollektiv rational, denn der Zustand, der erreicht wird, wenn beide Personen individuell rational in Interaktionssituationen vom Typ des Gefangenen-Dilemmas interagieren, ist für beide schlechter als ein Zustand, der erreicht worden wäre, wenn sie die andere Entscheidungsalternative gewählt hätten (der Gleichgewichtspunkt in Gefangenen-Dilemma-Situationen ist nicht pareto-optimal).

Dieser Rettungsversuch leidet an zwei Schwächen, die bei genauerer Betrachtung miteinander eng verbunden sind: Zum einen kann die Empfehlung in Interaktionssituationen vom Typ des Gefangenen-Dilemmas die dominante Strategie als Bestandteil des einzigen Gleichgewichtspunktes des Spieles zu wählen, deshalb nicht gut durch den Verweis auf den individuellen Charakter dieser Empfehlung gerechtfertigt werden, weil die Auszeichnung von Gleichgewichtspunkten bei der rationalen Entscheidung in Interaktionssituationen, wie wir oben gesehen haben, erst durch die Annahme gemeinsamen Wissens, also einer kollektiven Komponente legitimiert ist. Diese Art der Rechtfertigung nicht-kooperativer Strategiewahl in Interaktionssituationen vom Typ des Gefangenen-Dilemmas macht daher die methodischen Grundlagen der Spieltheorie insgesamt zunichte und wirft das *rational choice*-Paradigma auf das im ersten Abschnitt

dargestellte Grundmodell rationaler Entscheidung („Spiele gegen die Natur“) zurück. Zum anderen muß auch eine Theorie individueller Rationalität – sofern sie nur normativ und nicht deskriptiv verstanden wird – bestimmte minimale Regeln der Universalisierbarkeit beachten (das Problem der Abtrennbarkeit von Sollenssätzen). Es ist fraglich, ob die generelle Empfehlung in Interaktionssituationen vom Typ des Gefangenen-Dilemmas, die nicht-kooperative Strategie zu wählen, diesen Regeln genügt.

Schon in den 60er Jahren wurden empirische Untersuchungen darüber angestellt, wie sich Personen in Interaktionssituationen vom Typ des Gefangenen-Dilemmas verhalten.<sup>13</sup> Dabei stellte sich heraus, daß Personen auch dann, wenn sie über die Struktur der Interaktionssituation informiert sind und wissen, daß es sich um eine einmalige Interaktion handelt („One-shot-PD-game“), sich zu einem erstaunlich hohen Prozentsatz kooperativ, also nicht individuell optimierend, verhalten. Wenn es auch nicht die einzig mögliche ist, liegt doch folgende Interpretation eines kooperativen Verhaltens in Interaktionssituationen vom Typ des Gefangenen-Dilemmas nahe: Die Person kooperiert, weil sie weiß, daß beidseitig kooperatives Verhalten kollektiv rational ist. Nur unter sehr speziellen Annahmen ließe sich die Vermutung anstellen, daß altruistische Motive eine Rolle spielen. (Durch entsprechende Variationen der Auszahlungen ließe sich sogar empirisch eruieren, ob altruistische Motive dabei eine Rolle spielen oder nicht.) Zumindest scheint es intuitiv in keiner Weise abwegig zu sein anzunehmen, daß eine Person sich bewußt für die kooperative Strategie entscheidet, obwohl ihre subjektiven Bewertungen der vier möglichen Ergebnisse einer vorgegebenen Interaktionsstruktur dem Typ des Gefangenen-Dilemmas entsprechen. Um eine *petitio principii* auszuschließen, darf das Gefangenen-Dilemma nicht über die Rangordnung der Entscheidungskombinationen, sondern ausschließlich über die Rangordnung der Konsequenzen der Entscheidungskombinationen definiert werden. In der ursprünglichen Geschichte, die zum Namen „Gefangenen-Dilemma“ geführt hat, waren die Konsequenzen Gefängnisjahre für die beiden Interaktionsbeteiligten. Derjenige, der nicht gestand, erhielt etwa zwei Jahre Gefängnis, wenn der andere ebenfalls nicht gestand, während er als Nicht-Geständiger besonders hart bestraft wurde, wenn der andere gestand und damit beide ihres Verbrechens überführt werden konnten, etwa 12 Jahre. Derjenige, der gestand, wurde als Kronzeuge der Anklage freigesprochen, wenn der andere nicht gestand, und erhielt 10 Jahre, wenn der andere

---

<sup>13</sup> Vgl. Rapoport/Guyer/Gordon (1976) sowie Kern/Räder (1988).

ebenfalls gestand. „Gestehen“ war hier die nicht-kooperative Strategie, im folgenden mit „n“ abgekürzt, und „nicht gestehen“ war die kooperative Strategie, im folgenden mit „k“ abgekürzt. Die Konsequenzenmatrix (Konsequenzen in Gefängnisjahren) sieht also folgendermaßen aus:

		B	
		n	k
A	n	10/10	0/12
	k	12/0	2/2

*gestehen nicht = kooperativ*

Wir sagten, die Struktur des Gefangenen-Dilemmas definiere sich über die je subjektiven Präferenzen über Konsequenzen. Insofern handelt es sich bei dieser Interaktionsstruktur mit den gegebenen Konsequenzen (in Gefängnisjahren) nur dann um eine Interaktionsstruktur vom Typ des Gefangenen-Dilemmas, wenn A es auch dann vorzieht, freizukommen, wenn B dadurch 10 Jahre länger sitzen muß, und wenn A es auch dann vorzieht, seine Gefängniszeit um zwei Jahre zu verkürzen (von 12 auf 10), wenn dadurch die Gefängniszeit von B von 0 auf 10 Jahre zunimmt. Mit  $C_A$  für die A betreffenden Konsequenzen und  $C_B$  für die B betreffenden Konsequenzen ergibt sich also folgende, eine Interaktionsstruktur vom Typ des Gefangenen-Dilemmas definierende Präferenzrelation:

A:  $C_A(n, k) / C_B(n, k)$   
 $C_A(k, k) / C_B(k, k)$   
 $C_A(n, n) / C_B(n, n)$   
 $C_A(k, n) / C_B(k, n)$

B:  $C_A(k, n) / C_B(k, n)$   
 $C_A(k, k) / C_B(k, k)$   
 $C_A(n, n) / C_B(n, n)$   
 $C_A(n, k) / C_B(n, k)$

Diese umständliche Notation wurde bewußt gewählt, um zu vermeiden, daß die Wahl der symbolischen Darstellung eine theoretische Vorentscheidung für das *revealed preference*-Konzept, also die Annahme, daß sich die Präferenzen im Entscheidungsverhalten offenbaren, beinhaltet. Wenn wir uns bei der Darstellung der Präferenzen bei A auf die linke und bei B auf die rechte Spalte, also auf die Konsequenzen für A bzw. für B, beschränkt

hätten, dann wäre damit schon implizit die Annahme ausschließlich eigen-orientierter Präferenzen eingeflossen.

