

Julian Nida-Rümelin

# Entscheidungs- theorie und Ethik

tuduv  Studie

7440422 \* 9 E

**tuduv-Studien**

Reihe Sozialwissenschaften

Band 31



CIP-Kurztitelaufnahme der Deutschen Bibliothek

**Nida-Rümelin, Julian:**

Entscheidungstheorie und Ethik / Julian

Nida-Rümelin. – München : tuduv-Verlagsgesellschaft,  
1987.

(tuduv-Studien : Reihe Sozialwissenschaften ;  
Bd. 31)

ISBN 3-88073-154-3

NE: tuduv-Studien / Reihe Sozialwissenschaften

© 1987 beim Autor. Alle Rechte vorbehalten.

Kommissionsverlag: tuduv-Verlagsgesellschaft mbH,

8000 München 2, Gabelsbergerstraße 15, Telefon 0 89/2 80 90 95

Gesamtherstellung: Fotodruck Frank GmbH, 8000 München 2

Printed in Germany

D 27/6952

*Die Genauigkeit, Kraft und Sicherheit des mathematischen Denkens erfüllte ihn fast mit Schwermut...*

*Wann immer man ihn bei der Abfassung mathematischer und logischer Abhandlungen oder bei der Beschäftigung mit den Naturwissenschaften gefragt haben würde, welches Ziel ihm vor-schwebe, so würde er geantwortet haben, daß nur eine Frage das Denken wirklich lohne, und das sei die des rechten Lebens...*

*aus "Der Mann ohne Eigenschaften"  
von Robert Musil*



## VORBERMERKUNG

Seit diese Arbeit als Dissertation der Fakultät für Philosophie, Wissenschaftstheorie und Statistik der Universität München unter dem Titel "Rationalität und Moralität - Eine Untersuchung der Bedeutung rationaler Entscheidung für die Sozialethik" vorgelegt wurde, sind drei Jahre vergangen. Ich hatte vor, ihre Ergebnisse in einen ersten Teil einer breiter angelegten Abhandlung zum Thema 'Rationalität und Moralität' zu integrieren. Dieses Vorhaben besteht fort, aber es war in der bis zur Publikation gesetzten Frist nicht mehr zu verwirklichen. Dieses Buch ist daher gegenüber der Fassung von 1983 kaum verändert. Ich habe lediglich einige sehr spezielle, für das Verständnis unwesentliche Passagen herausgenommen und eine Schlußbemerkung angefügt.

Das Thema 'Rationalität und Moralität' hat verschiedene Aspekte: Unter einem entscheidungstheoretischen Aspekt ist die zentrale Frage, ob konsequentialistische Handlungsmodelle zur Analyse moralischen Verhaltens geeignet sind; den sprachanalytischen hat mit besonderer Sorgfalt Richard Hare herausgearbeitet. Der staatsphilosophische (oder institutionentheoretische) Aspekt steht im Zentrum der zeitgenössischen Renaissance normativer Vertragstheorie.

Dieses Buch befaßt sich fast ausschließlich mit dem ersten der genannten Aspekte. Um diese thematische Eingrenzung deutlicher werden zu lassen, habe ich für die Buchveröffentlichung den Titel "Entscheidungstheorie und Ethik" gewählt.



## VORWORT

Seit Anfang der 70er Jahre wird auch im deutschen Sprachraum von einer 'Rehabilitierung der praktischen Philosophie' gesprochen. Und tatsächlich ist die philosophische Diskussion normativ-ethischer Fragen während der letzten Jahre spürbar aufgelebt. Das seit den 50er Jahren in der moralphilosophischen Forschung bestehende Ungleichgewicht zugunsten des metaethischen Bereichs wurde dabei merklich abgebaut.

Ein wesentliches Merkmal dieser Entwicklung ist ihr fachgebietsübergreifender Charakter:

Normativ-ethische Fragen sind heute wieder verstärkt Gegenstand aktueller Forschungsbemühungen nicht nur in der Moralphilosophie, sondern auch in der politischen Theorie, der Sozialwahltheorie, der theoretischen Ökonomie, der Spiel- und Entscheidungstheorie und natürlich weiterhin in Disziplinen wie der Rechtsphilosophie und der Theologie.

Und dennoch fällt auf, daß - von wenigen Ausnahmen abgesehen - die interdisziplinäre wechselseitige Befruchtung in der ethischen, speziell in der sozial-ethischen Diskussion nicht sehr weit gediehen ist. Um zwei Beispiele zu nennen: Philosophische Gerechtigkeitstheorien könnten an Klarheit und Differenziertheit gewinnen, wenn sie das sozialwahltheoretische Instrumentarium nutzen würden. Andererseits leiden die meisten Beiträge des 'collective-choice'-Ansatzes an einer mangelnden Klärung ihrer intuitiven Grundlagen. Man kann den Eindruck gewinnen, Jahrhunderte moralphilosophischen Bemühens um die motivationalen Grund-

lagen ethischer Entscheidung hätten in der Sozialwahltheorie keine Spuren hinterlassen.

Diese Arbeit versucht, Grundlagen der Sozialethik vor dem Hintergrund moderner Theorien rationaler Entscheidung zu behandeln. Und dabei werden Revisionen in beide Richtungen erforderlich: Traditionelle Fragen der Sozialethik müssen neu gestellt, und Grundbegriffe der rationalen Entscheidungs-, Spiel- und Sozialwahltheorie anders als üblich interpretiert werden.

Insofern soll die Arbeit auch dazu beitragen, diesem o.g. Mißstand abzuhelpfen.

Viele Impulse verdankt diese Arbeit der Diskussion auf dem 5. Internationalen Wittgenstein-Symposium zu Grundproblemen der Ethik im August 1980 in Österreich, sowie den Seminaren von Prof. Dr. Wolfgang Stegmüller, Prof. Dr. Wilhlem K. Essler, Prof. Dr. Otfried Höffe, Prof. Dr. Walter Hoering (Tübingen), Prof. Dr. Max Drömmner und Dr. Lucian Kern.

Herrn Dr. Spohn bin ich für zahlreiche Gespräche und wertvolle Kritik sehr dankbar.

Meinem Doktorvater Herrn Prof. Stegmüller danke ich für die teilnahmevolle Betreuung der Dissertation, Herrn Prof. Essler (jetzt Frankfurt) für erste Anregungen und Gespräche und Herrn Prof. Drömmner für gemeinsame Seminare zu zentralen Themen der Arbeit.

Nicht vergessen sind die Anregungen, die mir die Gespräche mit Herrn Prof. Konrad (+) über Jahre hinweg gegeben haben.

München, im März 1983

J. N-R.



## INHALTSVERZEICHNIS

Vorbemerkung

Vorwort

§ 1 Einführung und Übersicht

### A ENTSCHEIDUNGSRATIONALITÄT: KONSISTENZ INDIVIDUELLER UND KOLLEKTIVER PRÄFERENZEN

§ 2 Rationales Verhalten in Risikosituationen ohne Interaktion 15

§ 3 Rationales Verhalten in Unsicherheits-  
situationen ohne Interaktion 32

§ 4 Interaktionssituationen 42

§ 5 Rationalität kollektiver Präferenzen 48

### B PRÄFERENZTHEORIE UND MORALPHILOSOPHIE

§ 6 Zur interpersonellen Vergleichbarkeit  
subjektiver Bewertung 55

§ 7 Die Interdependenzproblematik normativer  
Aggregationstheorien 80

§ 8 Praktischer versus theoretischer Konsensus 102

§ 9 Zur Differenzierung des Präferenzbegriffs 110

§ 10 Moral und Nutzen 137

### C KOLLEKTIVE RATIONALITÄT UND MORALITÄT

§ 11 Zur Interpretation normativer Theorien  
der Interessenaggregation 150

§ 12 Rationalität und Moralität als konkurrie-  
rende Prinzipien 167

§ 13 Begriffsbestimmung: Moralisches System 200

§ 14 Interaktionssituationen vom PD-Typ 208

§ 15 Begriffsbestimmung: Kollektive Rationali-  
tät 219

§ 16 Moralität als Metapräferenztheorie 224

§ 17 Kollektive Rationalität als Minimalmoral 239

Schlußbemerkung 263

## ANHANG

- Mathematische und logische Hilfsmittel	266
- Allais' Kritik am Bayesianismus	269
- Zur Interdependenzproblematik	281
- Zur Aggregation individueller Urteile	287
LITERATURVERZEICHNIS	300
PERSONENREGISTER	327
SACHREGISTER	330
SYMBOLVERZEICHNIS	338

- \* Literaturverweise erfolgen das erste Mal unter Nennung von Autor und Titel, Ort, Jahr etc.; erneute (und beiläufige) Verweise werden abgekürzt: z.B.: 'Sen (1970), S.12'. Seitenangaben mit 'S.' geben die Referenzstelle wieder, ohne 'S.' Anfang bzw. Ende des Artikels.

Das Literaturverzeichnis führt die zitierte Literatur in alphabetischer, mehrere Arbeiten eines Autors in chronologischer Reihenfolge auf. Zusätzlich enthält es einige für das Thema wichtige Arbeiten, auf die im Text nicht Bezug genommen wurde.

## 1. EINFÜHRUNG UND ÜBERSICHT

Seit einigen Jahrzehnten behandelt die rationale Entscheidungstheorie als Spieltheorie, als Theorie kollektiver Entscheidungen und als Entscheidungstheorie im engeren Sinne<sup>1</sup> zahlreiche Probleme mit großer Sorgfalt und formaler Präzision, die traditionell Probleme der philosophischen Ethik und der politischen Theorie waren und sind,<sup>2</sup> ohne daß das einen angemessenen Niederschlag in diesen Disziplinen gefunden hätte. Das gilt besonders für den europäischen Kontinent.

Umgekehrt haben lange Zeit die über Entscheidungstheorie arbeitenden Ökonomen die philosophischen (begrifflichen, metaethischen, methodologischen) Aspekte ihrer Theoriebildung kaum thematisiert und es daher Philosophen allzu leicht gemacht, die normative Relevanz dieser Untersuchungen zu bestreiten.

Seit einigen Jahren sind Ansätze zu erkennen, diese Situation zu beenden: Entscheidungstheoretiker gehen mit ihren Mitteln traditionell philosophische Problemstellungen an. Philosophen setzen sich mit der Frage der normativen Relevanz entscheidungstheoretischer Er-

---

<sup>1</sup> Entscheidungstheorie im engeren Sinne meint die Theorie rationaler Entscheidung, in der sich der Entscheidende nur einer von seiner Entscheidung unabhängigen Umwelt gegenüber sieht. Die Tradition der modernen rationalen Entscheidungstheorie geht zumindest bis auf Daniel Bernoulli's *Specimen theoriae novae de mensura sortis* (1780) zurück.

<sup>2</sup> Das bekannteste Beispiel ist vermutlich das 'general possibility theorem' von Kenneth Arrow, das zeigt, daß vier - scheinbar sehr bescheidene - Verfassungsprinzipien von keiner Regel kollektiver Entscheidung zugleich erfüllt werden können, wenn diese nur minimalen Rationalitätsbedingungen (Transitivität, Reflexivität und Vollständigkeit ihres Wertebereichs) genügen soll. (*Social choice and individual values*, New York u.a. 1963).  
Auch Überlegungen zur 'collective choice' sind nichts neues: Condorcet stellte 1785 fest, daß die Mehrheitsentscheidung einer elementaren Rationalitätsbedingung nicht genügt: Sie aggregiert individuelle Präferenzen nicht immer zu einer transitiven (kollektiven) Präferenzrelation.

gebnisse auseinander,<sup>1</sup> bedienen sich der formalen Hilfsmittel mathematisch präzisierter Entscheidungstheorie<sup>2</sup> oder verwenden im Aufbau ihrer Theorie entscheidungstheoretische Modelle.<sup>3</sup>

Verkürzt kann man sagen, das alte philosophische Problem der Beziehung von Rationalität und Moralität wird zunehmend zum Thema einer interdisziplinären Forschergemeinschaft.<sup>4</sup> Dies scheint mir eine wesentliche und für eine systematische - nicht nur philosophiehistorische - Rehabilitierung praktischer Philosophie erfolgversprechende Entwicklung.<sup>5</sup>

Die Tatsache, daß die analytische Philosophie die rationale Begründbarkeit normativer Urteile nicht mehr generell bestreitet, kann dabei als eine Form methodologischer Absicherung dieser Entwicklung gelten. Analytische Ethik hat eine Reihe metaethischer Theorien hervorgebracht, die sich u.a. mit dem Begründungsproblem moralischer Urteile auseinandersetzen.<sup>6</sup> Das auffallende Mißverhältnis zwischen detaillierten Untersuchungen zu den methodologischen Problemen normativer Theorien und dem Mangel an modernen normativen Theorien selbst wird sich durch eine gegenseitige Befruchtung von Entscheidungstheorie und Ethik am erfolgversprechendsten beheben lassen. Hier setzt die Problemstellung dieses Dissertationsvorhabens ein:

---

<sup>1</sup> Besonders anregend die Habilitationsschrift von Otfried Höffe, *Strategien der Humanität. Zur Ethik öffentlicher Entscheidungsprozesse*, Freiburg/München 1975.

<sup>2</sup> Vgl. etwa die Vorträge des Int.Symposiums zu Entscheidungstheorie und Sozialethik (Schloß Reisingburg 1976), teilweise abgedruckt in: H.W. Gottinger/W. Leinfellner (Hg.), *Decision theory and social ethics*, Dordrecht 1978.

<sup>3</sup> z.B. John Rawls, *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975.

<sup>4</sup> Sie hat verschiedene Diskussionsforen, vgl. bes. die Zeitschrift *Theory and decision*.

<sup>5</sup> Vgl. M. Riedel (Hg.), *Rehabilitierung der praktischen Philosophie*, Freiburg 1972.

<sup>6</sup> Wobei sich die analytische Ethik unterdessen völlig von ihrem emotivistischen Ausgangspunkt entfernt hat. Diese Entwicklung zeichnete sich schon in den 50er Jahren mit den Arbeiten von Nowell-Smith (1954), Baier (1958) und Brandt (1959) ab.

Wie sind die begrifflichen Grundlagen zu fassen, damit entscheidungstheoretische Präzisierungen ethischer Theorien überhaupt sinnvoll vorgenommen werden können, und welche Beziehungen ergeben sich zwischen rationaler und moralischer Entscheidung und auf der Metaebene zwischen den Theorien rationaler Entscheidung und den Theorien moralischer Entscheidung?

Dieses Problem wird nicht in seiner ganzen Breite, sondern auf die sozialetische Dimension eingeschränkt behandelt.

Die Arbeit gliedert sich in zwei Teile.

Im ersten Teil geht es zunächst um eine kurze Darstellung der entscheidungs-, spiel- und sozialwahltheoretischen Explikationen rationaler Präferenzen und rationaler Entscheidung. (Teil I, Kap. A)

In §2 wird ein entscheidungstheoretisches Modell eingeführt, daß sich sowohl von Savages Grundmodell wie von Jeffreys 'holistischer' Konzeption unterscheidet.

Aufgrund von Allais' Kritik der 'amerikanischen Schule' schlagen wir eine Modifikation des Bernoulli-Prinzips vor, die die Risikobereitschaft einer Person adäquater zu berücksichtigen erlaubt.

§3 beschäftigt sich mit der Explikation rationalen Verhaltens in Unsicherheitssituationen. Dabei geht es um die Frage der Abgrenzung von Risiko- und Unsicherheitssituationen, die Darstellung konkurrierender Entscheidungsregeln und den Ansatz einer generellen Theorie rationalen Verhaltens.

Die Beziehung rationalen und moralischen Verhaltens in Interaktionssituationen beschäftigt uns unter verschiedenen Aspekten in allen Teilen der Arbeit.

§4 beschränkt sich deshalb auf einige spieltheoretische Grundbegriffe, Darstellungsformen und Klassifikationen.