- A:  $C_A(n, k)$   
 $C_A(k, k)$   
 $C_A(n, n)$   
 $C_A(k, n)$
- B:  $C_B(k, n)$   
 $C_B(k, k)$   
 $C_B(n, n)$   
 $C_B(n, k)$

Wenn wir die Rangfolge der Strategienkombinationen zur Charakterisierung einer Interaktionssituation vom Typ des Gefangenen-Dilemmas herangezogen hätten, dann wäre damit eine Vorentscheidung für die individuell-optimierende Rationalitätskonzeption getroffen worden.

- A: n, k  
k, k  
n, n  
k, n
- B: k, n  
k, k  
n, n  
n, k

Wir betrachten nun erneut die Person, die in Kenntnis der Interaktionsstruktur, in der sie sich befindet, und in Kenntnis der C-Präferenzen (der Präferenzen, die sich auf die Konsequenzen der Interaktion beziehen), sich für die kooperative Strategie entscheidet. Als Handlungsgrund (als Rechtfertigung dieser Entscheidung) gelte dieser Person die Tatsache, daß beidseitige Kooperation beidseitiger Nicht-Kooperation vorzuziehen ist, was vor dem Hintergrund einer symmetrischen Interaktionssituation für die Wahl der kooperativen Entscheidung spreche. Die Präferenzen einer solchen Person lassen sich konsistent offensichtlich nicht durch die zuletzt angegebene Präferenzstruktur wiedergeben. Denn wenn wir annehmen, daß alle wesentlichen Informationen der vier Alternativen durch die jeweilige Strategienkombination gegeben sind, dann hieße dies, daß etwa A sich bei der Alternative n, k und k, k für k, k entscheidet, was der

angegebenen Präferenzordnung zuwiderläuft. Die Präferenzen einer solchen Person wären also in der Tat inkonsistent.

*Die Absicht zu kooperieren kann einen guten Handlungsgrund darstellen. Es wäre jedoch völlig abwegig, kooperative Handlungsgründe als Umwertung von Konsequenzen zu interpretieren.*<sup>14</sup> Ein rationaler Akteur kann sich auch dann für eine kooperative Strategie in Interaktionssituationen vom Typus des Gefangenen-Dilemmas entscheiden, wenn er die Konsequenzen, die Nicht-Kooperation für eine gegebene Entscheidung des anderen Interaktionsbeteiligten haben, jeweils vorzieht. In diesem Beispiel heißt das: Es ist möglich, daß sich eine rationale Person für die kooperative Strategie entscheidet (nicht gestehen), auch wenn sie das Ergebnis 0 Gefängnisjahre für sich selbst und 12 Gefängnisjahre für den anderen dem Ergebnis 2 Gefängnisjahre für sich selbst und 2 Gefängnisjahre für den anderen ebenso vorzieht wie 10 Gefängnisjahre für sich selbst und 10 Gefängnisjahre für den anderen dem Ergebnis 12 Gefängnisjahre für sich selbst und 0 Gefängnisjahre für den anderen. Ja, noch mehr: Eine Person kann aus logischen Gründen kooperative Handlungsabsichten nur haben, wenn sie keine Umwertung der Konsequenzen vornimmt, die die Gefangenen-Dilemma-Struktur der Interaktionssituation aufheben. Zugespitzt formuliert: Die entscheidungstheoretische Orthodoxie des *revealed preference*-Konzeptes macht Kooperation als Handlungsgrund logisch unmöglich, so daß man (aus logischen Gründen) gezwungen ist, Kooperation als handlungsgrund auszuschließen oder das *revealed preference*-Konzept der Entscheidungstheorie aufzugeben. Da Kooperation als Handlungsgrund unbezweifelbar eine wichtige Rolle für unser alltägliches Verhalten spielt und nicht anzunehmen ist, daß dieses Verhalten in toto jeweils irrational ist, haben wir damit ein starkes Argument für die Inadäquatheit des *revealed preference*-Konzeptes der entscheidungstheoretischen Orthodoxie.

### 3. Kollektive Akteure

Das Problem der Rationalität von Gruppen-Entscheidungen (Kollektiventscheidungen) ist Gegenstand einer eigenständigen Subdisziplin der Entscheidungstheorie: der Logik kollektiver Entscheidungen (collective

---

<sup>14</sup> Diese Interpretation entspräche dem Ansatz der Metapräferenzen, wie sie von Sen (1974) in die Diskussion um das Gefangenen-Dilemma eingeführt wurde. Ich kritisiere diesen Ansatz am Beispiel des Gefangenen-Dilemmas in Nida-Rümelin (1991 b).

choice, public choice, social choice).<sup>15</sup> Unabhängig von den konkreten Bedingungen kollektiven Handelns (implizite und explizite Regeln der Entscheidungsfindung, institutionelle Rahmenbedingungen, individuelle Rechte etc.) lassen sich kollektive Entscheidungen als Aggregationsfunktion darstellen: Aufgrund der individuellen Präferenzrelationen der Referenz-Gruppe wird aufgrund einer Aggregationsfunktion eine kollektive Präferenzrelation gebildet. Es hat sich rasch herausgestellt, daß die Konstitution eines kollektiven Akteurs qua Aggregation unerwartet große Schwierigkeiten macht. Wenn für kollektive Akteure die gleichen Bedingungen für die Kohärenz der Präferenzen angelegt werden, wie bei individuellen Akteuren (vgl. Abschnitt 1), dann ist eine Aggregation mit vernünftigen Eigenschaften ausgeschlossen.<sup>16</sup> Wenn man beliebige individuelle Präferenzrelationen zuläßt – eingeschränkt nur durch die Kohärenzbedingungen des Abschnitts 1 – dann läßt sich zeigen, daß *die einzige Aggregationsregel, die sowohl die Irrelevanz- wie die Pareto bedingung erfüllt, diktatorisch* ist, d. h., daß der kollektive Akteur durch die Präferenzen einer einzigen Person festgelegt ist.<sup>17</sup> Die Irrelevanzbedingung verlangt, daß eine kollektive Rangfolge zweier Alternativen nicht allein dadurch verändert werden darf, daß eine dritte Alternative hinzutritt, *ceteris paribus*. Die Pareto bedingung verlangt, daß Alternativen, die von allen Personen mindestens schwach vorgezogen werden, auch kollektiv mindestens schwach vorgezogen werden.