§5 befaßt sich mit den Grundbegriffen subjektivistischer Aggregationstheorien und den Rationalitätsbedingungen kollektiver Präferenzen.

Mit §5 sind die einführenden Darlegungen der Arbeit abgeschlossen.

Das Kapitel B des ersten Teils beschäftigt sich mit den präferenztheoretischen Grundlagen sozialetischer Theorien.

Während §6 ihre Informationsbasis, das vieldiskutierte Problem der interpersonellen Vergleichbarkeit subjektiver Bewertungen behandelt, sollen die Überlegungen von §7 und §8 deutlich machen, welche Differenzierung des Präferenzbegriffs notwendig ist, um das Instrumentarium der Spiel- und Sozialwahltheorie für die Sozialethik nutzbar zu machen.

§9 zieht daraus die Konsequenzen und §10 kritisiert anhand einer axiomatischen Darstellung eine mögliche Alternative zu diesem Ansatz.

In §6 wird aufgezeigt, daß die Problematik interpersoneller Vergleiche durch die Möglichkeit intrapersoneller Präferenzeninterdependenz wesentlich abgeschwächt werden kann; insofern erfolgt hier ein Vorgriff auf die Ergebnisse der nachfolgenden Abschnitte.

Der Vorschlag einer Aufspaltung des Präferenzbegriffes (§9) stützt sich auf zwei Ergebnisse: Einmal die 'doppelte Aporie' der Aggregation (uniformer) interdependenter Präferenzen (§7) und zum zweiten die Notwendigkeit, die theoretische und die praktische Ebene der Aggregation auseinanderzuhalten (§8).

Der zweite Teil der Arbeit befaßt sich schwerpunktmäßig mit dem Zusammenhang individueller und kollektiver Rationalität einerseits und kollektiver Rationalität und Moralität andererseits.

Ein erster Abschnitt (§11) steckt den metatheoretischen Rahmen ab. Die Unterscheidung zwischen axiologischer und nicht-axiologischer Ethik und zwischen normativ-ethischen Theorien erster und zweiter Ordnung erlaubt es zwar einerseits, die allgemeine ethische Begründungsproblematik speziell für normative Aggregationstheorien abzuschwächen, zwingt aber andererseits zu einer hypothetischen Interpretation ihrer Handlungsempfehlungen.

In §12 beschäftigen wir uns mit dem Konflikt individueller Rationalität und sozialer Optimalität vor dem Hintergrund der spieltheoretischen Konzeption rationalen Verhaltens in Interaktionssituationen.

Das legt die Frage nahe, ob es einen begrifflichen Zusammenhang zwischen Moralität, individueller Rationalität und kollektiver Rationalität gibt. Damit befassen sich §13 und §14.

Aufgrund der Ergebnisse dieser beiden Abschnitte kommt der Analyse von Situationen, in denen individuell rationales Verhalten zu kollektiver Irrationalität führt, eine zentrale Rolle für die Themenstellung dieser Arbeit zu - solche Situationen nennen wir Interaktionssituationen vom PD-Typ (nach dem bekannten spieltheoretischen 'prisoner's dilemma'). Dieser Analyse sind daher die Abschnitte §15, §16 und §17 gewidmet.

Der darauf folgende Abschnitt (§18) weitet die Konzeption kollektiver Rationalität in einer Weise aus, die es in größerem Umfang erlaubt, normative Prinzipien der Interessenaggregation unter dem Aspekt kollektiver Rationalität zu diskutieren.

Abschließend wird die Erkenntnis aus §17 aufgegriffen, daß eine rein teleologische Sozialethik die Rolle der Verhaltensrationalität nicht adäquat klären kann und daher durch deontologische Prinzipien ergänzt und korrigiert werden muß (§19).

A ENTSCHEIDUNGSRATIONALITÄT:

KONSISTENZ INDIVIDUELLER UND KOLLEKTIVER  
PRÄFERENZEN

§ 2 Rationales Verhalten in Risikosituationen  
ohne Interaktion: S. 14

§ 3 Rationales Verhalten in Unsicherheits-  
situationen ohne Interaktion: S. 31

§ 4 Interaktionssituationen: S. 41

§ 5 Rationalität kollektiver Präferenzen: S. 47



## 2. RATIONALES VERHALTEN IN RISIKOSITUATIONEN OHNE INTERAKTION

Wir sprechen auch im Alltagsleben und ohne Bezug auf eine spezielle Theorie rationaler Entscheidung von rationalem, konsistentem, in sich logischem, oder aber von irrationalem, inkonsistentem, unlogischem Verhalten, Handeln, Entscheiden. Können solche Formulierungen mit einer präzisen Bedeutung versehen werden?

Zunächst scheint eine gewisse Normierung des Sprachgebrauchs sinnvoll:

- (1) Das Prädikat "logisch" oder "unlogisch" sollte nur auf Sätze oder Mengen von Sätzen angewendet werden. Eine Erweiterung des Anwendungsbereiches als Eigenschaft des Denkens einer Person oder gar der Person selbst ist möglicherweise zweckmäßig. Verhalten, Handlungen oder Entscheidungen sollten jedoch besser nicht "logisch" oder "unlogisch" genannt werden. Insofern ist auch "Logik der Entscheidung" oder "Entscheidungslogik" zur Bezeichnung einer Theorie rationaler Entscheidung eher irreführend.
- (2) Inkonsistentes Verhalten ist irrational. Aber irrationales Verhalten ist nicht notwendigerweise inkonsistent, es mag neben der Inkonsistenz noch andere Kriterien für die Irrationalität einer Entscheidung geben. Es stellt sich also die Frage: Was genau ist inkonsistentes Verhalten? Die Bayessche Entscheidungstheorie gibt darauf eine Antwort.

Eine Klasse von Entscheidungen bzw. eine Präferenzrelation von Handlungen einer Person ist konsistent genau dann, wenn es eine Wahrscheinlichkeitsfunktion über die Menge der Ereignisse und eine subjektive Bewertungsfunktion der Konsequenzen gibt, so daß der subjektive Erwartungswert

der präferierten Entscheidung jeweils höher ist als derjenige der nicht-präferierten.<sup>1</sup>

Rationales Verhalten können wir also in einer ersten Annäherung als Verhalten in Einklang mit den eigenen Präferenzen betrachten. Unter bestimmten einschränkenden Annahmen ist das entgegen erstem Anschein eine begrifflich klare und überzeugende Festlegung, sofern damit Verschärfungen der Rationalitätsbedingungen und Einschränkungen ihres Anwendungsbereiches nicht ausgeschlossen werden. Rationales Verhalten in einer Risikosituation folgt demnach dem Bernoulli-Prinzip bzw. dem Bayesschen Rationalitätskriterium, das empfiehlt, Handlungsalternativen nach der Größe des Erwartungswertes der subjektiven Wünschbarkeit in eine Präferenzordnung zu bringen.<sup>2</sup>

Nun gibt es eine Reihe von entscheidungstheoretischen Modellen<sup>3</sup>, in deren Kontext das Bayessche Rationalitätskriterium formuliert werden kann, die wir hier nicht diskutieren wollen.<sup>4</sup> Dennoch soll im folgenden ein Grundmodell skizziert werden, das geeignet ist, rationales Verhalten in Risikosituationen ohne Interaktion zu explizieren.

Es entspricht keinem der mir bekannten entscheidungstheoretischen Ansätzen vollständig. Die Abweichungen ergeben sich nicht in erster Linie wegen der Anwendungsprobleme im intendierten Bereich (Risikosituationen

---

<sup>1</sup> Vgl. Jeffrey (1965), S.533f, der Präferenzrelationen dieser Art "coherent" nennt.

<sup>2</sup> Es gibt zahlreiche gute Darstellungen der rationalen Entscheidungstheorie, speziell der Theorie rationaler Entscheidungen in Risikosituationen, z.B.: Chernoff/Moses (1959), Luce/Suppes (1963), Schneeweiß (1967), Krelle (1968), Raiffa (1968), Bühlmann/Loeffel/Nievergelt (1969 u. 1975), Stegmüller (1973), Gäfgen (1974), Göttinger (1974a), Egle (1975), Spohn (1978).

<sup>3</sup> Vgl. besonders die Konzeptionen von Savage (1972), Jeffrey (1967), Fishburn (1964, 1966, 1970, 1972).

<sup>4</sup> Eine detaillierte Analyse und Kritik alternativer Entscheidungsmodelle: W. Spohn, *Grundlagen der Entscheidungstheorie*, Kronberg/Ts. 1978. Vgl. a. W. Stegmüller, *Entscheidungslogik (rational-Entscheidungstheorie)* [Wissenschaftstheorie Bd.4] Berlin u.a. 1973.

ohne Interaktion), sondern weil wir die Begriffe möglichst unverändert auch bei der Diskussion unterschiedlicher Arten individueller Präferenzen und bei der Klärung sozialetischer Fragen - besonders der kollektiven Rationalität und der gerechten Interessenaggregation - also bei der Analyse von Situationen gebrauchen wollen, die jenseits der entscheidungstheoretischen Problematik im engeren Sinne liegen.

Die wesentlichen Begriffe des Modells lassen sich folgendermaßen charakterisieren:

### 1. Zustände:

"Zustand" ist ein wichtiger Grundbegriff nicht nur der Entscheidungs-, sondern auch der Wahrscheinlichkeits- und Sozialwahltheorie. Savage definiert "Zustand" als "a description of the world, leaving no relevant aspect undescribed"<sup>1</sup> Gleichbedeutend ist der Begriff der "möglichen Welten". Jeffrey versucht, diesen Begriff wie folgt zu erläutern:

" Wir wollen eine nichtleere Menge von Sätzen einen *Bericht* nennen. Ein Bericht ist *widerspruchsfrei*, wenn seine Elemente nicht in logischem Widerspruch miteinander stehen; ein Bericht ist *vollständig*, wenn ihm kein Satz der Sprache zugefügt werden kann, ohne daß er widersprüchlich würde. Gleichbedeutend können wir sagen, ein *vollständiger, widerspruchsfreier Bericht* ist eine Menge von Sätzen, die genau einen von jedem Paar *kontradiktorischer Sätze* enthält, die in der Sprache des Subjektes vorkommen. Ein *vollständiger, widerspruchsfreier Bericht* beschreibt eine *mögliche Welt* (...) so genau, wie es im Hinblick auf die sprachlichen Möglichkeiten des Subjektes angeht. Wenn es jedoch als allzu metaphysisch erscheint, von möglichen Welten zu sprechen, dann können wir unsere Aufmerksamkeit auch den Berichten selbst zuwenden und sagen, ein *vollständiger, widerspruchsfreier Bericht* sei nichts anderes als eine *mögliche Welt*."<sup>2</sup>

'Zustände' spielen allerdings in den unterschiedlichen entscheidungstheoretischen Modellen eine verschiedene

---

<sup>1</sup> Savage (1972), S.9. Allerdings habe ich strenggenommen weder subjektive Wahrscheinlichkeitsannahmen, noch subjektive Wünschbarkeitsannahmen über *Beschreibungen*, sondern höchstens über das Beschriebene.

<sup>2</sup> Jeffrey (1967), S.264.

Rolle. In unserem Modell sind die 'möglichen' Welten', die 'Zustände' besonders zentral, denn wir nehmen an, daß die individuellen subjektiven Bewertungsfunktionen (Nutzenfunktionen)  $u_i$  als Definitionsbereich die Menge der Zustände haben:  $U_i : \exists \exists \xi \rightarrow x \in \mathbb{R}$

## 2. Ereignisse

Ereignisse sind Mengen von Zuständen.

Die Menge der Ereignisse ist die Potenzmenge von  $\Xi$ :  
 $E := \text{Pot}(\Xi)$ .

Ein Ereignis ist die Menge derjenigen Zustände, in denen die betreffende Proposition wahr ist.

$E$  ist  $\sim \cup$ -abgeschlossen, d.h. es gilt

$$(1) \ \varepsilon \in E \Rightarrow \sim \varepsilon \in E.$$

$$(2) \ \varepsilon, \varepsilon' \in E \Rightarrow \varepsilon \cap \varepsilon' \in E.$$

$$(3) \ \varepsilon, \varepsilon' \in E \Rightarrow \varepsilon \cup \varepsilon' \in E.^1$$

$\varepsilon \cap \varepsilon'$  ist dasjenige Ereignis, das sich ergibt, wenn  $\varepsilon$  und  $\varepsilon'$  zugleich eintreten.

$\varepsilon \cup \varepsilon'$  ist dasjenige Ereignis, das darin besteht, daß mindestens eines der beiden Ereignisse  $\varepsilon$  und  $\varepsilon'$  eintritt.

$\sim \varepsilon$  ist dasjenige Ereignis, das darin besteht, daß  $\varepsilon$  nicht eintritt.

$$\sim \varepsilon = \Xi \setminus \varepsilon.$$

Zwei Ereignisse heißen entgegengesetzt, wenn gilt

$$\varepsilon \cup \varepsilon' = \Xi \wedge \varepsilon \cap \varepsilon' = \emptyset, \text{ das ist genau dann der Fall, wenn } \varepsilon' = \sim \varepsilon.$$

Zwei Ereignisse heißen disjunkt, wenn gilt  $\varepsilon \cap \varepsilon' = \emptyset$ .

Eine Menge von Ereignissen  $M$  heißt vollständig, wenn die Vereinigung der Ereignisse dieser Menge gleich  $\Xi$  ist.

---

<sup>1</sup> E können wir auch als Ereignisfeld oder Möglichkeitsraum bezeichnen.  $E$  ist ein boolescher Verband (boolesche Algebra); vgl. F. Jehle, *Boolesche Algebra*, München 1974.

### 3. Entscheidungssituation

Eine Entscheidungssituation (im engeren Sinne)<sup>1</sup> ist durch sieben Größen gekennzeichnet:

$$\langle \iota, \Xi, U, C, \Sigma^1, \rho_1, \nu_1 \rangle$$

- $\iota$  ist der Entscheidungsträger.
- $\Xi$  ist die Menge der Zustände.
- $U$  ist eine vollständige Menge<sup>2</sup> disjunkter Ereignisse. Die Elemente von  $U$  nennen wir Umstände  $e$ .
- $C$  ist eine Menge von Ereignissen. Elemente von  $C$  bezeichnen wir mit  $c, c'$  etc., bzw.  $c_{1\alpha}, c_{1\beta}, c_{1\gamma}$  etc.<sup>3</sup>
- $\Sigma^1$  ist die Menge der offenstehenden Handlungen.  $\Sigma^1$  ist eine Menge von Funktionen von  $U$  in  $C$ . Elemente von  $\Sigma^1$  bezeichnen wir mit  $\sigma, \sigma'$ , im abzählbaren Fall häufig mit  $\sigma_{1\alpha}, \sigma_{1\beta}, \sigma_{1\gamma}$  etc.<sup>3</sup>
- $\rho_1$  ist eine subjektive handlungsbedingte Wahrscheinlichkeitsfunktion auf  $\text{Pot}(\Xi)$
- $\nu_1$  ist eine Funktion von  $\Xi$  in  $\mathbb{R}$ .

### 4. Entscheidungskriterium

Ein Entscheidungskriterium ordnet jeder Entscheidungssituation  $\langle \iota, \Xi, U, C, \Sigma^1, \rho_1, \nu_1 \rangle$  zwei Funktionen zu:

eine Funktion  $\hat{\nu}_1$  (subjektive Bewertung der Konsequenzen) von  $C$  in  $\mathbb{R}$  und

eine Funktion  $u_1$  (subjektive Bewertung der eigenen Handlungen) von  $\Sigma^1$  in  $\mathbb{R}$ .

Die Bayessche Entscheidungsregel ordnet einer Entscheidungssituation  $\langle \iota, \Xi, U, C, \Sigma^1, \rho_1, \nu_1 \rangle$  in Abhängigkeit von der subjektiven Bewertung der Konsequenzen folgende Bewertungsfunktion der Handlungen zu:

---

<sup>1</sup> Eine Entscheidungssituation i.e.S. umfaßt Risiko- und Sicherheits-situationen ohne Interaktion.