In der Literatur gibt es eine umfangreiche Diskussion darüber, wie durch Abschwächung einer oder mehrerer dieser Bedingungen doch noch ein befriedigendes Resultat erreicht werden kann. Die Ergebnisse sind ernüchternd. Eine Abschwächung der Bedingungen an die Kohärenz der kollektiven Präferenzen auf bloße Azyklizität etwa führt unter den gleichen Bedingungen zwar nicht mehr zur Ein-Personen-Diktatur, aber es sind nun ausschließlich Aggregationsfunktionen möglich, die *einer Gruppe ein*

<sup>15</sup> Als die beiden Standardwerke zur Logik kollektiver Entscheidungen können nach wie vor Arrow (1951) und Sen (1970) gelten; eine Übersicht bietet Sen (1986), vgl. auch Mueller (1979), Rothkirch (1981) und Kern/Nida-Rümelin (1994).

<sup>16</sup> Obwohl die logische Analyse kollektiver Entscheidungen bis ins 18. Jahrhundert zurückreicht (Condorcet), kann die Schrift von Arrow (1951) als Ursprung der Logik kollektiver Entscheidungen gelten. In Kap. 5 dieser Schrift wird das sog. ‚Possibility-Theorem‘ entwickelt, das eigentlich ‚Impossibility-Theorem‘ heißen müßte, da es eine vernünftigen Bedingungen folgende Aggregation *prima facie* ausschließt.

<sup>17</sup> Zur Problematik der Interpretation von Aggregationsfunktionen vgl. Nida-Rümelin (1986, 1987, 1991 a) und Tanguiane (1992), der für eine Art Rehabilitation diktatorischer Entscheidungsregeln im Sinne demokratischer Repräsentation argumentiert.



*Vetorecht* zugestehen, was einer Art *Gruppendiktatur* gleichkommt (Vetogruppen-Theorem).<sup>18</sup>

Wie problematisch die Ausweitung des *rational choice*-Paradigmas auf kollektive Akteure ist, zeigt mit besonderer Schärfe das Liberale Paradoxon:<sup>19</sup> Die Minimalbedingung kollektiver Rationalität, die Paretoinklusivität der Aggregationsfunktion, kommt für bestimmte Präferenzstrukturen in Konflikt mit der Minimalbedingung individueller Entscheidungsfreiheit. Die Forderung, daß jedes Individuum über mindestens eine Alternative frei entscheiden kann, und die Forderung, daß Alternativen, die von allen Individuen (schwach) vorgezogen werden, auch von der Aggregationsfunktion vorgezogen werden, kann zu zyklischen Präferenzen führen.

Von besonderer Brisanz für das *rational choice*-Paradigma ist das Theorem der Strategie- und Manipulationsanfälligkeit kollektiver Entscheidungsverfahren.<sup>20</sup> Das *rational choice*-Paradigma geht von konsequentialistisch optimierenden Einzelindividuen aus. Folgerichtig muß sich jede Theorie kollektiver Rationalität die Frage stellen, ob ihre Entscheidungsregeln damit vereinbar sind, daß Individuen je individuell optimierend handeln. Das Theorem der Strategieanfälligkeit besagt nun, daß individuell optimierende Rationalität dazu führt, daß die Individuen für beliebige Aggregationsfunktionen jeweils Grund haben, ihre Präferenzen zu verfälschen. Tatsächlich gibt es nur eine Aggregationsregel, die strategiefrei ist, d. h., bei der kein Individuum einen Anreiz hat, seine Präferenzen verfälscht in den Aggregationsprozeß einzuspeisen: die Diktatur. Ein analoges Ergebnis ergibt sich für die Manipulationsanfälligkeit kollektiver Entscheidungsprozesse. Alle Aggregationsfunktionen bis auf die diktatorischen sind für bestimmte Präferenzstrukturen manipulationsanfällig.

Diese ausschließlich negativen Resultate der Logik kollektiver Entscheidungen dürfen nicht zu voreiligen Schlußfolgerungen verleiten. Die logische Analyse kollektiver Entscheidungsprozesse ist damit in keiner Weise obsolet geworden, im Gegenteil, sie bietet ein analytisch fruchtbares Instrumentarium, um die hohe Komplexität realer kollektiver Entscheidungsprozesse zu erfassen. Allerdings erschüttern diese Ergebnisse die Vorstel-

<sup>18</sup> Vgl. Kern/Nida-Rümelin (1994), Kap. 4.

<sup>19</sup> Das Liberale Paradoxon wurde von Sen (1970), Kap. 6 entdeckt und hat seitdem eine verzweigte Diskussion ausgelöst, deren hervorstechende Markierungspunkte der libertäre Vorschlag, individuelle Freiheitsrechte aus der Aggregation herauszunehmen von Nozick (1973, 1974), das Theorem von Gibbard (1974) und die Kritik des sog. ‚welfarism‘ von Sen (1979) sind; vgl. Breyer (1978), Krüger/Gärtner (1983), Wriglesworth (1985), Riley (1986), Mezzetti (1987), Barry (1987) und Kern/Nida-Rümelin in diesem Band, Kap. XIV.

<sup>20</sup> Zum Problem der Strategie- und Manipulationsanfälligkeit vgl. Gibbard (1973), Satterthwaite (1975), Schmeidler/Sonnenschein (1978).

lung, daß sich gesellschaftliches Handeln aus je individuell optimierendem Handeln im Sinne des ursprünglichen *rational choice*-Paradigmas ergibt und daß kollektives Handeln einen optimierenden kollektiven Akteur konstituiert.

#### 4. Strukturelle Rationalität

Die Ausweitung des *rational choice*-Paradigmas auf Interaktionssituationen und kollektive Entscheidungen bringt Probleme mit sich, die man als Probleme der Koordination im weiten Sinne interpretieren kann. Koordinationsprobleme im engeren Sinne liegen dort vor, wo die unterschiedlichen Handlungsoptionen jeweils vergleichbar gute Ergebnisse zeitigen und es eine bloße Frage der Koordination ist, welche dieser Handlungsoptionen gewählt wird. *Interpersonelle* Koordination im weiten Sinne liegt immer dann vor, wenn es darum geht, die individuellen Handlungsoptionen aufeinander abzustimmen. *Intrapersoneller* Koordination geht es dagegen um die Abstimmung der Handlungsoptionen einer einzigen Person.

So wie je individuelle Optimierung in Situationen vom Typ des Gefangendilemmas dazu führt, daß die beteiligten Personen schlechter gestellt werden, als es der Fall wäre, wenn sie je individuell nicht optimierend gehandelt hätten, so ist die Gestaltung eines guten Lebens – intrapersonell betrachtet – davon abhängig, daß die handelnde Person nicht in jeder Einzelsituation optimierend handelt, sondern sich selbst einschränkende Bedingungen auferlegt, die ihrem Leben erst Struktur und Kohärenz geben.<sup>21</sup> Die Vorstellung, daß das notwendige Mindestmaß kollektiver Rationalität allein durch Sanktionen sichergestellt werden könnte, ist sicherlich abwegig. Diese Form der Harmonisierung optimierender Rationalität und kollektiver Rationalität (man könnte es das ‚Hobbes’sche Programm‘ nennen) ist zumindest für den Bereich der Alltagsinteraktionen nur sehr eingeschränkt anwendbar. Neben Sanktionen spielen ohne Zweifel Tugenden bzw., moderner gesprochen, erworbene Verhaltensdispositionen für die Abschwächung des Koordinationsproblems optimierender Rationalität eine große Rolle. Während die Methode der Sanktionierung struk-

<sup>21</sup> Eine interessante, letztlich wohl aus einem puritanisch geprägten Ethos herrührende Kritik der an Optimierung ausgerichteten Rationalitätskonzeptionen formuliert Slotte (1989). Einige Aporien optimierender Rationalität schildert anschaulich und unter Bezug auf soziologische und biologische Forschungsergebnisse Elster (1979, 1983, 1987); vgl. auch Campell/Sowden (1985). Vgl. auch die interessante Konzeption der ‚bounded rationality‘ von Simon (1983).

turell rationalen Verhaltens spieltheoretisch der Veränderung der Auszahlungen entspricht, schränken Dispositionen die Klasse der möglichen Strategien ein. Die vielfältigen empirischen Phänomene der Selbstbindung weisen darauf hin, daß Personen auch (mehr oder weniger bewußt) sich selbst Sanktionen auferlegen, um Handlungsoptionen auszuschließen, von denen sie wissen, daß ihre Wahl auch dann, wenn sie je punktuell optimierend wäre, insgesamt, d. h. unter dem Aspekt der intrapersonellen Koordination betrachtet, schädlich wäre.

Neben gesellschaftlicher oder juridischer Sanktionierung (Sanktionierung durch Ethos oder Legislation), dispositionellen Prägungen und Akten der Selbstbindung gibt es jedoch auch das empirische Phänomen der intentionalen Regelbefolgung. Mit intentionaler Regelbefolgung ist folgendes gemeint: Es gibt Situationen, in denen eine rationale Person einer Regel folgt, obwohl sie bei Nichtbefolgung keine Sanktionen zu befürchten hätte (und ihr dieses auch bewußt ist) und obwohl sie keinerlei dispositive Prägung erfahren hat, die sie dazu veranlassen könnte, dieser Regel zu folgen. Eine solche Regelbefolgung ist intentional, wenn die Handlungswahl in der Absicht getroffen wird, dieser Regel zu folgen. Die Rede von „Regeln“ in diesem Zusammenhang mag das Mißverständnis nahelegen, es handle sich um explizite, erlassene oder moralisch etablierte Regeln. „Regel“ ist dabei in einem umfassenden Sinne gemeint. So wäre es z. B. eine Form von Regelbefolgung, in bestimmten Varianten des Gefangenendilemmas die kollektiv rationale Handlung zu wählen. Da es sich bei inter- oder intrapersoneller Koordination darum handelt, jeweils eine Option zu wählen, die sich in eine gewünschte *Struktur von Handlungen* oder Strategien einfügt, kann man hier auch von *struktureller Rationalität* sprechen. Die Tatsache, daß der Verweis auf eine z. B. moralisch befürwortete Regel einen guten Grund abgibt, eine Handlung zu wählen, die mit dieser Regel konform ist, verweist darauf, daß hier die Rede von „Rationalität“ sinnvoll ist: *Rational ist eine Handlung dann, wenn es einen guten Grund gibt, diese Handlung zu wählen*. Strukturelle Rationalität käme etwa auch dann zum Ausdruck, wenn eine Person sich entscheidet, ihre tatsächlichen Präferenzen in ein Aggregationsverfahren einzuspeisen, obwohl sie weiß, daß nur die verfälschte Wiedergabe dieser Präferenzen die von ihr verfolgten Ziele optimieren würde.

Koordination ist ein guter Handlungsgrund, der danach verlangt, sich individuell und punktuell so zu verhalten, daß die einzelnen Handlungen sich in eine befürwortete Struktur individuellen bzw. gesellschaftlichen Handelns einfügen. Die Identifikation von praktischer Rationalität mit (punktuellem) Folgenoptimierung würde dagegen nur zwei Erklärungs-

modelle für das empirische Phänomen koordinativen und kooperativen Handelns zulassen: Das Modell des Leviathan und das Modell der dispositionellen Prägung bzw. des homo sociologicus. Das *Modell des Leviathan erklärt koordinatives Verhalten durch manifeste oder verborgene Sanktionen* (im Falle, daß es sich um Sanktionen in foro interno handelt, wäre es angemessener, statt vom Modell des Leviathan vom Modell Freud zu sprechen), und das *Modell der dispositionellen Prägung erklärt koordinatives Verhalten durch erworbene Dispositionen, die die Handlungsfreiheit der Personen beschränken* (in der soziologischen Sprechweise würde man von Rollenmustern reden). Die Lücke, die auch eine geschickte Kombination beider Erklärungsmodelle läßt, verweist auf ein Defizit in der Theorie praktischer Rationalität: Auch genuin koordinatives Verhalten kann rational sein. Das Motiv, deshalb in einer bestimmten Weise zu handeln, weil diese Handlung sich in die gewünschte Struktur gesellschaftlichen oder individuellen Handelns einfügt, kann ein rationales Motiv sein. Eine entsprechende Handlung kann wohlbegründet sein: sie wäre strukturell rational. *Strukturell rationales Handeln ist mit der Kohärenz-Variante des rational choice-Paradigmas kompatibel – es ist allerdings nicht kompatibel mit der konsequentialistischen Standardinterpretation des rational choice-Paradigmas.*

### 5. Rational choice und Ethik

Das (konsequentialistisch) optimierende Rationalitätskonzept überträgt sich auf die ethische Theorie, wenn man von der Annahme ausgeht, moralisches Handeln sei idealiter rational. Natürlich gibt es die Möglichkeit, den Anwendungsbereich des *rational choice*-Konzeptes bereichsspezifisch einzuschränken und (in kantischer Sprechweise) nur pragmatische und technische Imperative diesem Konzept zu unterstellen, während moralische Imperative anderen Kriterien, etwa solchen der Universalisierbarkeit, zu genügen hätten.<sup>22</sup> Naheliegender aber ist es, die Versöhnung von rationalitätstheoretischer und ethischer Betrachtung über eine entsprechende Modifizierung der zu optimierenden Wertfunktion sicherzustellen. Moralisches und egoistisches Handeln würde sich dann nicht an der Frage der (konsequentialistischen) Optimierung scheiden, sondern an der Gestalt der Wertfunktion. Rationales moralisches Handeln würde im Gegensatz zu rationalem, aber nicht moralischem Handeln eine unparteiische, im

<sup>22</sup> Einen solchen Vorschlag unterbreitet Suppes (1973); zu seiner Kritik vgl. Nida-Rümelin (1987), § 10.