<sup>2</sup> Alle Mengen seien finit und nicht leer.  
#U, #C, # $\Sigma^1$ , #c, #e  $\in \mathbb{N} \wedge U \neq \emptyset, C \neq \emptyset, \dots$

<sup>3</sup>  $\alpha, \beta, \gamma \in \mathbb{N}$ .

$$(I) \quad u(\sigma) = \sum_{e \in U} \hat{v}_1(\sigma(e)) \cdot \rho_1(e|\sigma)$$

bzw. wenn  $\rho_1(e)$  entscheidungsunabhängig ist:

$$(I') \quad u_1(\sigma) = \sum_{e \in U} \hat{v}_1(\sigma(e)) \cdot \rho_1(e)$$

Daher ist es naheliegend, den Zusammenhang von Konsequenzen und Zustandsbewertung in ähnlicher Weise zu beschreiben:

$$\hat{v}_1(c) = \sum_{\xi \in C} v_1(\xi) \cdot \rho_1(\{\xi\}|\sigma) \quad (*) \quad \text{wäre allerdings}$$

inadäquat, denn während für die Wahrscheinlichkeiten der

Umstände  $\sum_{e \in U} \rho_1(e|\sigma) = 1$ , gilt natürlich im allgemeinen nicht:  $\sum_{\xi \in C} \rho_1(\{\xi\}|\sigma) = 1$ .<sup>1</sup>

Und das hat zur Konsequenz, daß die Gewichtung von (\*) durch die entscheidungsbedingte Gesamtwahrscheinlichkeit von c:  $\rho_1(c|\sigma)$  'verzerrt' wird: Während (I) zur Konsequenz hat, daß

$$\min(\{\hat{v}_1(\sigma(e))\}) \leq u(\sigma) \leq \max(\{\hat{v}_1(\sigma(e))\})$$

gilt nicht entsprechend

$$\min(\{v_1(\xi)\}) \leq \hat{v}_1(c) \leq \max(\{v_1(\xi)\}), \text{ sondern}$$

$$\min(\{v_1(\xi)\}) \cdot \sum_{\xi \in C} \rho_1(\{\xi\}|\sigma) \leq \hat{v}_1(c) \leq \max(\{v_1(\xi)\}) \cdot \sum_{\xi \in C} \rho_1(\{\xi\}|\sigma)$$

daher muß (\*) korrigiert werden zu:

$$(II) \quad \hat{v}_1(c) = \frac{\sum_{\xi \in C} v_1(\xi) \cdot \rho_1(\{\xi\}|\sigma)}{\sum_{\xi \in C} \rho_1(\{\xi\}|\sigma)} = \rho_1(c|\sigma)$$

(II) ähnelt Jeffreys Wünschbarkeitsdefinition für Propositionen.<sup>2</sup>

$$(**) \quad u_1(p \cup p') = \frac{v_1(p) \cdot \rho_1(p) + v_1(p') \cdot \rho_1(p')}{\rho(p \cup p')}, \text{ wobei } p, p' \text{ Propositionen.}$$

Aber es ist wichtig zu sehen, daß II nicht den gleichen

<sup>1</sup> Allerdings gilt sowohl  $\forall e, e' \in U : \rho_1(e \cap e') = 0$  (da die Umstände einer Entscheidungssituation disjunkt sind), als auch  $\forall \xi, \xi' \in C : \rho_1(\xi \cap \xi') = 0$ .

<sup>2</sup> s. Jeffrey (1967), S.91ff.

Einwänden wie (\*\*)<sup>1</sup> ausgesetzt ist, denn nach Definition der Entscheidungssituation gilt für alle Konsequenzen einer Handlung  $c = \sigma(e)$ ,  $c' = \sigma(e')$ :  $\rho_1(c \cap c') = 0$ , wenn  $c \neq c'$ .<sup>2</sup>

In den meisten Fällen allerdings wäre es völlig abwegig, anzunehmen, daß  $\iota$  tatsächlich eine subjektive Wahrscheinlichkeitsbewertung für alle möglichen Welten<sup>3</sup> aus  $C$  vornehmen kann. Realistischer ist es,  $\iota$  eine Wahrscheinlichkeitsbewertung einiger Teilmengen von  $C$  zuzuordnen - während wir idealisiert annehmen wollen, daß er innerhalb dieser Teilmengen  $c$  völlig ratlos ist, welche Wahrscheinlichkeiten er ihren Elementen zuschreiben soll.<sup>4</sup>

Ein für die Anwendung geeigneterer Übergang von der Zustands- zur Konsequenzenbewertung enthält daher zwei Schritte:

1. Bildung einer (bzgl.  $c$ ) vollständigen Menge  $M$  disjunkter Teilmengen von  $c$  :

$$\forall \mu, \mu' \in M : \mu \cap \mu' = \emptyset \wedge \bigcup_{\mu \in M} \mu = c \quad \text{dergestalt,}$$

daß  $\iota$  eine subjektive Wahrscheinlichkeitsbewertung für alle  $\mu$  aus  $M$  hat.

Die subjektive Bewertung der Elemente von  $M$  könnte dann etwa nach der Regel  $\check{\nu}_1(\mu) = \min(\{\nu_1(\xi) \mid \xi \in \mu\})$  erfolgen, wobei sich die Einteilung von  $c$  in  $M$  insofern an der Bewertungsfunktion  $\nu_1$  orientieren kann, als die Elemente von  $M$  vorzugsweise solche Zustände zusammenfassen, die bewertungsmäßig keinen großen Unterschied für  $\iota$  machen.

---

<sup>1</sup> Vgl. Spohn (1978), S.77ff.

<sup>2</sup> Wobei natürlich  $e, e' \in U$  und damit - da  $\sigma$  Funktion ist - auch  $e \neq e'$ .

<sup>3</sup> Wortsprachlich sprechen wir von der subjektiven Wahrscheinlichkeit eines Zustandes, meinen damit aber genaugenommen nicht  $\rho_1(\xi)$ , sondern  $\rho_1(\{\xi\})$ .

<sup>4</sup> Genauer: denjenigen Teilmengen von  $c$ , die Einermengen sind.

2. Der Übergang von  $\tilde{v}_1$  zu  $\hat{v}_1$  wäre dann erneut 'bayesianisch':

$$\hat{v}_1(c) = \sum_{\mu \in M} \tilde{v}_1(\mu) \cdot \rho_1(\mu | \sigma).$$

Dieser Ansatz unterscheidet sich von Savages Konzeption u.a. darin, daß die Bewertung der Konsequenzen eine 'abgeleitete Bewertung' ist,<sup>1</sup> daß Konsequenzen und Umstände Mengen von Zuständen sind und daß die Konsequenzen eine objektivistische Interpretation haben.<sup>2</sup>

Der Konzeption Jeffreys gleicht der Ansatz darin, daß Konsequenzen und Umstände Teilmengen von  $\Xi$  sind. Er unterscheidet sich jedoch auch von diesem 'holistischen' Ansatz in zwei Punkten: Einmal darin, daß Handlungen keine Teilmengen von  $\Xi$  sind, sondern im Rahmen einer gegebenen Entscheidungssituation als Funktion von der Menge der Umstände in die Menge der Konsequenzen definiert werden,<sup>3</sup> und zum anderen darin ( und das ist für unser Thema wichtig), daß es keine primäre Bewertung von Propositionen gibt.

Erst dieser Unterschied zu Jeffreys Modell erlaubt es, einige für die sozialetische Debatte wichtige Differenzierungen des Präferenzbegriffs vorzunehmen. (vgl. §9)

## 5. Propositionen

Wir haben schon oben von "Propositionen" gesprochen. Wünsche, Hoffnungen, Vermutungen, Annahmen beziehen sich auf Propositionen.

---

<sup>1</sup> So wie in allen Entscheidungsmodellen die Bewertung der Handlungen keine primäre, sondern eine (aufgrund der Konsequenzen- und Wahrscheinlichkeitsbewertung) abgeleitete Bewertung ist.

<sup>2</sup> Konsequenzen sind nicht 'Zustände der Person' bzw. Beschreibungen von Merkmalen, die  $\iota$  für seine Nutzenbewertung relevant erscheinen.

<sup>3</sup> Werden also auch der Forderung von Spohn gerecht, daß sich die subjektive Wahrscheinlichkeitsfunktion  $\rho_1$  nicht über die Handlungsalternativen ( $\Sigma'$ ) erstrecken sollte; vgl. Spohn (1978), S.72-78, S.186-193.



Er wünscht sich, daß morgen die Sonne scheint; sie hofft, daß er sie besucht; wir vermuten, daß es keine größte Primzahl gibt; wir wollen annehmen, daß sich Wünsche auf Propositionen beziehen ...

"Ich wünsche mir mehr Muße" läßt sich auch 'propositional' formulieren: "Ich wünsche mir, daß ich mehr Muße habe". Grammatikalisch sind die Gegenstände von Glaubens- und Wünschensdispositionen in der Alltagssprache häufig andere Dinge als Propositionen. Jeden Satz dieser Art kann man jedoch immer ohne Bedeutungsveränderung auch propositional formulieren. Sprachlich formulierte Präferenzen (jeglicher Art) beziehen sich auf Propositionen: "Er macht heuer lieber in Italien Urlaub als in Spanien" D.h. er bevorzugt die eine Proposition - nämlich, daß er heuer in Italien Urlaub macht gegenüber der anderen Proposition - nämlich, daß er heuer in Spanien Urlaub macht. In ähnlicher Weise läßt sich so gut wie jede Präferenzbehauptung propositional interpretieren.

In welcher Beziehung stehen nun Propositionen und propositionale Präferenzen zu den bisher eingeführten Begriffen?

Eine Proposition läßt sich jeweils durch eine Menge von Zuständen charakterisieren. Mathematisch können wir daher die Menge der Propositionen mit der Potenzmenge von  $\Xi$ , mit  $\text{Pot}(\Xi)$  identifizieren. Da die subjektive Wahrscheinlichkeitsfunktion  $\rho_1$  auf dem Ereignisraum definiert ist, bekommt auf diese Weise jede Proposition auch eine subjektive Wahrscheinlichkeit zugeordnet. Das ist jedoch i.a. eine zu starke Idealisierung, die wir schon oben dadurch abgeschwächt haben, daß wir zuließen, gegebenenfalls den Definitionsbereich von  $\rho_1$  auf eine Teilmenge von  $\text{Pot}(\Xi)$  einzuschränken.

---

<sup>1</sup> Kennt man die Teilmenge von  $\text{Pot}(\Xi)$ , auf der  $\rho_1$  sinnvoll definiert werden kann, hat man immer noch die Freiheit, unterschiedliche Einteilungen von  $\text{Pot}(\Xi)$  bzw.  $c$  zu wählen, deren Elemente zum Definitionsbereich von  $\rho_1$  gehören. Natürlich erlaubt nicht jede Teilmenge von  $\text{Pot}(\Xi)$ , eine solche Einteilung vorzunehmen, aber wir können fordern, daß jeder Definitionsbereich einer subjektiven Wahrscheinlichkeitsfunktion diese Eigenschaft hat.

Wenn wir nicht nur sagen, daß alle Propositionen Elemente von  $Pot(\Xi)$ , also Mengen von möglichen Welten sind, sondern daß jede Menge von Zuständen auch einer Proposition entspricht, so lassen wir zu, daß es möglicherweise Propositionen gibt, die wortsprachlich nicht formuliert werden können: Denn wer garantiert, daß unsere Sprache so reichhaltig ist, daß eine beliebige Menge von Zuständen durch einen Satz der Art "Es ist der Fall, daß ..." beschrieben werden kann.<sup>1</sup> Insofern ist es vielleicht zweckmäßiger,  $Pot(\Xi)$  mit der Menge der Sachverhalte zu identifizieren. Dann entspricht zwar jeder Proposition ein Sachverhalt, aber nicht jedem Sachverhalt eine Proposition, und es kann dann analytisch bleiben, daß jede Proposition einem 'Es ist der Fall, daß -Satz' entspricht. Präferenzen haben wir auch über Sachverhalte, die keiner Proposition entsprechen: Wir wissen oft genau, was wir wollen, ohne genau sagen zu können, was wir wollen.

Nachdem wir diesen Vorbehalt gemacht haben, wollen wir uns wieder der 'schlampigen' Ausdrucksweise bedienen, die die Menge der Propositionen mit  $Pot(\Xi)$  identifiziert.

Umstände und Konsequenzen sind in diesem Sinne Propositionen, die eigenen Handlungen nicht.<sup>2</sup>

Während die Wünschbarkeit von Konsequenzen jeweils letztlich unsicher bleibt und von den subjektiven Wahrscheinlichkeitsbewertungen abhängt, kann man sich vorstellen, daß die Wünschbarkeit von Zuständen völlig unabhängig von Wahrscheinlichkeitsbewertungen (und anderen epistemischen Voraussetzungen über die tatsächliche Welt)

---

<sup>1</sup> Auf diese rhetorisch gemeinte Frage gibt es sogar eine andere Antwort als "Niemand", nämlich "Jeffrey" - siehe das Zitat oben, in dem Zustände mit *vollständigen Berichten* identifiziert werden. Dieser Identifikation sollten wir uns aber nicht anschließen, denn damit wird der Sprache eine übermäßige Kraft der Weltgestaltung zugemutet.

<sup>2</sup> Natürlich lassen sich auch die eigenen Handlungen i.a. propositionale formulieren; dennoch sind sie in unserem Modell keine Elemente des Ereignisraumes, sondern Funktionen von  $U$  in  $C$ , womit wir

bestimmt werden kann. Man könnte sagen: Je kleiner die Zustandsmenge, desto genauer die subjektive Bewertung.<sup>1</sup> Je größer die Zustandsmenge, desto weniger läßt sich sinnvoll angeben, welchen subjektiven Wert die betreffende Proposition hat. Genau umgekehrt verhält es sich mit der subjektiven Wahrscheinlichkeitsbewertung: Je größer die Zustandsmenge, desto leichter ist es, ihr eine subjektive Wahrscheinlichkeit zuzuordnen. Der Zusammenhang von propositionalen Präferenzen und Zustandspräferenzen wird uns später noch eingehender beschäftigen und vielleicht manche Fragen aufhellen, die sich hier (etwas verfrüht) aufdrängen.

Es ist oft zweckmäßig, eine Entscheidungssituation i.e.S. in der Gestalt von drei Matrizen folgendermaßen darzustellen:

### 1. Die Konsequenzenmatrix

Für jede Entscheidung  $\sigma_k$  ergibt sich bei dem Umstand  $e_1$  die Konsequenz  $c_{k1}$ . Die Elemente der Konsequenzenmatrix ( $c_{kl}$ ) sind Teilmengen von  $\Xi$ .

1	$e_1$	$e_2$	$e_3$	$e_4$	...
$\sigma_1$	$c_{11}$	$c_{12}$	$c_{13}$	$c_{14}$	
$\sigma_2$	$c_{21}$	$c_{22}$	...	...	
$\sigma_3$	$c_{31}$	...	...	$c_{34}$	...
...	...				

wobei  $\sigma_j \in \Sigma^1$ ,  $e_j \in U$ ,  $U \subset \text{Pot}(\Xi)$ ,  $c_{kl} \in C$ ,  $C \subset \text{Pot}(\Xi)$

und  $\sigma_k(e_1) = c_{k1}$ .

---

das Problem umgehen, einzelne Arten von Propositionen aus dem Definitionsbereich der subjektiven Wahrscheinlichkeitsfunktion herauszunehmen.

<sup>1</sup> Das gilt für zwei Mengen  $\mu$  und  $\mu'$  aus  $\text{Pot}(\Xi)$ , wenn  $\mu \subset \mu' \wedge \#\mu < \#\mu'$  und nicht für beliebige  $\mu, \mu' \in \text{Pot}(\Xi)$ :  $\#\mu < \#\mu'$ .

## 2. Die Bewertungsmatrix

Die Person 1 hat eine subjektive Bewertungsfunktion über  $\Xi$ . Nach welcher Regel sich daraus die subjektive Bewertung der Konsequenzen ergeben sollte, ist nach unserer Konzeption schon Gegenstand der Theorie rationalen Verhaltens. Die meisten Darstellungen der Entscheidungstheorie orientieren sich jedoch primär an Savage und nehmen die Konsequenzenbewertung als gegeben an. Die Wünschbarkeitsfunktion hätte hier als Definitionsbereich die Menge der Konsequenzen. Nach der ersten Stufe des Entscheidungskriteriums stehen uns jedoch ebenfalls Konsequenzenbewertungen zur Verfügung.

Die subjektiven Bewertungen der Konsequenzen bilden die Bewertungsmatrix.

1	$e_1$	$e_2$	$e_3$	$e_4$	...
$\sigma_1$	$0(c_{11})$	$0(c_{12})$	$0(c_{13})$	$0(c_{14})$	
$\sigma_2$	$0(c_{21})$	$0(c_{22})$	...	...	
$\sigma_3$	$0(c_{31})$	...	...	$0(c_{34})$	...
...	...				

## 3. Die Wahrscheinlichkeitsmatrix

Wir haben nicht angenommen, daß die subjektiven Wahrscheinlichkeiten der Umstände einer Entscheidungssituation entscheidungsunabhängig sind. Es ergibt sich daher die Wahrscheinlichkeitsmatrix:

1	$e_1$	$e_2$	$e_3$	$e_4$	
$\sigma_1$	$\rho_1(e_1 \sigma_1)$	$\rho_1(e_2 \sigma_1)$	$\rho_1(e_3 \sigma_1)$	$\rho_1(e_4 \sigma_1)$	...
$\sigma_2$	$\rho_1(e_1 \sigma_2)$	$\rho_1(e_2 \sigma_2)$	...	...	
$\sigma_3$	$\rho_1(e_1 \sigma_3)$	...		$\rho_1(e_4 \sigma_3)$	...
...	...				

Die Wahrscheinlichkeitsbewertungen der Umstände sind

entscheidungsunabhängig, wenn

$\rho_1(e_k|\sigma) = \rho_1(e_k|\sigma')$  für  $\forall \sigma, \sigma' \in \Sigma^1$ , in diesem Fall werden die Wahrscheinlichkeitsbewertungen der Umstände häufig als letzte Zeile an die Konsequenzenmatrix angehängt.

Das Bayessche Entscheidungskriterium empfiehlt diejenige Handlung  $\sigma_k$ , für die

$$u_1(\sigma_k) = \sum_{l=1}^{\#U} \hat{O}_1(c_{kl}) \cdot \rho_1(e_l|\sigma_k) \text{ maximal ist. } *$$

\* Nach Daniel Bernoulli wird das Bayessche Kriterium auch Bernoulli-Prinzip genannt.

Bernoulli hatte als erster die alte Entscheidungsregel der "mathematischen Erwartung" als unvernünftig kritisiert und die heute akzeptierten Argumente für das Gesetz des abnehmenden Grenznutzens vorgebracht. Als Alternative zum mathematischen Erwartungswert (eine mit der Geldmenge lineare Nutzenfunktion) schlug Bernoulli eine Nutzenfunktion des Geldes vor, deren erste Ableitung der Geldmenge umgekehrt proportional ist;

$$\frac{du}{dq} = \frac{1}{q} \Rightarrow \int du = \int \frac{dq}{q} \Rightarrow u(q_2) - u(q_1) = \ln q_2 - \ln q_1,$$

also die typische logarithmische Nutzenfunktion, die auch heute noch in der ökonomischen Theorie weithin angenommen wird.<sup>1</sup>

In dieser Schrift wurde auch zum ersten Mal das 'St. Petersburger Paradoxon' veröffentlicht - so genannt, weil die Schrift von der St. Petersburger Akademie der Wissenschaften herausgegeben wurde.<sup>2</sup>

"Mein sehr verehrter Oheim, der berühmte Nicolaus Bernoulli, Professor beider Rechte, legte einmal dem bekannten Montmart fünf Probleme vor, die man in dem Buche 'Analyse sur les yeux de hazard de M. de Montmart', p.402 findet."<sup>3</sup> Das letzte dieser Probleme ist das St. Petersburger Paradoxon, das Bernoulli mithilfe seiner Theorie der Wertbestimmung von Glücksfällen gelöst zu haben glaubte.<sup>4</sup>

Dieses Paradoxon spricht für eine Beschränktheit der subjektiven Bewertungsfunktionen - darauf kommen wir später noch einmal zurück (§6).

<sup>1</sup> Vgl. D. Bernoulli, *Specimen theoriae novae de mensura sortis* (1738) §§5 und 10 bzw. in der deutschen Übersetzung von Pringsheim (1896), S.27f und 33-37. Dennoch behielt die klassische Wertlehre mit ihrer Gebrauchswert-Tauschwert-Unterscheidung mindestens bis zum Erscheinen der Arbeiten von Jevons, Menger und Walras in den 70er und 80er Jahren des vergangenen Jahrhunderts die Oberhand.

<sup>2</sup> *Commentarii academiae scientiarum imperialis Petropolitanae*. Tomus V, Petrop. 1738, S.175-192.

<sup>3</sup> Bernoulli (1896), S.46.

<sup>4</sup> Vgl. Bernoulli (1896), S.45-60.

$u_1(\sigma_k)$  ist der Erwartungswert der subjektiven Konsequenzenbewertung, wir nennen  $u_1(\sigma_k)$  auch Nutzenerwartungswert von  $\sigma_k$  :  $NEW(\sigma_k)$ .

Damit lautet die Bayessche Definition rationalen Verhaltens : Eine rationale Entscheidung maximiert den Nutzenerwartungswert; und die Bayessche Definition konsistenten Verhaltens: Eine Menge von Entscheidungen einer Person ist genau dann konsistent, wenn es eine Wahrscheinlichkeitsfunktion  $\rho_1$  und eine Nutzenfunktion  $u_1$  gibt, so daß jede der Entscheidungen aus dieser Menge den Nutzenerwartungswert maximiert.

Debreu<sup>1</sup> hat bewiesen, daß das Postulat nutzenmaximierenden Verhaltens im Sicherheitsfall<sup>2</sup> aus zwei Postulaten deduzierbar ist, nämlich der Konnexität der individuellen Präferenzordnung im Güterraum  $\mathbb{R}^m$  (der unserem  $\Xi$  entspricht) und der Kontinuität der individuellen Präferenzordnungen.

- (1) Konnexität: Die schwachen individuellen Präferenzen konstituieren eine vollständige Ordnung über  $\Xi$ .
- (2) Kontinuität: Konvergiert eine Folge von Zuständen  $\xi_1, \xi_2, \xi_3, \dots$  gegen  $\xi_0$  (bei Debreu sind die Zustände (=Situationen) Punkte im m-dimensionalen Güterraum  $\mathbb{R}^m$ ) und eine andere Folge von Zuständen  $\xi_1', \xi_2', \xi_3', \dots$  gegen  $\xi_0'$  und ist  $\zeta_k \succ_1 \xi_k'$  für  $\forall k = 1, 2, 3, \dots$ , so gilt:  $\xi_0 > \xi_0'$ .<sup>3</sup>

Wenn das Entscheidungsverhalten eines Individuums den Postulaten (1) und (2) entspricht, so maximiert jede seiner Entscheidungen eine wohldefinierte ordinale oder kardinale Nutzenfunktion  $u_1$ . Allerdings ist fraglich, ob man

---

<sup>1</sup> Vgl. Debreu (1959), S.55-59.

<sup>2</sup>  $\forall e \in U \quad \forall \sigma \in \Sigma^1 : \rho_1(e|\sigma) \in \{0, 1\}$ .

<sup>3</sup> Diese beiden Axiome werden bisweilen als 'basic utility axioms' angeführt, vgl. J.C. Harsanyi, *Bayesian decision theory, rule utilitarianism, and Arrow's impossibility theorem*, in: H.W. Brock, *Game theory, social choice, and ethics*, Teil 2, Theory and Decision 11 (1979) 289-317.

diesen beiden Postulaten einen begründenden Status einräumen kann.

Im Risikofall sind zwei weitere Rationalitätspostulate erforderlich, um das Bayessche Postulat rationaler Entscheidung zu deduzieren:

(3) Die probabilistische Äquivalenz von Lotterien

Es wird gefordert, daß der Entscheidungsträger dann indifferent zwischen zwei Lotterien ist, wenn ihre Auszahlungen und Wahrscheinlichkeiten gleich sind, unabhängig von der Art der Zufallsereignisse und dem verwendeten Zufallsmechanismus.<sup>1</sup>

(4) Die Wahl der dominanten Entscheidung

Sind die Wahrscheinlichkeiten der Umstände entscheidungsunabhängig und sind die Konsequenzen der Umstände bei einer Entscheidung in jedem Fall für den Entscheidungsträger nützlicher als bei einer anderen Entscheidung, wird die erste Entscheidung gewählt.

Diese vier Postulate zusammen erlauben ausschließlich NEW-maximierendes Verhalten.

Ein Verhalten ist konsistent, wenn es eine Wahrscheinlichkeitsfunktion und eine Nutzenfunktion gibt, die jede Entscheidung zu einer NEW-maximierenden Entscheidung machen.

Bayesianische Verhaltenskonsistenz kann allerdings auch solches Entscheidungsverhalten umfassen, das irrational bezüglich der Bayesschen Entscheidungstheorie ist. Von einer Person  $\iota$  mit einer subjektiven Wahrscheinlichkeitsfunktion  $\rho_\iota$  und einer Bewertungsfunktion  $\hat{U}_\iota$  (Bernoullinutzen) erwarten wir ein Verhalten, das den Nutzen-erwartungswert der Entscheidungsalternativen maximiert.

---

<sup>1</sup> Eine Lotterie ist ein Zufallsexperiment mit bekannter Wahrscheinlichkeitsverteilung, das je nach Ergebnis eine bestimmte Auszahlung hat:  $\langle \rho(e_1) | \pi(e_1), \rho(e_2) | \pi(e_2), \rho(e_3) | \pi(e_3), \dots, \rho(e_n) | \pi(e_n) \rangle$ , wobei  $\rho(e_k | e_l) = 0$  für  $k, l \in \{1, 2, \dots, n\} \wedge k \neq l$  und  $\sum_{k=1}^n \rho(e_k) = 1$ .

Wenn wir einmal davon ausgehen, daß sowohl die Wahrscheinlichkeitsfunktion  $\rho_1$  wie die Bewertungsfunktion  $\hat{u}_1$  unabhängig vom tatsächlichen Entscheidungsverhalten der Person bestimmt werden kann, so ist es durchaus möglich, daß das tatsächliche Entscheidungsverhalten dieser Person konsistent in dem Sinne ist, Jaß es eine Wahrscheinlichkeit  $\rho_1^*$  und eine Nutzenfunktion  $\hat{u}_1^*$  gibt, die ihr tatsächliches Entscheidungsverhalten als NEW-maximierend auszeichnet, aber ohne daß  $\rho_1^*$  mit  $\rho_1$  und  $\hat{u}_1^*$  mit  $\hat{u}_1$  übereinstimmen. M.a.W.:  $\iota$  verhält sich bezüglich seiner tatsächlichen Wahrscheinlichkeits- und Bewertungsfunktionen nicht NEW-maximierend, obwohl sein Entscheidungsverhalten konsistent im Sinne der Bayesschen Entscheidungstheorie ist.

Es gibt eine Reihe weiterer Gründe, aus denen heraus konsistentes Verhalten dennoch irrational sein kann. Man spricht etwa auch dann von einer irrationalen Entscheidung, wenn diese zwar bezüglich der subjektiven Wahrscheinlichkeiten und Bewertungen der Person NEW-maximierend ist, aber diese Wahrscheinlichkeitsbewertungen selbst wiederum als irrational empfunden werden (z.B. angesichts der Erfahrung dieser Person über die relative Häufigkeit eines Ereignisses in der Vergangenheit). Das kann natürlich auch dann der Fall sein, wenn die subjektiven Wahrscheinlichkeiten den Kolmogoroff-Axiomen genügen.<sup>1</sup>

Auch eine Irrationalität der Konsequenzenbewertung ist nicht ausgeschlossen. Die sog. 'kognitive Psychologie' befaßt sich ja zentral mit der Rationalität von Wünschen.

---

<sup>1</sup> Shimony's Rationalitätsbedingung ist demgegenüber weit schwächer. Sie verlangt von einer rationalen Person daß dessen Glaubensgrad ('Glaubensgrad' wird von Shimony dem maximalen Wettquotienten gleichgesetzt) der jeweiligen Hypothese (hier für das Eintreffen des jeweiligen entscheidungsrelevanten Ereignisses) dergestalt ist, daß es keine Klasse von Hypothesen gibt, so daß eine Wette auf das Eintreffen dieser Hypothese mit Sicherheit einen Verlust bedeutet oder mit Sicherheit keinen Gewinn bringt, obwohl ein Verlust möglich ist. Vgl. A. Shimony, *Coherence and the axiom of confirmation*, Journal of symbolic logic 20 (1955) 1-28.



R.B. Brandt stellt in seiner Arbeit *A theory of the good and the right*<sup>1</sup> einige Typen irrationaler Bewertungen, 'irrtümlischer' Wünsche, Zielsetzungen, Aversionen dar.

Irrational wird man ein Entscheidungsverhalten auch dann nennen, wenn der Entscheidungsträger die Informationsmöglichkeiten nicht nutzt, deren 'Kosten' unbedeutend im Vergleich zu dem möglichen Schaden einer Fehleinschätzung sind.<sup>2</sup> Auch dieser Aspekt einer rationalen Entscheidung ist unabhängig von der Konsistenz des Entscheidungsverhaltens.<sup>3</sup>

---

<sup>1</sup> Vgl. Brandt (1979), Kap.VI, vgl. dazu a. Stegmüller (1977)

<sup>2</sup> Vgl. dazu auch P. Gärdenfors, *Forecasts, decisions and uncertain probabilities*, Erkenntnis 14 (1979) 159-181.

<sup>3</sup> Konsistentes Verhalten braucht nicht rational zu sein, aber man kann von einem adäquaten Konsistenzbegriff erwarten, daß rationales Verhalten auch konsistent ist. Es ist fraglich, ob der Bayessche Konsistenzbegriff dem gerecht wird, wie die Kritik von Allais am Bayesianismus zeigt. Vgl. M. Allais, (1953) und zur Diskussion dazu Allais/Hagen (1979). Ausgangspunkt ist die Beobachtung, daß Versuchspersonen in Wahlexperimenten systematisch der Bayesschen Konsistenzforderung zuwiderhandeln (vgl. dazu a. Schneeweiß (1967), S.79). Um Allais' Kritik gerecht zu werden, scheint es mir notwendig zu sein, die übliche Interpretation der BRLT-Zuordnung zu modifizieren; die Identifikation von Risikoscheu und abnehmen-dem Grenznutzen im Bayesschen Entscheidungskriterium läßt sich nicht aufrechterhalten (vgl. Raiffa (1978), S.57f u. S.66-70). Ein eigener Parameter für das Maß der Risikoscheu (bzw. Risikobereitschaft) verlangt allerdings, das Axiom der Lotterien-substituierbarkeit aufzugeben. Besonders bei positionellen Gütern (zur Theorie der positionellen Güter vgl. Hirsch (1976)) ist es durchaus vernünftig nicht-konvexe Nutzenfunktionen anzunehmen. Bei entsprechender Risikoscheu wäre in einem solchen Fall dann dennoch eine konvexe BRLT-Zuordnung möglich. Damit hätte man schon für Risikosituationen ohne Interaktion eine nicht-bayesianische Konsistenzdefinition individuellen Entscheidungsverhaltens, die manche Mängel der v. Neumann/Morgensternschen Nutzentheorie beheben könnte (vgl. dazu a. Dreyfus/Dreyfus (1978), Arnaszus (1974), Höffe (1975a)).