Grenzfall interpersonell invariante Wertfunktion optimieren. Harsanyi<sup>23</sup> ist diesen Weg mit besonderer Stringenz gegangen. Man kann diesen Ansatz wegen seiner ethischen Fortführung der Bayes'schen Entscheidungs rationalität auch als ‚ethischen Bayesianismus‘ bezeichnen.<sup>24</sup>

Das moralisch Richtige als Optimierung einer gemeinsamen Wertfunktion zu bestimmen, hat eine große intuitive Attraktivität: Auf diese Weise kann man an einer (konsequentialistisch) optimierenden Rationalitätskonzeption festhalten und die inhaltlichen moralischen Fragen auf die Diskussion der angemessenen Werte verlagern. Tatsächlich zeigt jedoch eine genauere Betrachtung, daß die Koordinations- und Kooperationsprobleme selbst bei der starken Annahme interpersonell gleicher Wertfunktionen fortbestehen und sich für interpersonell variante Wertfunktionen sogar noch verschärfen.<sup>25</sup>

Eine ethische Theorie muß sich auch daran messen, welche Eigenschaften eine andere Gesellschaft von Personen hätte, die dieser Theorie in idealer Weise genügen. In einer Gesellschaft konsequentialistisch optimierender Personen steht die Gestaltung des eigenen Lebens sowohl äußeren wie inneren Ansprüchen gegenüber: Die *äußeren Ansprüche* ergeben sich aus der Forderung, alle abschätzbaren Handlungsfolgen zu berücksichtigen, und das schließt die Wirkung eigenen Handelns für das Handeln anderer Personen ein. Diese äußeren Ansprüche machen eine vernünftige Lebensplanung unter normalen Umständen unmöglich, da sie die Verfolgung persönlicher Ziele und Projekte sowie die Aufrechterhaltung persönlicher Bindungen unterminieren. Die *inneren Ansprüche* konsequentialistisch rationalen Handelns sprengen in Gestalt punktueller Optimierung die Strukturen der eigenen Lebensgestaltung. Die beiden Orientierungspunkte des ethischen Diskurses, der eines *guten Lebens* und der einer *wohlgeordneten Gesellschaft*, sind in einer komplexen Weise miteinander verknüpft, der der konsequentialistische Ansatz nicht gerecht wird.

Die struktursprengende Wirkung optimierender Rationalität ist nicht nur ein Problem der gesellschaftlichen Kooperation, sondern auch der individuellen Lebensgestaltung. Die individuelle Lebensform ist weitgehend von den Strukturen bestimmt, in die sich individuelles und punktuelles Handeln einfügt. Diese in vielen Fällen selbst gewählten oder zumindest beeinflussbaren Strukturen bestimmen die Qualität der indivi-

<sup>23</sup> Harsanyi hat die reinste Form einer rationalen Ethik dieses Typs in verschiedenen Varianten seit 1953 entwickelt und in einer Reihe von Aufsätzen dargelegt (vgl. Harsanyi 1953; 1955; 1958; 1976; 1977 a; 1977 b; 1978; 1979; 1980; 1985).

<sup>24</sup> Vgl. Stegmüller (1977), Nida-Rümelin (1992) und Schmidt in diesem Band, Kap. VIII.

<sup>25</sup> Vgl. dazu Nida-Rümelin (1993), Kap. 11–14.

duellen Lebensform ebenso wie die der Gesellschaftform als ganzes, die man sich als eine Vernetzung individueller Lebensformen durch unterschiedliche Formen der Interaktion denken kann. Zwischen den Strukturen des Handelns gibt es wiederum eine strukturelle Beziehung: Die einzelnen Strukturen fügen sich zu Strukturen größerer Allgemeinheit und schließlich vernetzen sich individuelle Lebensformen zu einer Gesellschaftsform. Wenn man die einzelne Person nicht als Konglomerat von Augenblicksempfindungen, Wahrnehmungen und Reaktionen begreift<sup>26</sup>, dann ist sie Träger der intrapersonellen Strukturierung ihres (je punktuellen) Handelns und der Koordination dieser individuellen Lebensform mit den Lebensformen anderer Personen. Die moralische Person befindet sich, bildlich gesprochen, an der Nahtstelle zweier unterschiedlicher Ebenen der Handlungskoordination.

Die Strukturen des Handelns sind weder intra- noch interpersonell (durch Erziehung, Sanktionen, Konventionen, anthropologische Determinanten etc.) in toto vorgegeben, sondern zu einem guten Teil Gegenstand freier Entscheidung — sie können geformt, entwickelt, verändert oder bewahrt werden. Es stellt sich daher eine doppelte Aufgabe: Kriterien für ethisch wünschbare Strukturen und Kriterien für die Einfügung der individuellen Handlung in diese Strukturen zu entwickeln — die konsequentialistische Theorie praktischer Rationalität und die auf ihr beruhenden konsequentialistischen Ethiken werden dieser doppelten Aufgabe nicht gerecht. Das *rational choice*-Paradigma bedarf daher einer grundlegenden Revision im Sinne struktureller Rationalität. Diese Revision betrifft seine konsequentialistische Standardinterpretation — die Kohärenzvariante des *rational choice*-Paradigmas ist davon, wie oben dargestellt, nicht betroffen.

### *Literatur*

- Abell, P. (Hg.): *Rational Choice Theory*. Cambridge 1991.
- Anand, P.: Are the preference axioms really rational? *Theory and Decision* 23 (1987), S. 189–214.
- Arrow, K. J.: *Social Choice and Individual Values*. New York 1951, 2. Aufl. 1963.
- Barry, B.: Lady Chatterley's Lover and Doctor Fisher's Bomb Party: Liberalism, pareto optimality, and the problem of objectionable preferences. In: *Foundations of Social Choice Theory*. Hg. von J. Elster/A. Hylland. Cambridge u. a. 1987.
- Becker, G. S.: *The Economic Approach to Human Behavior*. Chicago 1976.
- Bicchieri, C.: Self-refuting theories of strategic interaction: a paradox of common knowledge. *Kenntnis* 30 (1989), S. 69–85.

---

<sup>26</sup> Vgl. die Diskussionsbeiträge in Elster (1986).