### 3. RATIONALES VERHALTEN IN UNSICHERHEITSSITUATIONEN OHNE INTERAKTION

Nachdem wir uns im letzten Abschnitt mit individuell rationalem Verhalten in solchen Situationen befaßt haben, in denen es sinnvoll ist, eine (entscheidungsbedingte) Wahrscheinlichkeitsbewertung der Umstände -  $\rho_1(e|\sigma)$ ,  $e \in U$  - vorzunehmen, geht es jetzt um die Frage, welches Verhalten in solchen Situationen individuell rational ist, in denen eine subjektive Wahrscheinlichkeitsbewertung der Umstände unmöglich ist. Solche Entscheidungssituationen sind Unsicherheitssituationen im engeren Sinne. Eine Entscheidung erfolgt in einer Sicherheitssituation, wenn der Entscheidungsträger weiß, welcher Umstand vorliegt. Hat er nur eine subjektive Wahrscheinlichkeitsbewertung der Umstände, liegt eine Risikosituation vor. Fühlt er sich auch außerstande für eine (subjektive) Wahrscheinlichkeitsbewertung der Umstände, so handelt es sich um eine Unsicherheitssituation im engeren Sinne.

Sicherheits- und Risikosituationen unterscheiden sich nur in den Werten der (bedingten) subjektiven Wahrscheinlichkeitsfunktion über  $U$ . Kommt im Wertebereich eine Eins vor, handelt es sich um eine Sicherheitssituation, kommt darin keine Eins vor, um eine Risikosituation. In einer Unsicherheitssituation dagegen gibt es überhaupt keine subjektive Wahrscheinlichkeitsfunktion. Daher haben wir die Sicherheitssituation, ohne sie zu erwähnen, schon in §2 behandelt, da wir den Wertebereich von  $0$  nicht auf  $[0,1)$  eingegrenzt haben, während das Modell für die Analyse einer Unsicherheitssituation nicht anwendbar ist.

In der Anwendung gibt es einen fließenden Übergang von der Risiko- zur Sicherheitssituation, der sich in den Entscheidungskriterien darin widerspiegelt, daß die Sicherheitssituation zum Grenzfall wird.

Wesentlich schwieriger ist es, die Beziehung von Risikosituationen (einschließlich des Sicherheitsfalls) einerseits und Unsicherheitssituationen andererseits zu klären.

In einer Entscheidungssituation, in der es sinnvoll ist, den Umständen objektive, aber unbekannte Wahrscheinlichkeiten zuzuordnen, kann eine Person Vermutungen darüber anstellen, in welchem Bereich sich diese Wahrscheinlichkeiten bewegen. Glaubt die Person, die objektiven Wahrscheinlichkeiten zu kennen - z.B. bei der Wahl zwischen Lotterien -, handelt es sich um eine reine Risikosituation; ist sie sich über die objektiven Wahrscheinlichkeiten jedoch völlig im unklaren, ist es eine Unsicherheitssituation.<sup>1</sup>

Stellen wir uns vor, die Welt wäre streng deterministisch, d.h. wenn zu einem Zeitpunkt  $t_0$  der Weltzustand  $\xi_0$  besteht, dann sind damit auch alle anderen Weltzustände  $\xi(t)$  eindeutig festgelegt (für alle  $t$ ,  $t > t_0$  und  $t < t_0$ ), und in der Beschreibung der Zustände kommt keine statistische Proposition notwendig vor (m.a.W.: Jede statistische Proposition ist durch eine Menge von nicht-statistischen Propositionen substituierbar). Gibt es in einer solchen Welt objektive Wahrscheinlichkeiten?

Sicher gibt es jedenfalls auch in einer solchen Welt relative Häufigkeiten. Aber Wahrscheinlichkeiten sind dann dennoch 'Ausdruck mangelnden Wissens'. Wahrscheinlichkeitsaussagen sprechen in einer solchen Welt offensichtlich nicht über die Natur, sondern über eine subjektive Erwartungshaltung angesichts vorliegender Informationen. Wie reichhaltig die Information ist, ist nicht entscheidend, wir haben jeweils relativ zu ihr eine Er-

---

<sup>1</sup> Man sieht, daß metatheoretische Fragen der Wahrscheinlichkeitstheorie einen Einfluß darauf haben, welches Kriterium man zur Bestimmung von Unsicherheitssituationen akzeptiert, denn ein Wahrscheinlichkeitssubjektivist würde diesen Klärungsversuch vermutlich zurückweisen.

wartungshaltung, und die schlägt sich in entsprechenden subjektiven Wahrscheinlichkeiten nieder. Für einen Subjektivisten ist daher die Unterscheidung Risikosituation und Unsicherheitssituation weniger selbstverständlich. Denn - so könnte dieser entgegenhalten - einen 'Grad des Für-wahr-haltens' gibt es immer.

Eine mögliche Konsequenz des Wahrscheinlichkeitssubjektivismus ist es daher, an dem Bayesschen Entscheidungskriterium festzuhalten und für Situationen, in denen es mangels Information keinen Grund gibt, eines der Ereignisse für wahrscheinlicher als ein anderes zu halten, gleiche subjektive Wahrscheinlichkeiten anzusetzen. Damit erhält man das Laplace-Kriterium:<sup>1</sup>

Bevorzuge eine Entscheidung  $\sigma_\kappa$  gegenüber  $\sigma_\lambda$ , gdw.

$$\sum_{e \in U} \hat{O}_1(\sigma_\kappa(e)) > \sum_{e \in U} \hat{O}_1(\sigma_\lambda(e)).$$

Es gibt dabei zwei mögliche Interpretationen des Laplace-Kriteriums:

---

<sup>1</sup> Das Argument des unzureichenden Grundes, auch "Indifferenzprinzip" genannt, geht auf Jacob Bernoulli zurück. Es kann mit der Wahrscheinlichkeitsinterpretation minimaler Wettquotienten gestützt werden. Die Explikation des Wahrscheinlichkeitsbegriffes durch Wettquotienten geht bis ins 16. Jahrhundert auf Gerolamo Gardano zurück. Zur Theorie der rationalen Wettquotienten vgl. B. de Finetti, *Initial probabilities. A prerequisite for any valid deduction*, *Synthese* 20 (1969) 2-16; R.B. Braithwaite, *Why is it reasonable to base a betting rate upon an estimate of chance?*, in Bar-Hillel (1965) 263-274. Die Bertrandschen Paradoxa (*Calcul des probabilités*, Paris 1889) zeigen jedoch, daß das Indifferenzprinzip kontraintuitive Konsequenzen hat. Auch der Einwand, Bertrand bringe in seinen Beispielen unterschiedliche Aufgabenstellungen (vgl. z.B. Gnedenko (1978), S. 21-34) ist nicht stichhaltig, wie die Übertragung auf physikalische Beispiele deutlich machen kann. Zur Kritik des Indifferenzprinzips vgl. a. Keynes (1921), Kap.4. Auch Carnap versuchte, das Prinzip zu umgehen (vgl. Carnap (1962)), s.a. Arrow (1951) und Savage (1954).

- (1) Die (hier angeführte) 'bayesianische' Interpretation:  
Wenn man keinen Grund hat, einen der Umstände (unabhängig von der gewählten Entscheidung) für wahrscheinlicher als einen anderen zu halten, hält man die Umstände offensichtlich im gleichen Grade für wahr, und in diesem Falle ist die Bayessche Entscheidungsregel äquivalent mit dem Laplace-Kriterium.
- (2) Wenn man keine Wahrscheinlichkeitsinformation hat, muß man seine Entscheidung ausschließlich auf die subjektive Bewertung der Konsequenzen stützen. Es gibt keinen Grund, eine dieser Konsequenzen besonders zu berücksichtigen, daher ist die Nutzensumme der möglichen Konsequenzen einer Handlung ausschlaggebend.

Daneben sind weitere Kriterien für rationale Entscheidungen in Unsicherheitssituationen vorgeschlagen worden.

Das Maximin-Kriterium<sup>1</sup>:

Bevorzuge eine Entscheidung  $\sigma_k$  gegenüber  $\sigma_\lambda$  gdw.

$$\min(\{0_1(\sigma(e))\}) > \min(\{0_1(\sigma(e))\}), e \in U$$

bzw.

Wähle diejenige Entscheidung, für die der Nutzen der ungünstigsten Folge maximal ist.

(I)

	$e_1$	$e_2$	$e_3$
$\sigma_1$	2	8	12
$\sigma_2$	4	5	4
$\sigma_3$	15	6	0

$$\text{Maximin}(\sigma_1) = 2$$

$$\text{Maximin}(\sigma_2) = 4$$

$$\text{Maximin}(\sigma_3) = 0$$

$$\sigma_2 > \sigma_1 > \sigma_3$$

Weniger überzeugend wirkt dieses Kriterium in Entscheidungssituationen, in denen geringe Unterschiede in den Bewertungen der ungünstigsten Konsequenzen, aber ansonsten große Unterschiede in den Konsequenzenbewertungen vorkommen.

---

<sup>1</sup> Auch Wald-Regel genannt, vgl. A. Wald, *Statistical decision functions* New York 1950.

(II)

	e <sub>1</sub>	e <sub>2</sub>
σ <sub>1</sub>	0,99	10
σ <sub>2</sub>	1	

$$\text{maximin}(\sigma_2) > \text{maximin}(\sigma_1).$$

Die sich daraus ergebende Kritik ist die Motivation für das Minimax-Verlustkriterium<sup>1</sup>:

Bevorzuge eine Entscheidung  $\sigma_\kappa$  gegenüber  $\sigma_\lambda$  gdw.

$$-\max(\{\max(\{0(\sigma_\mu(e_\omega)) - 0(\sigma_\kappa(e_\omega))\})_\mu\})_\omega >$$

$$-\max(\{\max(\{0(\sigma_\mu(e_\omega)) - 0(\sigma_\lambda(e_\omega))\})_\mu\})_\omega$$

bzw.:

Wähle diejenige Entscheidung, für die der maximale Verlust (gegenüber einer anderen möglichen Entscheidung) minimal ist.

(III)

	e <sub>1</sub>	e <sub>2</sub>
σ <sub>1</sub>	0,99	10
σ <sub>2</sub>	1	1

→  
Verlust-  
Matrix

	e <sub>1</sub>	e <sub>2</sub>
σ <sub>1</sub>	0,01	0
σ <sub>2</sub>	0	9

$$\text{MIMAV}(\sigma_1) = -0,01; \quad \text{MIMAV}(\sigma_2) = -9$$

(IV)

	e <sub>1</sub>	e <sub>2</sub>	e <sub>3</sub>
σ <sub>1</sub>	3	7	4
σ <sub>2</sub>	10	2	6
σ <sub>3</sub>	1	4	12

	e <sub>1</sub>	e <sub>2</sub>	e <sub>3</sub>
σ <sub>1</sub>	7	0	8
σ <sub>2</sub>	0	5	6
σ <sub>3</sub>	9	3	0

$$\text{MIMAV}(\sigma_1) = -8$$

$$\text{MIMAV}(\sigma_2) = -6$$

$$\text{MIMAV}(\sigma_3) = -9$$

( Verlustmatrix )

$$\sigma_2 > \sigma_1 > \sigma_3$$

Während das Minimax-Kriterium in (II) die Entscheidung  $\sigma_2$  als rational auszeichnet, empfiehlt das Minimax-Verlust-Kriterium  $\sigma_1$ , was der Intuition eher entspricht.

<sup>1</sup> Englisch: minimax-regret-criterion oder minimax-risk-criterion; auch Savage-Kriterium genannt, vgl. Savage (1951).

Während das Maximin-Kriterium ausschließlich die schlechteste Folge berücksichtigt, berücksichtigt das Maximax-Kriterium ausschließlich die beste Folge einer Entscheidung:

Bevorzuge eine Entscheidung  $\sigma_\kappa$  gegenüber  $\sigma_\lambda$  gdw.

$$\max(\{0(\sigma_\kappa | e_\mu)\})_\mu > \max(\{0(\sigma_\lambda | e_\mu)\})_\mu$$

bzw.

Wähle diejenige Entscheidung, für die der Nutzen der günstigsten Folge maximal ist.

Zwischen diesen beiden Extremen der ausschließlichen Berücksichtigung der schlechtesten bzw. der besten Folge im Minimax- bzw. Maximax-Kriterium, gibt es ein Kontinuum von Entscheidungsregeln, das Hurwicz-Kriterium<sup>1</sup>:

Bevorzuge eine Entscheidung  $\sigma_\kappa$  gegenüber  $\sigma_\lambda$  gdw.

$$r \cdot \min(\{0(\sigma_\kappa | e_\mu)\})_\mu + (1-r) \cdot \max(\{0(\sigma_\kappa | e_\mu)\})_\mu > \\ > r \cdot \min(\{0(\sigma_\lambda | e_\mu)\})_\mu + (1-r) \cdot \max(\{0(\sigma_\lambda | e_\mu)\})_\mu, \text{ wobei } r \in [0,1]$$

Zur Bestimmung von  $r$  (Optimismus-Pessimismus-Index) eignet sich eine Entscheidungssituation, bei der aus der Indifferenz des Entscheidenden zwischen den Entscheidungen  $\sigma_1$  und  $\sigma_2$  für einen bestimmten Wert  $c$  von  $x$  auf dessen Index  $r$  geschlossen werden kann:

(V)	$e_1$	$e_2$
$\sigma_1$	$0(a)$	$0(b)$
$\sigma_2$	$0(x)$	$0(x)$

o.B.d.A.:  $0(a) > 0(b)$ .

$$r \cdot 0(a) + (1-r) \cdot 0(b) = r \cdot 0(x) + (1-r) \cdot 0(x) \Rightarrow r = \frac{0(x) - 0(b)}{0(a) - 0(b)}$$

Für jede beliebige einzelne Entscheidung läßt sich - vorausgesetzt, sie verletzt nicht das Dominanzkriterium - ein Optimismus-Pessimismus-Index angeben, der sie nach der Hurwicz-Regel bei gegebener Bewertungsfunktion  $0$  als rational auszeichnet. Jedoch gibt es nicht für jede

<sup>1</sup> Vgl. Luce/Raiffa (1957), S.282ff.

Präferenzrelation über die Menge der Entscheidungen einen entsprechenden durchgängigen Indexwert.

Entscheidungspräferenzen, die sich nach dem Minimax-Risiko-Kriterium richten, haben z.B. keinen gemeinsamen Indexwert, der sie nach der Hurwicz-Regel als rational auszeichnet.

Handelt eine Person nach dem Laplace-Kriterium, hat sie einen Optimismus-Pessimismus-Index von 0,5, denn sie ist indifferent zwischen  $\sigma_1$  und  $\sigma_2$  gdw. in der Entscheidungssituation der Matrix V gilt:

$$0,50(a) + 0,50(b) = 0,50(x) + 0,50(x).$$

$$\text{Damit ist } x = 0,5 \cdot (0(a) + 0(b)) \text{ und } r = \frac{0(x) - 0(b)}{0(a) - 0(b)} = 0,5.$$

Handelt eine Person nach dem Maximax-Kriterium, hat sie einen Optimismus-Pessimismus-Index von 1, denn sie ist indifferent zwischen  $\sigma_1$  und  $\sigma_2$ , gdw.  $0(x) = 0(a)$ .

Handelt eine Person nach dem Maximin-Kriterium, hat sie einen Optimismus-Pessimismus-Index von 0, denn sie ist indifferent zwischen  $\sigma_1$  und  $\sigma_2$  gdw.  $0(x) = 0(b)$ .

Eine Entscheidung gemäß dem Laplace-Kriterium ist also rational bezüglich eines Optimismus-Pessimismus-Index von 0,5, und in gleicher Weise ist eine Entscheidung gemäß dem Maximin-Kriterium rational bezüglich eines Index von 0, etc..

Da ein behavioristisches Modell mit zeitabhängigem Indexwert jede beliebige tatsächlich getroffene Entscheidung einer Person für rational im Hurwicz-Sinn erklären kann, sind entweder Annahmen über eine gewisse zeitliche Invarianz nötig, oder man wendet das Rationalitätskriterium nicht auf einzelne Entscheidungen, sondern auf Relationen von Entscheidungen an. Konkretes Entscheidungsverhalten legt nur dann Präferenzordnungen über Entscheidungen fest, wenn man eine gewisse zeitliche Invarianz der individuellen Präferenzen voraussetzen kann.



Milnor hat gezeigt, daß diese Kriterien rationaler Entscheidung in Unsicherheitssituationen gegen bestimmte, intuitiv überzeugende, Minimalbedingungen verstoßen.<sup>1</sup> Nun mag die reine Unsicherheitssituation als ein Grenzfall betrachtet werden, in dem die Informationsbasis nicht ausreicht, um individuelle Verhaltensrationalität zu bestimmen. Da reale Entscheidungssituationen jedoch im allgemeinen weder als reine Unsicherheitssituation noch als reine Risikosituation interpretiert werden können, müßte es Ziel der Theorie rationaler Entscheidung sein, ein generelles Rationalitätskriterium zu entwickeln als dessen Grenzfälle sich das Entscheidungskriterium für Risiko und andererseits das Entscheidungskriterium für Unsicherheit ergeben. Ein solches generelles Kriterium ließe sich auch sehr viel besser an den Desiderata unseres intuitiven Vorverständnisses prüfen.

Gärdenfors<sup>1</sup> hat einen ersten Schritt in diese Richtung gemacht.

Jeder Umstand erhält nicht eine Wahrscheinlichkeit, sondern ein Wahrscheinlichkeitsintervall  $I_{\rho}(e|\sigma)$  zugeordnet.  $\forall e \in U: I_{\rho}(e|\sigma) \subset [0,1]$ .

Im Risikofall ist die Weite dieser Intervalle  $\max(I_{\rho}(e|\sigma)) - \min(I_{\rho}(e|\sigma))$  auf Null, oder fast Null geschrumpft. Im Unsicherheitsfall dehnt sich dieses Intervall bis auf die maximale Weite  $\max(I_{\rho}(e|\sigma)) - \min(I_{\rho}(e|\sigma)) = 1$  aus.<sup>2</sup>

---

<sup>1</sup> vgl. Milnor (1954). Von den zehn Metakriterien rationaler Entscheidung, die Milnor aufstellt, können nicht alle eine Selektionsfunktion beanspruchen, die übrigen dienen eher der Charakterisierung konkurrierender Kriterien.

<sup>2</sup> vgl. P. Gärdenfors, *Forecasts, decisions and uncertain probabilities*, Erkenntnis 14 (1979), S. 159-187.

Anders als Gärdenfors nehmen wir auch hier *handlungsbedingte* Wahrscheinlichkeitsintervalle an.

Als ein generelles Kriterium rationaler Entscheidung wird das 'Maximin-Kriterium für Nutzenerwartungswerte' vorgeschlagen:

Ein möglicher Nutzenerwartungswert in einer Entscheidungssituation zwischen den beiden Grenzfällen wäre:

$$\sum_{e \in U} \hat{0}_1(\sigma(e)) \cdot \rho(e|\sigma), \text{ wobei } \forall e \in U : \rho(e|\sigma) \in I_\rho(e|\sigma)$$

Das Maximin-Kriterium für Nutzenerwartungswerte lautet: Diejenige Entscheidung  $\sigma$  ist rational, für die der minimale Nutzenerwartungswert maximal ist.

Damit ist die Forderung erfüllt, daß sich die Kriterien rationalen Verhaltens in reinen Risiko- und reinen Unsicherheitssituationen als Grenzfälle eines allgemeinen Rationalitätskriteriums ergeben, denn es gilt:

$$(1) \quad \lim_{\|\rho\|^* \rightarrow 0} \min(\{\sum_{e \in U} \hat{0}_1(\sigma(e)) \cdot \rho(e|\sigma)\})_{\rho(e|\sigma) \in I_\rho(e|\sigma)} = \sum_{e \in U} \hat{0}_1(\sigma(e)) \cdot \rho(e|\sigma)$$

Wobei:

$$\|I_\rho(e|\sigma)\| = \max(I_\rho(e|\sigma)) - \min(I_\rho(e|\sigma)) ; ^1$$

$$\|I_\rho\|^* = \max(\{\|I_\rho(e|\sigma)\|\})_{e \in U}$$

und

$$(2) \quad \lim_{\|\rho\|^o \rightarrow 1} \min(\{\sum_{e \in U} \hat{0}_1(\sigma(e)) \cdot \rho(e|\sigma)\})_{\rho(e|\sigma) \in I_\rho(e|\sigma)} = \min(\{\hat{0}(\sigma(e))\})_{e \in U} ; \text{ wobei:}$$

$$\|I_\rho\|^o = \min(\{\|I_\rho(e|\sigma)\|\})_{e \in U}$$

<sup>1</sup> Gibt es in  $I_\rho(e|\sigma)$  kein kleinstes Element, sei die größte untere Schranke:  $\inf(I_\rho(e|\sigma))$  eingesetzt - ebenso für  $\max(I_\rho(e|\sigma)) : \sup(I_\rho(e|\sigma))$

Die Wahrscheinlichkeitsintervalle können als Grenzfall von Wahrscheinlichkeiten zweiter Ordnung interpretiert werden: Meine (subjektive) Wahrscheinlichkeit dafür, daß sich  $\rho(e|\sigma)$  außerhalb  $I_\rho(e|\sigma)$  befindet, ist Null und dafür, daß  $\rho(e|\sigma) \in I_\rho(e|\sigma)$ , ist meine (subjektive) Wahrscheinlichkeit Eins. Solche Wahrscheinlichkeiten zweiter Ordnung machen natürlich nur Sinn, wenn  $\rho(e|\sigma)$  als objektive Wahrscheinlichkeit interpretiert wird.

Komplizierter wird es, wenn die Wahrscheinlichkeiten zweiter Ordnung stetig sind, was der Normalfall sein dürfte. Gärdenfors macht auch dazu einen interessanten Vorschlag. Die Wahrscheinlichkeitsintervalle werden ersetzt durch sog. Vertrauensintervalle. Für jeden Umstand  $e$  wird das Vertrauensintervall  $\tilde{I}_\rho(\sigma|e)$  so fortgesetzt, daß die Wahrscheinlichkeit (zweiter Ordnung) dafür, daß die tatsächliche (objektive) Wahrscheinlichkeit von  $e$  in  $\tilde{I}_\rho(\sigma|e)$  liegt, gleich einer Größe  $\alpha \in [0,1]$  ist. Je größer  $\alpha$  gewählt wird, desto größer die Risikoscheu der betreffenden Person. Es läßt sich zeigen, daß im Grenzfall von  $\alpha=0$  dieses ' $\alpha$ -Maximin-Kriterium' dem Bayesschen Kriterium äquivalent ist, und im Grenzfall von  $\alpha=1$  (falls alle Wahrscheinlichkeiten zweiter Ordnung für beliebige objektive Wahrscheinlichkeiten größer Null sind) dem üblichen Maximin-Kriterium entspricht.

Interessant ist, daß damit Kriterien wie das Maximax- und das Hurwicz-Kriterium für  $r > 0,5$  nicht erfaßt werden.

Dieser Ansatz einer generellen Konzeption rationalen Verhaltens im Kontinuum der Entscheidungssituationen zeigt m.E. den Weg, der aus der Aporie der Theorie rationaler Entscheidung in Unsicherheitssituationen herausführen könnte. Und mit dieser Skizze wollen wir es hier bewenden lassen.

#### 4. INTERAKTIONSSITUATIONEN

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 - etwas irreführend - als Spieltheorie bezeichnet. In §2 und §3 haben wir uns mit zwei Grenzfällen von Entscheidungssituationen befaßt, den Risiko- und den Unsicherheitssituationen. Wir haben gesagt, man könne sich zwischen diesen beiden Grenzfällen ein Kontinuum von Entscheidungssituationen vorstellen, wobei ein solches Kontinuum unter einem rein entscheidungstheoretischen Aspekt (d.h. ohne Berücksichtigung von Interaktionsproblemen) erschöpfend wäre: jede beliebige Entscheidungssituation hat auf diesem Kontinuum seinen Platz.

Der Spieltheorie und der Entscheidungstheorie geht es um die individuelle rationale Entscheidung. Entscheidungstheoretisch wird die rationale Entscheidung relativ zu der gegebenen subjektiven Bewertungsfunktion der Konsequenzen und der subjektiven bedingten Wahrscheinlichkeit der Umstände bestimmt. Von welcher Art diese Ereignisse sind, ist dabei zunächst nicht wesentlich. In einer Interaktionssituation sind die 'Ereignisse' die Entscheidung anderer Personen. Auch diese haben eine subjektive Wahrscheinlichkeit, warum sollte also nicht die Spieltheorie als Theorie rationaler Entscheidung in Interaktionssituationen eine schlichte Spezialisierung der generellen Theorie rationalen Verhaltens, der Entscheidungstheorie sein?

Und tatsächlich kann die gleiche (reale) Entscheidungssituation häufig sowohl entscheidungs- wie spieltheoretisch analysiert werden.

Ein Beispiel:

Gegenstand sei die Entscheidung eines Kapitalanlegers. Welche seiner möglichen Entscheidungen rational ist, hängt davon ab, welche Inflationsrate zu erwarten ist, welche Aktien in der nächsten Zeit in welchem Umfang verkauft werden, etc. Alle diese Umstände sind im wesentlichen Folgen menschlichen Handelns. Der Kapitalanleger wird für diese Umstände bestimmte Wahrscheinlichkeiten haben, und abhängig von der Gestalt der Graphen seiner Wahrscheinlichkeitsfunktionen zweiter Ordnung nimmt diese Entscheidungssituation ihren Platz irgendwo zwischen Risiko- und Unsicherheitssituationen ein. Einer rein entscheidungstheoretischen Bestimmung steht also nichts im Wege.

Zugleich jedoch könnte man die Situation z.B. als ein n-Personen-Spiel behandeln, mit k Kapitalanlegern, a Aktienkäufern ... und versuchen, die individuell rationalen Verhaltensweisen in diesem 'Spiel' zu bestimmen.

Eine besondere Voraussetzung, die Spieltheoretiker i.a. machen (und die erst die Spieltheorie als eigene Disziplin etabliert), ist es nun, nur solche n-Tupel individueller Strategien in den Kandidatenkreis für individuelle Rationalität aufzunehmen, die Gleichgewichtspunkte der Interaktionsstruktur sind, d.h. für die gilt, daß kein Beteiligter ein Interesse daran hat, von seiner Strategie abzurücken, wenn sich alle anderen an ihre Strategien halten. Man kann diese Voraussetzung auch als Konsequenz der potentiellen Publizität spieltheoretischer Rationalitätsempfehlungen interpretieren: Nur solche Rationalitätsempfehlungen werden zugelassen, deren allgemeine Bekanntheit kein Grund sein kann, von ihnen abzuweichen.

Eine Interaktionssituation ist durch fünf Größen gekennzeichnet:

$$\langle K, E, \langle \Sigma^1 \rangle, C, \langle U \rangle \rangle$$

$K$  ist die Menge der an der Interaktion Beteiligten.

$\Xi$  bleibt die Menge der Zustände.

$\langle \sigma^1 \rangle$  ist das  $n$ -Tupel der individuellen Entscheidungspotentiale.

$C$  ist die Menge der Konsequenzen.

$\langle v_1 \rangle$  ist das  $n$ -Tupel subjektiver Bewertungsfunktionen.

Der Zusammenhang von  $O_1$  mit dem Definitionsbereich  $C$  und  $v_1$  mit dem Definitionsbereich  $\Xi$  ist auch hier

$$O_1(c) = \frac{\sum_{\xi \in C} v_1(\xi) \cdot \rho_1(\xi | \langle \sigma^1 \rangle)}{\rho_1(c | \langle \sigma^1 \rangle)}$$

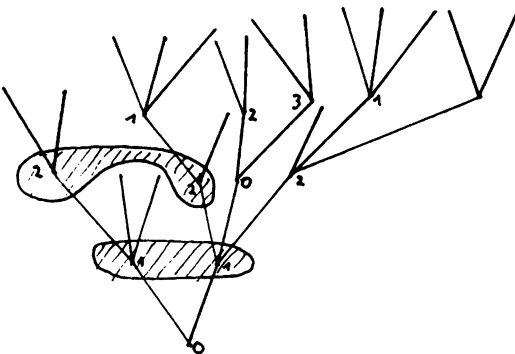
und damit im Sicherheitsfall, d.h. dann wenn die Konsequenzen eindeutig durch die Strategienkombination bestimmt sind:

$$O_1(c) = \sum_{\xi \in C} v_1(\xi) \cdot \rho_1(\xi | \langle \sigma^1 \rangle) .$$

### Darstellungsformen

#### 1. Extensive Form

In der extensiven Form wird das Spiel durch einen Spielbaum sozusagen Zug um Zug dargestellt.



Es gibt Zufallszüge, die durch "0" gekennzeichnet sind. Die einzelnen Zugmöglichkeiten der Spieler ergeben sich aus den Verzweigungspunkten. In unserem Beispiel ist zunächst der erste Spieler am Zug, unabhängig davon,

was der Zufall gewählt hat. Es stehen ihm drei Alternativen offen. Der erste Spieler weiß jedoch nicht, an welcher der beiden Verzweigungen er sich befindet. Dieser Sachverhalt wird durch entsprechende 'Informationsmengen' zum Ausdruck gebracht, in denen jeweils diejenigen Verzweigungspunkte zusammengefaßt werden, bei denen der Spieler, der am Zug ist, nicht weiß, welche von ihnen seiner tatsächlichen Situation entspricht.

Als nächstes ist in unserem Beispiel der zweite Spieler am Zug, außer die erste Person wählt den Zug, der zu einem Verzweigungspunkt des 'Zufalls' führt ...

Ist  $r_1, r_2, r_3, \dots$  die Folge der Spielzüge, so gibt es vier Funktionen dieser Folge:

1. die Spieler

Die Folge  $r_1, r_2, r_3, \dots, r_{k-1}$  bestimmt (naheindeutig) den Spieler  $i_{k \in K}$ , der den k-ten Zug ausführt,

2. die Zug-Alternativen

Die Folge  $r_1, r_2, r_3, \dots, r_{n-1}$  bestimmt die möglichen Züge als k-ten Zug.

3. die Informationsmengen im k-ten Zug  $V_k$

4. die Wahrscheinlichkeitsverteilungen der Zufallszüge im k-ten Zug.

In der extensiven Darstellung sind individuelle Strategien  $\sigma_1$  Funktionen, die jeder Informationsmenge des Spielers einen Zug, der aus der Informationsmenge herausführt, zuordnen. Den Endpunkten  $c$  des Spielbaums werden 'Auszahlungen' (= subjektive Bewertungen)  $\langle \sigma_1(c) \rangle$  zugeordnet.

2. Die Normalform

Eine Interaktionssituation läßt sich auch in Normalform darstellen. In dieser Darstellungsweise beschränkt man sich auf die Angabe der möglichen Strategiekombinationen, ihrer Konsequenzen und die subjektiven Bewertungen dieser Konsequenzen. Sind die Konsequenzen eindeutig

von den Strategiekombinationen bestimmt, gibt es also keine Zufallszüge, kann man die Bewertungsfunktionen  $0_1$  statt über  $C$  auch über  ${}_1\Pi_K\Sigma^1$  gehen lassen. In diesem 'Sicherheitsfall' kann die Interaktionssituation auf  $\langle K, \langle \Sigma^1 \rangle, \langle 0_1 \rangle \rangle$  reduziert werden:

$0_1$  Funktion von  $\prod_{i=1}^{\#K} \Sigma^i$  in  $R$ .