- Breyer, F.: *Das Liberale Paradox. Eine wohlfahrts- und spieltheoretische Untersuchung*. Meisenheim 1978.
- Campbell, R./Sowden, L. (Hgg.): *Paradoxes of Rationality and Cooperation*. Vancouver 1985.
- Elster, J.: *Ulysses and the Sirens. Studies in Rationality and Irrationality*. Cambridge u. a. 1979.
- Elster, J.: *Sour Grapes. Studies in the Subversion of Rationality*. Cambridge 1983.
- Elster, J. (Hg.): *The Multiple Self*. Cambridge 1986.
- Elster, J.: *Subversion der Rationalität*. Frankfurt a. M., New York 1987.
- Gibbard, A.: Manipulations and voting schemes: A general result. *Econometrica* 41 (1973), S. 587–601.
- Gibbard, A.: A pareto-consistent libertarian vlam. *Journal of Economic Theory* 7 (1974), S. 388–410.
- Harsanyi, J. C.: Cardinal utility in welfare economics and the theory of risk-taking. *Journal of Political Economy* 61 (1953), S. 434–435. (Auch in Harsanyi (1976)).
- Harsanyi, J. C.: Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 62 (1955), S. 309–321. (Auch in Harsanyi 1976)).
- Harsanyi, J. C.: Ethics in terms of hypothetical imperatives. *Mind* 67 (1958), 305–316. (Auch in Harsanyi 1976).
- Harsanyi, J. C.: *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht 1976.
- Harsanyi, J. C.: Rule utilitarianism and decision theory. *Erkenntnis* 11 (1977 a), S. 25–53.
- Harsanyi, J. C.: Morality and the theory of rational behavior. *Social Research* 44 (1977), S. 623–656. (= 1977 b).
- Harsanyi, J. C.: Bayesian decision theory and utilitarian ethics. *American Economic Review, Papers and Proceedings* 68 (1978), S. 223–228.
- Harsanyi, J. C.: Rule utilitarianism, and Arrow's impossibility theorem. *Theory and Decision* 11 (1979), S. 289–317.
- Harsanyi, J. C.: Rule utilitarianism, rights, obligations, and the theory of rational behavior. *Theory and Decision* 12 (1980), S. 115–133.
- Harsanyi, J. C.: Rule utilitarianism, equality and justice. *Social Philosophy and Policy* 2 (1985).
- Heal, J.: Common knowledge. *Philosophical Quarterly* 28 (1978), S. 116–131.
- Holler, M. J./Illing, G.: *Einführung in die Spieltheorie*. Berlin, Heidelberg u. a. 1991.
- Kern, L./Räder, H.-G.: Private goods and public bads. In: *Theory and Decision – in Honour of W. Leinfellner*. Hrsg. von G. L. Eberlein/H. Berghel. Dordrecht 1988.
- Kern, L./Nida-Rümelin, J.: *Logik kollektiver Entscheidungen*. Wien, München 1994.
- Krüger, L./Gärtner, W.: Alternative claims and Sen's paradox. *Theory and Decision* 15 (1983), S. 211–229.
- Lewis, D. K.: *Convention: A Philosophical Study*. Cambridge, Mass. 1969.
- Luce, R. D./Raiffa, H.: *Games and Decisions*. London, New York u. a. 1957.
- Mezzetti, C.: Paretian efficiency, Rawlsian justice and the Nozick theory of rights. *Social Choice and Welfare* 4 (1987), S. 25–37.
- Mueller, D.: *Public Choice*. Cambridge u. a. 1979.
- Munier, B. (Hg.): *Risk, Decision and Rationality*. Dordrecht 1987.

- Nida-Rümelin, J.: Wir-Intention und die Logik kollektiver Entscheidungen. In: *Analyse und Kritik* 8 (1986), S. 96–108.
- Nida-Rümelin, J.: *Entscheidungstheorie und Ethik* (1987).
- Nida-Rümelin, J.: Zur Philosophie der Demokratie: Arrow-Theorem, Liberalität und strukturelle Normen. *Analyse und Kritik* 13 (1991), S. 184–203. (= 1991 a).
- Nida-Rümelin, J.: Practical reason or metapreferences? An undogmatic defense of kantian morality. *Theory and Decision* 30 (1991), S. 133–162. (= 1991 b).
- Nida-Rümelin, J.: Rationale Ethik. In: *Geschichte der neueren Ethik*. Hrsg. von A. Pieper, Bd. 2. Tübingen 1992, S. 154–172.
- Nida-Rümelin, J.: *Kritik des Konsequentialismus*. Wien, München 1993.
- Nozick, R.: Distributive justice. *Philosophy and Public Affairs* 3 (1973), S. 45–126.
- Nozick, R.: *Anarchy, State and Utopia*. Oxford 1974; dt. Übers.: *Anarchie, Staat, Utopia*. München o. J. (1976).
- Person, I.: Rationality and maximization of satisfaction. *Nous* 22 (1988), S. 537–554.
- Ramsey, F. P.: Truth and probability. In: ders.: *Foundations of Mathematics and Other Logical Essays*. London 1931.
- Rapoport, A./Guyer, M. J./Gordon, D. G.: *The 2 × 2 Game*. Ann Arbor 1976.
- Riley, J.: Generalized social welfare functionals: Welfarism, morality and liberty. *Social Choice and Welfare* 3 (1986), S. 233–254.
- Rothkirch, C. von: *Rationales Handeln im öffentlichen Interesse*. Frankfurt a. M., New York 1981.
- Satterthwaite, M.: Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10 (1975), S. 187–217.
- Savage, L.: *The Foundations of Statistics*. New York 1954, 2. Auflage 1972.
- Schmeidler, D./Sonnenschein, H.: Two proofs of the Gibbard-Satterthwaite theorem on the possibility of a strategy proof social choice function. In: *Decision Theory and Social Ethics*. Hg. von H. W. Gottinger/W. Leinfellner. Dordrecht 1978.
- Sen, A. K.: *Collective Choice and Social Welfare*. San Francisco u. a. 1970.
- Sen, A. K.: Choice, ordering and morality. In: *Practical Reason*. Hrsg. von S. Körner. Oxford 1974, S. 54–67.
- Sen, A. K.: Personal utilities and public judgements. Or what is wrong with welfare economics? *The Economic Journal* 89 (1979), S. 537–558.
- Sen, A. K.: Social choice theory. In: *Handbook of Mathematical Economics*. Bd. III. Hg. von K. Arrow/M. Intriligator. Amsterdam u. a. 1986.
- Simon, H. A.: *Models of Bounded Rationality*. 2 Bde. Cambridge 1983.
- Slote, M.: *Beyond Optimizing. A Study of Rational Choice*. Cambridge, London 1989.
- Spohn, W.: *Grundlagen der Entscheidungstheorie*. Kronberg/Ts. 1978.
- Stegmüller, W.: *Rationale Entscheidungstheorie (Entscheidungslogik)*. Berlin, Heidelberg, New York 1973 (= *Studienausgabe Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Bd. 4: *Personelle und statistische Wahrscheinlichkeit*, Teil I).
- Stegmüller, W.: On the interrelations between ethics and other fields of philosophy and science. *Erkenntnis* 11 (1977), S. 55–80.
- Suppes, P.: The concept of obligation in the context of decision theory. In: *Science, Decision and Value. Proceedings of the 5th Univ. of Western Ontario Philos. Coll. 1969*. Hrsg. von J. Leach/R. Butts/G. Pearce. Dordrecht 1973.