In dieser Darstellungsweise kommen die Details der einzelnen Züge und der Informationsstand der Spieler nicht zum Ausdruck. Wie es zur Wahl einer bestimmten Strategie kommt, bleibt unberücksichtigt. Die Interaktionssituation kann in  $n$ -dimensionalen Matrizen<sup>1</sup> wiedergegeben werden.<sup>2</sup>

<sup>1</sup>Matrixdarstellung einer Drei-Personen-Interaktionssituation in Normalform im Sicherheitsfall

$\sigma_{31}$	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 0_1(c_{111}), 0_2(c_{111}), 0_3(c_{111}) \rangle$	$\langle 0_1(c_{121}), 0_2(c_{121}), 0_3(c_{121}) \rangle$
$\sigma_{12}$	$\langle 0_1(c_{211}), 0_2(c_{211}), 0_3(c_{211}) \rangle$	$\langle 0_1(c_{221}), 0_2(c_{221}), 0_3(c_{221}) \rangle$
$\sigma_{13}$	$\langle 0_1(c_{311}), 0_2(c_{311}), 0_3(c_{311}) \rangle$	$\langle 0_1(c_{321}), 0_2(c_{321}), 0_3(c_{321}) \rangle$

$\sigma_{32}$	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 0_1(c_{112}), 0_2(c_{112}), 0_3(c_{112}) \rangle$	$\langle 0_1(c_{122}), 0_2(c_{122}), 0_3(c_{122}) \rangle$
$\sigma_{12}$	$\langle 0_1(c_{212}), 0_2(c_{212}), 0_3(c_{212}) \rangle$	$\langle 0_1(c_{222}), 0_2(c_{222}), 0_3(c_{222}) \rangle$
$\sigma_{13}$	$\langle 0_1(c_{312}), 0_2(c_{312}), 0_3(c_{312}) \rangle$	$\langle 0_1(c_{322}), 0_2(c_{322}), 0_3(c_{322}) \rangle$

$$\Sigma^1 = \{\sigma_{11}, \sigma_{12}, \sigma_{13}\}, \quad \Sigma^2 = \{\sigma_{21}, \sigma_{22}\}, \quad \Sigma^3 = \{\sigma_{31}, \sigma_{32}\}.$$

$$K = \{1, 2, 3\}.$$

$$c_{1K\lambda} = c(\langle \sigma_{11}, \sigma_{2K}, \sigma_{3\lambda} \rangle).$$

<sup>2</sup> Fußnote s. nächste Seite



## Klassifikation von Interaktionssituationen

- Man unterscheidet kooperative von nicht-kooperativen Interaktionssituationen. In kooperativen Interaktionssituationen sind bindende Abmachungen zwischen den Beteiligten möglich.
- In extensiver Darstellung lassen sich Interaktionssituationen in solche mit vollständiger und solche mit unvollständiger Information unterscheiden.
- Da Probleme der Koalitionsbildung eine besondere Analyse erfordern, unterscheidet man außerdem Zwei-Personen-Spiele, in denen diese Probleme keine Rolle spielen, von Mehr-Personen-Spielen.
- Je nach der Gestalt der Auszahlungsfunktionen spricht man von Konstantsummen-Spielen und von Nicht-Konstantsummenspielen (oft auch 'Nullsummenspiele' und 'Nicht-Nullsummenspiele' genannt). In Konstantsummen-Spielen bedeutet jeder zusätzliche Vorteil des einen einen entsprechenden Nachteil für mindestens einen anderen. Solche Interaktionssituationen nennen wir auch reine Konfliktsituationen. <sup>1</sup>

---

<sup>2</sup> Eine dritte, die kooperative Darstellungsform eignet sich besonders für n-Personen-Spiele, mit  $n \geq 3$ , in denen es um Koalitionsbildung geht, ohne daß die spezifischen Informationsvoraussetzungen der individuellen Strategiewahl interessieren. Zentraler Begriff ist hier die charakteristische Funktion  $v$ , die jeder Teilmenge von Spielern ihren minimalen Nutzen aus dieser Koalitionsbildung zuordnet. Zur Diskussion dieses Begriffes vgl. Shubik (1981). Koalitionsbildungen werden uns in dieser Arbeit jedoch nicht beschäftigen.

<sup>1</sup> Als Lehrbuch eignet sich immer noch am besten R.D. Luce/ H. Raiffa, *Games and decisions*, New York u.a. 1957.

Vgl. a. Schelling (1958, 1963); Burger (1959); Chernoff/Moses (1959); Vogelsang (1963); Klaus (1968); Bühlmann/Loeffel/Nievergelt (1967, 1975); Owen (1971); Vorobjeff (1972); Rapoport (1960, 1976); Shubik (1981).

Zu experimentellen Spielen vgl. Vanacke (1957); Thibaut/Kelly (1959); Wilson (1962); Rapoport/Chammah (1965); Rapoport/Orwant (1965); Rapoport, Amnon /Kahan/Funk/Horowitz (1979),

## 5. RATIONALITÄT KOLLEKTIVER PRÄFERENZEN

"There is something in common between singing romantic songs about an abstract motherland and doing optimization exercises with an arbitrary objective function for a society. While both the activities are worthy, and certainly both are frequently performed, this book, I fear, will not be concerned with either. The subject of our study is the relation between the objectives of social policy and the preferences and aspirations of members of a society."<sup>1</sup>

In diesem Abschnitt sollen die grundlegenden Begriffe normativer subjektivistischer Aggregationstheorien eingeführt werden, um deren methodische Klärung wir uns später bemühen wollen.

Eine Sozialwahlsituation ist durch vier Größen charakterisiert:

$$\langle K, \Xi, g, f \rangle$$

- (1)  $K$  ist eine Menge (Kollektiv, Gruppe) von Personen  
 $\#K \geq 2$
- (2)  $\Xi$  bleibt die Menge der möglichen Welten ( $\#K \in \mathbb{N}$ )
- (3)  $g$  ist eine Funktion von  $K$  in  $\Xi \times \Xi$  (Präferenzstruktur von  $K$ )
- (4)  $f$  ist eine Funktion von  $g$  in  $\Xi \times \Xi$ , wobei  $G$  die Menge der logisch möglichen  $g$  ist (kollektive Präferenzzuordnung)

Eine Standard-Sozialwahlsituation ist darüberhinaus durch folgende Bedingung eingeschränkt:

Die Elemente von  $B_i(g)$  sind konnexe Ordnungsrelationen über  $\Xi \times \Xi$ , d.h.

- a)  $\forall i \in K \forall \xi \in \Xi: \langle \xi, \xi \rangle \in g(i)$  (Reflexivität)
- b)  $\forall i \in K \forall \xi_1, \xi_2, \xi_3 \in \Xi:$   
 $\langle \xi_1, \xi_2 \rangle \in g(i) \wedge \langle \xi_2, \xi_3 \rangle \in g(i) \Rightarrow \langle \xi_1, \xi_3 \rangle \in g(i)$   
(Transitivität)
- c)  $\forall i \in K \forall \xi_1, \xi_2 \in \Xi: \langle \xi_1, \xi_2 \rangle \in g(i) \vee \langle \xi_2, \xi_1 \rangle \in g(i)$   
(Konnexität)

---

<sup>1</sup> Sen (1970a), S.1

Die  $g(i)$  sind als (schwache) individuelle Präferenzrelationen zu interpretieren:

$\langle \xi_1, \xi_2 \rangle \in g(i)$  heißt: Person  $i$  empfindet  $\xi_1$  als mindestens so gut wie  $\xi_2$ ; bzw.  $i$  präferiert  $\xi_1$  gegenüber  $\xi_2$  schwach.

Für  $\langle \xi_1, \xi_2 \rangle \in g(i)$  schreiben wir auch  $\langle \xi_1, \xi_2 \rangle \in R_i$  oder  $\xi_1 \succsim_i \xi_2$ .

Die starke Präferenz  $P_i$  läßt sich dann definieren als:

$\langle \xi_1, \xi_2 \rangle \in P_i \iff \langle \xi_1, \xi_2 \rangle \in R_i \wedge \sim \langle \xi_2, \xi_1 \rangle \in R_i$  ;

die Indifferenz als

$\xi_1 \approx_i \xi_2 \iff \langle \xi_1, \xi_2 \rangle \in R_i \wedge \langle \xi_2, \xi_1 \rangle \in R_i$  .

In der Standardsituation sind die individuellen Präferenzrelationen reflexiv, transitiv und mit  $\approx_i$  antisymmetrisch, mithin (konnexe) Ordnungsrelationen.<sup>1</sup>

### Rationalitätsbedingungen kollektiver Präferenzen

$f$  ordnet jeder Präferenzstruktur  $g$  von  $K$  eine kollektive Präferenzrelation  $R \subset \Xi \times \Xi$  zu.<sup>2</sup>  $f: G \ni g \mapsto R \subset \Xi \times \Xi$

Je nach den Rationalitätsbedingungen, die dabei von  $R$  erfüllt sind, ergeben sich unterschiedliche Aggregationsarten.

Sei  $M(\Xi_0, f(g)) = \{ \xi \mid \sim \exists \xi' \in \Xi_0: \langle \xi', \xi \rangle \in f(g) \wedge \sim \langle \xi, \xi' \rangle \in f(g) \}$ .

Sei  $C(\Xi_0, f(g)) = \{ \xi \mid \forall \xi' \in \Xi_0: \langle \xi, \xi' \rangle \in f(g) \}$ . (Auswahlmenge)

Es gilt:  $C(\Xi_0, f(g)) \subset M(\Xi_0, f(g))$ .

Wenn  $f(g)$  reflexiv und transitiv, so gilt darüberhinaus:

$C(\Xi_0, f(g)) \neq \emptyset \Rightarrow C(\Xi_0, f(g)) = M(\Xi_0, f(g))$ .<sup>3</sup>

$c_f$  sei die der Aggregationsregel  $f$  entsprechende Auswahlfunktion:

$c_f : \text{Pot}(\Xi) \times G \ni \langle \Xi_0, g \rangle \rightarrow C(\Xi_0, f(g))$ .

<sup>1</sup> Leider hat sich in der Literatur kein gemeinsamer Sprachgebrauch für diese Begriffe durchgesetzt. Arrow und Sen nennen reflexive, transitive und konnexe binäre Relationen Ordnungen, während es in der Mathematik üblich ist, reflexive, transitive und antisymmetrische binäre Relationen als Ordnungen und bei Vollständigkeit als totale oder konnexe Ordnungen zu bezeichnen.

<sup>2</sup> statt  $\langle \xi_1, \xi_2 \rangle \in f(g)$  auch  $\langle \xi_1, \xi_2 \rangle \in R$  und  $\xi_1 \succsim \xi_2$  .

$\xi_1 \approx \xi_2 \iff \xi_1 \succ \xi_2 \wedge \xi_2 \succ \xi_1$

$\xi_1 \succ \xi_2 \iff \xi_1 \succ \xi_2 \wedge \sim \xi_2 \succ \xi_1$

<sup>3</sup> Beweis Sen (1970a) 1\*3

Rationalitätsbedingung I für kollektive Präferenzzuordnungen

$$\forall E_0 \in \text{Pot}(E) \wedge E_0 \neq \emptyset : c_f(E_0, g) \neq \emptyset$$

für beliebige Präferenzstrukturen  $g$ .

Aggregationsregeln, die diese Bedingung erfüllen, heißen soziale Entscheidungsfunktionen.<sup>1</sup>

(I) verlangt, daß die Aggregationsregel  $f$  in jeder beliebigen Alternativenmenge  $E_0 \subseteq E$  (bei beliebigen Präferenzstrukturen) ein bestes Element bestimmt. Der intuitive Gedanke ist, daß man von einer guten Aggregationsregel erwarten kann, daß sie einem in jeder Situation kollektiver Entscheidung zu einer Alternative rät. Wobei sich hier bemerkbar macht, daß probabilistische Fragen ausgeklammert sind, so daß es in diesem Rahmen keinen Sinn macht, zwischen kollektiven Zustands- und Entscheidungspräferenzen zu differenzieren.

Hinreichend für (I) ist, daß die Elemente  $R$  aus  $\text{Bi}(f)$  reflexiv, konnex und quasi-transitiv sind, wobei

$\succ$  quasi-transitiv  $\Leftrightarrow \xi_1, \xi_2, \xi_3 \in E : \xi_1 \succ \xi_2 \wedge \xi_2 \succ \xi_3 \Rightarrow \xi_1 \succ \xi_3$ .

Wenn  $R$  reflexiv und konnex ist, so gibt es sogar eine hinreichende und notwendige Bedingung für (I), nämlich daß  $R$  in  $E$  azyklisch ist (da  $\# \in \mathbb{N}$ ) - s.u..

Rationalitätsbedingung II für kollektive Präferenzzuordnungen

$$\forall E_1, E_2 \in \text{Pot}(E) :$$

$$\xi \in E_1 \wedge E_1 \subseteq E_2 \Rightarrow [\xi \in c_f(E_2, g) \Rightarrow \xi \in c_f(E_1, g)].$$

(II) fordert, daß eine Alternative, die von der Aggregationsregel in einer Zustandsmenge als die beste ausgewählt wird, auch weiterhin als die beste gilt, wenn die Zustandsmenge verkleinert wird (ohne daß dabei diese Alternative herausfällt).

Da sich zeigen läßt, daß (II) immer erfüllt ist, wenn die Funktionswerte der Aggregationsregel  $f$  (kollektive) Prä-

---

<sup>1</sup> Vgl. Rothkirch (1981), S.3.

ferenzrelationen sind, ist diese Rationalitätsbedingung keine Verschärfung von (I).

Wenn wir jedoch zulassen, daß sich eine Aggregationsregel  $f$  darauf beschränken kann, Präferenzstrukturen und Alternativenmengen anstatt kollektiver Präferenzen, (direkt)beste Alternativen bzw. eine Auswahlmenge  $C(E_0, g)$  zuzuordnen,

$$- f : \text{Pot}(E) \times G \ni \langle E_0, g \rangle \mapsto C(E_0, g) -$$

so wird (II) zu einer eigenständigen Rationalitätsbedingung für  $f$ .

### Rationalitätsbedingung III für kollektive Präferenzzuordnungen

Die Aggregationsregel  $f$  faßt die individuellen (konnexen) Präferenzordnungen (über  $E$ ) zu kollektiven Präferenzrelationen zusammen, die konnex (in  $E$ ) sind:

$$\forall g \in G : f(g) \subset E \times E \wedge \forall \xi_1, \xi_2 \in E : \langle \xi_1, \xi_2 \rangle \in f(g) \vee \langle \xi_2, \xi_1 \rangle \in f(g).^1$$

Bei einer wohlfahrtsökonomischen Interpretation kollektiver Präferenzen ist (III) eine naheliegende Bedingung.<sup>2</sup> Sind die kollektiven Präferenzen jedoch als moralische Präferenzen zu interpretieren, wäre (III) eine sehr starke Forderung. Bis auf wenige Ausnahmen haben sozialetische Theorien nicht den Anspruch, Kriterien für die Entscheidung zwischen beliebigen Alternativen zu entwickeln.

### Rationalitätsbedingung IV für kollektive Präferenzzuordnungen

Die Aggregationsregel  $f$  faßt die individuellen Präferenzordnungen zu kollektiven Präferenzordnungen zusammen, die reflexiv, transitiv und konnex sind.