- Tanguiane, A.: *Aggregation and Representation of Preferences. Introduction to Mathematical Theory of Democracy*. Berlin u. a. 1992.
- Trapp, R.: Utility theory and preference logic. *Erkenntnis* 22 (1985), S. 301–339.
- Tversky, A./Kahneman, D.: Choices, values and frames. *American Psychologist* 39 (1984), S. 341–350.
- Wiesenthal, H.: Die Ratlosigkeit des homo oeconomicus. In: Elster (1987), S. 7–20.
- Wriglesworth, J. L.: *Libertarian Conflicts in Social Choice*. Cambridge 1985.

## Register

- „Aggresso“-Spieltyp 337 ff.  
Absicht 111 f., 115 f., 120 f.  
Adverse Selektion 397 ff.  
Akrasia 116, 120 f., 132 f.  
Anerkennung 122 ff., 129 ff.  
Anonymität 392  
Anrechtsspiele 249  
Antiindividualismus 188 (Fußn.)  
Äquitätsprinzip 243  
Aufteilungsnormen 243  
Aufteilungsspiele 249 f.
- Bargaining 254, 289 f., 295, 304 ff.  
Bayes'sche Rationalität 265 ff.  
Bayesianismus 199 ff., 232 f.  
Befriedigung und Nutzen 178, 183, 185 f.  
Behaviorismus (behavioristisch) 78 f.  
Beschluß 109 ff., 136 ff.  
Brown-Robinson-Prozeß 228 ff.
- Chicken Game 426, 427 (Fußn.)
- Deliberation 46 ff.  
Diktatur 20 f.  
Direkte Bezugnahme 151 ff.  
Drohkoalition 337  
Drohpotential 326 ff.
- Egoismus 391  
Egoist, rationaler 350  
Eigenorientierung 444  
Entscheidung 109  
    unter Unsicherheit 207, 221  
Entscheidungsbaum 170 (Fußn.), 171, 180  
Entscheidungsregel 208 f. (s. a. Maximinregel, Prinzip von der Maximierung des erwarteten Nutzens)  
Entscheidungssituationen, kausale Struktur von 171 f., 176 f., 178
- Entscheidungstheorie 170, 172 (Fußn.), 173, 180, 183, 198 ff., 222, 235  
    auf komputationeller und semantischer Ebene 173, 234 f.  
    Zirkel in der 173 f., 188 f.
- Entschluß 109 ff., 131 ff.  
Epistemische Rationalität 223 ff., 228  
Ereignis-Kausalität 36 ff.  
Erwartungswertmaximierung 6 ff.  
Erweiterte Nutzenfunktion 272  
Erweiterte Präferenzordnung 272  
Essentielle Eigenschaften 156 f.  
Ethische Präferenz 273  
Evolutionärer Ansatz 395  
ex-ante-facto-Erwartungswert 145 ff.  
ex-post-facto-Beurteilung 143 ff.  
ex-post-facto-Erwartungswert 145 ff.
- Fairneß 292  
Fehlentscheidung 143 ff.  
Fiktives Spielen, s. Brown-Robinson-Prozeß  
Frage, praktische 111  
Friedenssicherung 256
- Gefangenendilemma (prisoner's dilemma) 14 ff., 137 f., 231 (Fußn.), 241, 248 ff., 288 f., 295 ff., 299 ff., 314, 324, 338, 340, 342, 352 ff., 392 ff., 415, 422, 433  
    laissez-faire-Lösung 415, 424, 428, 433  
    n-Personen 425, 426 ff.  
    2-Personen 425
- Gerechtigkeit 291  
Gerechtigkeitsinformation 343  
Gerechtigkeitsparameter (in Interessenaggregationsfunktion) 314, 343  
Gerechtigkeitsutilitarismus 322, 346  
Gewaltmonopol 255

- Gewichtung individueller Interessen 268 ff.  
 Gleichgewicht 10 ff., 241 f.  
   korreliertes 231 (Fußn.)  
 Gleichgewichtspunkt 203, 215, 227  
   in gemischten Strategien 211  
 Gleichgewichtsstrategie 203, 211, 213 f.  
 Gleichgewichtsprinzip für individuelle Entscheidungen 189, 192, 194, 232  
 Gleichheit von Handlungsergebnissen 86 ff.  
 Glück 178, 185
- Handeln, Handlung 111, 200 (Fußn.)  
   Handlungserklärung 35 ff., 53 ff.  
   Handlungsstrukturen 23 ff.  
 Haushaltsproduktionsfunktion 185 (Fußn.)  
 Hedonismus 171, 179 f., 184
- Implementierung individueller Rechte 443  
 Individualistisch-rationale Theorie der Moral 281  
 Induktionsproblem 224  
 Information 406  
 Informationssymmetrie 242, 253  
 Informationelle Rückkopplung 232  
 Intention 111 f., 115 f., 120 f.  
 Intentionalität 49 ff.  
 Interessen (und Meinungen) 246 f.  
 Interessenaggregationsfunktion (IAF) 313 ff.  
   interpersonal-kardinale 313  
   interpersonal-ordinale 314  
   kompensatorische 314, 318, 322, 334, 343, 345  
   nicht-kompensatorische 340  
   personal-kardinale 317  
   personal-ordinale 314  
 Interpersonelle Koordination 22  
 Interpersonelle Nutzenvergleiche 270  
 Intrapersonelle Koordination 22  
 invisible-hand-Erklärung 413, 434
- Kausale Abhängigkeit 174  
 Kausale Abschirmung 176  
 Kausale Struktur von Entscheidungssituationen 171 f., 176 f., 178  
 Klassensegregation 337  
 Kohärenzaxiome u. -theorie 6  
 Kollektive Akteure 19 ff., 239 ff.  
 Kompromiß 317, 330 ff., 343  
 Konditionalisierungsregel 228  
 Konsequentialismus 7 ff.  
 Kontinuität 6  
 Konvention 26  
 Konzession 318 ff.  
   Gauthier 320  
   Maximal (größtmögliche) 320  
   Zeuthen 320  
 Kooperation 15 ff., 239 ff., 288 ff., 396, 422 ff., 412 (Fußn.), 425 (Fußn.), 433 (Fußn.)  
 Koordination 413, 415, 423, 426  
   interpersonelle 22  
   intrapersonelle 22
- Liberalität 437 ff.
- Majorantenspiel 204, 210  
 Manipulation 21  
 Marktgesellschaft 391  
 Maximin-Regel 204, 207, 209  
 Mentales Idiom 35 ff.  
 Metrisierung 6 f., 201 (Fußn.)  
 Minimalmoral 315, 322, 330 f., 333, 340, 342, 344  
   erweiterte 315, 342  
 Minimax-Prinzip der relativen Konzession (Gauthier) 291, 305 ff.  
 Minorantenspiel 204, 210  
 Mögliche Welten 158 f.  
 Monotonie 6  
 Moral (Begriff) 283 ff.
- n-Personen-Spiel in Normalform 216  
 Nash-Lösung 290, 304 f., 241 ff.  
 Naturzustand 413 ff., 421, 423  
 Newcombs Problem 171 (Fußn.)  
 Nicht-Nutzen-Information 322, 329, 340, 345 f.

- Nichteliminierbarkeit direkt bezugnehmender Ausdrücke 144 ff.  
 Normen, soziale 245 ff.  
 Normenpaket-Verfahren 344 f.  
 Nutzen 4 ff., 62 ff., 171 f., 175 ff., 181 f., 183, 189, 190, 192 ff., 272  
  
 Operationalismus 185 (Fußn.), 200 (Fußn.)  
 Optimierung 42 ff.  
  
 Paket-Verfahren 344 f.  
 Pareto bedingung 263, 437 ff.  
 Partnersuche, vorteilsorientierte 350, 359  
 Payoff-Dominanz 241  
 Person, Begriff der 163 ff.  
 Positionen 440 ff.  
 Präferenzen 62 ff.  
 Praktische Rationalität 281, 310  
 Praktische Überlegung 170, 172, 179, 183, 192 f.  
   Dynamik 232  
   Zirkel 173 f., 189, 193  
 Praktischer Grund 187, 190 f.  
   und Ursache 173 (Fußn.), 187, 191  
 Prinzip von der Maximierung des erwarteten Nutzens 69 ff., 200, 222, 231 f.  
 Prisoner's Dilemma (Gefangenendilemma) 14 ff., 137 f., 231 (Fußn.), 241, 248 ff., 288 f., 295 ff., 299 ff., 314, 324, 338, 340, 342, 352 ff., 392 ff., 415, 422, 433  
   laissez-faire-Lösung 415, 424, 428, 433  
   n-Personen 425, 426 ff.  
   2-Personen 425  
 Protonorm, moralische 344 f.  
 Pseudo-Reputationen 401  
 Publizität 10, 415, 419 ff.  
  
 Rationale Moral 285 ff.  
 Rationalität 219 f. (s. a. Handlungsrationalität, epistemische Rationalität) und Nutzenfunktion als wechselseitiges Wissen 213 ff.  
   epistemische Zustände und Nutzenfunktion als wechselseitiges Wissen 215 ff.  
 Rationalitätsprinzipien 58 ff.  
   von Gefühlen 193  
   von Wünschen 193, 223  
 Recht(snorm), Begriff 415 ff.  
   holistische Struktur 415, 517 ff.  
 Rechte 411 ff., 434  
   auf Eigentum 412, 418 ff.  
   individuelle 443  
   natürliche 411 ff., 433 (Fußn.)  
   negative 416 f.  
   positive 417  
 Rechtsharmonie 441  
 Rechtsverteilung 441  
 Reduktion 6  
 Reduktion von Entscheidungssituationen 208 f.  
 Reflexivität 5  
 Regulationsmechanismen 397  
 Reichenbach-Axiom 229  
 Reputation 391, 399 ff.  
 Revealed Preference 7, 185  
 Risikoneigung 275  
  
 Satisfaktiver Zustand 177, 184  
 Selbstbewahrheitende Überzeugung 184, 187, 191  
 Selbstbindung 251, 253  
 Selbsterhaltungstreben 118 f.  
 Sicherheitsniveau, s. Maximin-Regel  
 Simulation 344 ff.  
 Small Worlds 172 (Fußn.), 207 f.  
 Solidarität 344 ff., 350 ff.  
 Soziale Wohlfahrtsfunktion  
   individualistische 261 ff.  
   lineare 267 ff.  
   utilitaristische 263 ff.  
 Spielernprozeß 229 f.  
 Spieltheorie 198 f., 201 f., 220, 224, 229, 235, 239 ff., 391 ff.  
 Status Quo 314 ff.  
   fiktiver 329 ff.  
   Gleichgewichts 325 ff.  
   Kaneko/Nakamura 329 f.  
   Nash 326, 328 f., 334 f., 339  
   natürlicher 325

- realer 330 ff.  
 Shapley 324 ff.  
 Straight Rule 228  
 Strategie, gemischte 205, 211, 216  
     rationalisierbare 214 (Fußn.)  
 Struktur von Handlungen 23 ff.  
 Strukturelle Rationalität 22 ff.  
 Substitution 6  
 Superspiel 295 ff., 353 ff., 393 ff.  
 Symmetrie 271  
 Sympathie 273 ff.
- Tauschkooperationen 394  
 Tracing Procedure 226  
 Transitivität 5
- Ungleichheit 358  
 Universalistische Theorie der Moral  
     309 f.  
 Unparteilichkeit 273, 286  
 Unsichtbare Hand 397 ff.  
 Ursachen (vs. Gründe) 126 ff.  
 Ursuppe, soziale 360  
 Utilitarismus 183 (Fußn.)
- Verdienste 322, 336, 340, 345  
 Verhandlungsmodell 313 ff.  
     distorsion 321 f.  
     iteratives 321  
     Koalitions 321  
     multi-period 321  
     oscillation cake 321  
     perfekte Welt 333, 341 ff.  
     preplay 321  
     risiko-empfindliches 321, 336 f.
- shrinking cake 321  
 trembling hand 322  
 von Gauthier (VM<sub>5</sub>) 217 f., 320  
 von Kalai/Smorodinsky (VM<sub>4</sub>) 317  
 von Kaneko/Nakamura (VM<sub>3</sub>)  
     317 ff.  
     von Nash (VM<sub>2</sub>) 317 ff., 327 f., 335  
     von Zeuthen (VM<sub>1</sub>) 317 f., 327 f.
- Verteilungsproblem, reines 325  
 Vetorecht 318, 345  
 Vollständigkeit 5  
 von Neumann/Morgenstern-Nutzenfunktion 266 ff.  
 Vorzugsrelationen 39
- Wahrheitsempfindlichkeit (von Entschlüssen) 110, 132 ff.  
 Wahrscheinlichkeit, subjektive 221  
 Wahrscheinlichkeitsverteilung 6  
 Wechselseitiges Wissen 213, 233  
 Weltverläufe 146 ff., 151 ff., 153 ff.  
 Wert der Existenz konkreter Menschen  
     161 ff.  
 Wertinvarianz 153 ff.  
 Wille, Willensmodell 112 ff.  
 Willensschwäche 116, 120 f., 132 f., 173  
 Wir-Intentionen 419  
 Wunsch 169, 184  
     Rationalität von 193 f., 223  
     wahrer 193 f., 186
- Zukünftige Generationen 341  
 Zwei-Personen-Nullsummenspiel 199,  
     203 f., 207, 215, 230