---

<sup>1</sup> Aggregationsregeln dieser Art werden in der social-choice-Literatur als 'decisive collective choice rules' bezeichnet; vgl. Sen (1970a)

<sup>2</sup> Vgl. Bergson (1938, auch 1966)

$$\begin{aligned} \forall g \in G : [ & f(g) \subset E \times E \wedge \\ & [\forall \xi \in E : \langle \xi, \xi \rangle \in f(g)] \wedge \\ & [\forall \xi_1, \xi_2, \xi_3 \in E : \langle \xi_1, \xi_2 \rangle \in f(g) \wedge \langle \xi_2, \xi_3 \rangle \in f(g) \Rightarrow \langle \xi_1, \xi_3 \rangle \in f(g)] \wedge \\ & [\forall \xi_1, \xi_2 \in E : \langle \xi_1, \xi_2 \rangle \in f(g) \vee \langle \xi_2, \xi_1 \rangle \in f(g) ]]. \end{aligned}$$

Solche Aggregationsregeln heißen soziale Wohlfahrtsfunktionen.<sup>1</sup>

Wenn  $E$  finit ist, ist (IV) eine hinreichende Bedingung für (I), und damit ist jede soziale Wohlfahrtsfunktion auch eine soziale Entscheidungsfunktion.

### Rationalitätsbedingung V für kollektive Präferenzzuordnungen

Die Aggregationsregel  $f$  faßt die individuellen Präferenzordnungen zu einer reflexiven und azyklischen Präferenzrelation zusammen.

$$\begin{aligned} \forall g \in G : [ & f(g) \subset E \times E \wedge \\ & [\forall \xi_1, \xi_2, \xi_3, \dots, \xi_\omega \in E : \\ & \langle \xi_1, \xi_2 \rangle \in f(g) \wedge \sim \langle \xi_2, \xi_1 \rangle \in f(g) \wedge \\ & \langle \xi_2, \xi_3 \rangle \in f(g) \wedge \sim \langle \xi_3, \xi_2 \rangle \in f(g) \wedge \\ & \dots \\ & \langle \xi_{\omega-1}, \xi_\omega \rangle \in f(g) \wedge \sim \langle \xi_\omega, \xi_{\omega-1} \rangle \in f(g) \\ & \Rightarrow \langle \xi_1, \xi_\omega \rangle \in f(g) ]]. \end{aligned}$$

Nur wenn die kollektive Präferenzrelation  $f(g)$  für beliebige Präferenzstrukturen  $g$  konnex ist, ist (V) hinreichende Bedingung für (I). (In diesem Fall ist sie sogar notwendig und hinreichend)

(V) kann als unverzichtbare Minimalbedingung kollektiver Rationalität angesehen werden.

In diesem Abschnitt würde zwar von Aggregationsregeln ( $f$ ) gesprochen, aber es ging uns genaugenommen nur um die Rationalität der Ergebnisse der Aggregation und nicht um die Rationalität bestimmter kollektiver Präferenzen

---

<sup>1</sup> Die Arrow-Bedingung 'unrestricted domain' (vgl. Arrow (1963) Kap.4) ist hier immer schon implizit vorausgesetzt.

in Relation zu gegebenen Präferenzstrukturen in K.

Um diese beiden Ebenen auseinanderzuhalten, ist dieser Abschnitt mit "Rationalität kollektiver Präferenzen" überschrieben, während wir im zweiten Teil der Arbeit in zwei Abschnitten unter dem Titel "Kollektive Rationalität I" und "Kollektive Rationalität II" auf die zweite o.g. Ebene zurückkommen.

B PRÄFERENZTHEORIE UND MORALPHILOSOPHIE

- § 6 Zur interpersonellen Vergleichbarkeit  
subjektiver Bewertungen: S. 54
- § 7 Die Interdependenzproblematik  
normativer Aggregationstheorien: S. 79
- § 8 Praktischer versus theoretischer  
Konsensus: S. 101
- § 9 Zur Differenzierung des Präferenzbegriffs: S. 109
- § 10 Moral und Nutzen: S. 136



## 6. ZUR INTERPERSONELLEN VERGLEICHBARKEIT SUBJEKTIVER BEWERTUNGEN

Der gerechte Ausgleich konfliktierender Interessen ist weiterhin das zentrale Thema der Sozialethik.

Insofern ist die interpersonelle Vergleichbarkeit persönlicher Interessen konstitutiv für die sozialetische Theorienbildung. Will man das Instrumentarium der rationalen Entscheidungs- und Spieltheorie und der Sozialwahltheorie für die Behandlung dieser Thematik heranziehen, so ist zweierlei erforderlich: Erstens die Klärung der Beziehung von subjektiven Präferenzen und persönlichen Interessen und zweitens eine Grundlage für den interpersonellen Vergleich persönlicher Interessen, der ihre gegenseitige wohlüberlegte Abwägung nach Grundsätzen sozialer Gerechtigkeit erst möglich macht. Beides hängt miteinander zusammen: Wir werden zuletzt zu dem Ergebnis kommen, daß es nur für die Wohlfahrtspräferenzen sinnvoll ist, volle interpersonelle kardinale Vergleichbarkeit anzuwenden.

Dennoch werden wir zunächst/<sup>allgemein</sup> die interpersonelle Vergleichbarkeit subjektiver Bewertungen diskutieren (§6), anschließend anhand der Möglichkeit zirkulärer Präferenzbestimmtheit aufzeigen, daß eine Revision des Präferenzbegriffs für die Zwecke normativer Sozialethik unumgänglich ist (§7), die These durch einige Überlegungen zur Unterscheidung von Interessen- und Urteilsaggregation untermauern (§8) und dann das Programm einer solchen Begriffsdifferenzierung - soweit es für die Behandlung des zweiten zentralen ethisch-materialen Problems der Sozialethik, nämlich das der kollektiven Rationalität, erforderlich erscheint - skizzieren (§9). Zum Ende des Abschnittes B des ersten Teils der Arbeit wird eine idealtypische Alternative zu unserem Ansatz diskutiert.

Dieser zweite Abschnitt des ersten Teiles der Arbeit beinhaltet also eine Art hermeneutischen Zirkels. Das Verständnis der Diskussion interpersoneller Vergleichbarkeit setzt das Verständnis für die Differenzierung des Präferenzbegriffs schon voraus - und umgekehrt. Und die Frage "Wozu soll das alles gut sein?" beantwortet sich leider erst im zweiten Teil der Arbeit.

Die Frage der interpersonellen Vergleichbarkeit individueller Präferenzen spielt eine große Rolle in der politischen Ökonomie, in der Wohlfahrts- und Sozialwahltheorie. Der Übergang von der klassischen zur paretianischen Tradition der Wohlfahrtstheorie etwa ist gekennzeichnet durch eine unterschiedliche Beurteilung dieser Vergleichbarkeitsfrage. Die Entwicklung der collective-choice-Theorie ist vor allem Konsequenz einer starken Informationsbeschränkung: Es gibt keine kardinale Vergleichbarkeit subjektiver Bewertungen.<sup>1</sup> Die Positionen scheiden sich an der Frage: Gibt es einen absoluten Nutzen?

Es gibt einen 'absoluten Nutzen', wenn die Nutzenfunktionen vollständig kardinal vergleichbar sind.

Messen die Individuen ihren Nutzen mit unterschiedlichen Maßeinheiten und unterschiedlichen Nullpunkten, so ist damit - entgegen dem, was man zuweilen in der ökonomischen Literatur liest - noch nicht Unvergleichbarkeit der Nutzenniveaus oder der Nutzendifferenzen die Folge: Die Temperatur läßt sich in Celsius und in Fahrenheit messen, ohne daß das die Temperaturvergleichbarkeit einschränkt, da uns die Transformationsfunktion  $n^{\circ}\text{C} = (\frac{9}{5}n+32)^{\circ}\text{F}$  bekannt ist. Die Anwendung interpersonell unterschiedlicher 'Nutzenmaßstäbe' ist also für

---

<sup>1</sup> Vgl. u.a. A. Bergson, *Essays in normative economics*, Cambridge 1966.

die interpersonelle Nutzenvergleichbarkeit nur dann problematisch, wenn wir keine empirische Grundlage haben, um eine Transformationsfunktion zu bestimmen.

Die v. Neumann/Morgensternsche Nutzentheorie<sup>1</sup> liefert aufgrund eines behavioristischen<sup>2</sup> Modells der konsistenten Wahl zwischen Lotterien individuelle kardinale Nutzenfunktionen.

Es handelt sich um eine behavioristische Methode der Nutzenbestimmung, da die Nutzenfunktionen ausschließlich aufgrund individuellen Wahlverhaltens - nicht aufgrund

---

<sup>1</sup> Eine einfache Darlegung der Bedingungen an eine Präferenzrelation, die eine bis auf lineare Transformationen festgelegte reellwertige Nutzenfunktion konstituieren, findet sich (mit Beweis) bei H. Chernoff/E.L. Moses, *Elementary decision theory*, New York 1959, S.350-352. Zu Nutzenbegriff und Meßbarkeit vgl. a. L.J. Savage, *The foundations of statistics*, New York/London 1954, Kap.5. Zum Zusammenhang von entscheidungstheoretischen Annahmen und Nutzenbegriff: M.Friedmann/L.J. Savage, *The expected utility hypothesis and the measurability of utility*, *Journal of economic theory* 60 (1952) 463-474. Sehr klare Darstellung der v. Neumann/Morgensternschen Nutzenaxiomatik und ihrer Implikationen in H.W. Gottinger, *Grundlagen der Entscheidungstheorie*, Stuttgart 1974, S.30-76. H.W. Schneeweiß bietet einen vergleichenden Überblick zu den unterschiedlichen Axiomensystemen, die die Existenz eines Bernoulli-Nutzens gewährleisten, *Nutzenaxiomatik und Theorie des Messens*, *Statistische Hefte* 4 (1963). Vgl. a. Marschak, *Rational behavior, uncertain prospects, and measurable utility*, *Econometrica* 18 (1950) 111-141, daher der Name "Marschak-Axiome"; einführend ders., *Nutzenmessung und Wahrscheinlichkeit*, in Shubik, *Spieltheorie und Sozialwissenschaft*, Frankfurt/M. 1965, 103-118; P.A. Samuelson, *Probability, utility and the independence axiom*, *Econometrica* 20 (1952) 670-678. In einem allgemeineren Kontext: G. Gäfgen, *Theorie der wirtschaftlichen Entscheidung*, Tübingen 1974; eine sehr elegante Formulierung der relativen Rationalitätspostulate, die eine Bernoulli-Nutzenfunktion konstituieren, bei J.N. Herstein/J. Milnor, *An axiomatic approach to measurable utility*, *Econometrica* 21 (1953). "Bernoulli-Nutzen" nach Daniel Bernoulli, *Specimen theoriae novae de mensura sortis*, (1738), statt "Bernoulli-Nutzen" auch "erwartungstreuer Nutzen", "Risiko-Nutzen" oder "linearer Nutzen". Schon lange vor dem bahnbrechenden Werk von J.v. Neumann/ O. Morgenstern (1947) hat Ramsey 1926 seine Nutzenaxiomatik entwickelt, vgl. F.P. Ramsey, *The foundations of mathematics and other logical issues*, London 1954, Kap.VII.

<sup>2</sup> Vgl. P. Suppes, *Behavioristic foundations of utility*, *Economica* 26 (1961) 186-202 und G. Debreu, *Stochastic choice and cardinal utility*, *Econometrica* 26 (1958) 440-444.

von z.B. Befragungen - festgelegt werden.<sup>1</sup> In diesem Zusammenhang wird in der englischsprachigen Literatur von "revealed preference" gesprochen, um den Unterschied zu Präferenzen aufgrund von Selbstbeurteilungen deutlich zu machen.<sup>2</sup>

Das Problem der konkreten Bestimmung dieser kardinalen Nutzenfunktionen soll uns hier nicht weiter beschäftigen.<sup>3</sup>

Da diese individuellen kardinalen Nutzenfunktionen aufgrund der v. Neumann/Morgenstern-Bedingungen<sup>4</sup> nur bis auf lineare Transformation bestimmt sind, erlaubt diese Informationsbasis noch keine interpersonellen Nutzen- niveau- oder Nutzendifferenzvergleiche.

---

<sup>1</sup> Einen guten Überblick zur Nutzentheorie geben R.D. Luce und P. Suppes im *Handbook of mathematical psychology*, Bd.3, S.249-411.

<sup>2</sup> Vgl. z.B. W.M. Gorman, *Preference, revealed preference and indifference*, in: Chipman/Hurwicz/Richter/Sonnenschein (1971) 81-113; W. Thomson, *Maximin strategies and elicitation of preference*, in: Laffont (1979) 245-268.

<sup>3</sup> Debreu gibt eine axiomatische Bestimmungsmethode an, die, da sie einen behavioristischen frequentistischen Nutzenbegriff expliziert, weit schwächere Konsistenzbedingungen für das individuelle Wahlverhalten zur Festlegung einer Nutzenfunktion erfordert. Vgl. G. Debreu, *Stochastic choice and cardinal utility*, *Econometrica* 26 (1958); vgl. der, *Topological methods in cardinal utility theory*, in: Arrow/Karlin/Suppes (1966) 16-26; ders., *Werttheorie: Eine axiomatische Analyse des ökonomischen Gleichgewichts*, Berlin/Heidelberg/New York 1976. Vgl. dazu auch: H.S. Houthakker, *Revealed preference and the utility fraction*, *Economica* 17 (1950) 159-174; D. Ellsberg, *Classic and current notions of measurable utility*, *Economic journal* 64 (1954) 528-556; M.P. Hurst/S. Siegel, *Prediction of decisions from a metric scale of utility*, *Journal of experimental psychology* 52 (1956) 128-134; M.H. De Groot, *Some comments on the experimental measurement of utility*, *Behavioral science* 8 (1963) 146-149; A.C. Michalos, *Estimated utility and corroboration*, *British journal of philosophical science* 16 (1966-67) 327-331; A. Camacho, *Maximizing expected utility and the rule of long run success*, in: Allais/Hagen (1979); ders., *On cardinal utility*, *Theory and decision* 10 (1979) 131-145; ders., *Approaches to cardinal utility*, *Theory and decision* 12 (1980) 359-379. Vgl.a. die empirischen Untersuchungen von P. Nozette/F. Mosteller, *An experimental measure of utility*, *Journal of political economy* 59 (1951) 371-404; D. Davidson/P. Suppes/S. Siegel, *Decision making: An experimental approach*, Stanford 1957; G.M. Becker/M.H. De Groot/J. Marschak, *Measuring utility by a single-response sequential method*, *Behavioral science* 9 (1964) 226-232.

<sup>4</sup> Oder der äquivalenten Marschak-Bedingungen, vgl. J. Marschak, *Rational behavior, uncertain prospects and measurable utility*. *Econometrica* 18 (1950) 111-141.

Die Kardinalität der subjektiven Bewertungsfunktionen sagt etwas über die intrapersonelle Vergleichbarkeit von Präferenzen aus, während eine ordinale Skala ausschließlich darüber Auskunft gibt, welcher Zustand welchem anderen vorgezogen wird. Die Größe der Differenz zweier Nutzenbewertungen hat in einer ordinalen Skala keine Bedeutung, während sie in einer kardinalen Skala sehr wohl eine Bedeutung hat, die bisweilen als "Intensität der Präferenz" interpretiert wird. Dieser Sachverhalt läßt sich auch so formulieren:

Zwei ordinale Bewertungsfunktionen, die durch eine positiv-monotone (intrapersonelle) Transformation ineinander übergehen, haben den gleichen Aussagegehalt. Zwei kardinale Bewertungsfunktionen, die durch eine positiv-monotone (intrapersonelle) Transformation ineinander übergehen, haben dagegen i.a. einen unterschiedlichen Aussagegehalt. (Nur ein Sonderfall der positiv-monotonen Transformationen ist davon ausgenommen: nämlich die Klasse der linearen positiv-monotonen Transformationen.)

Es gibt eine ganze Reihe von Variationen intra- und interpersoneller Bewertungsvergleichbarkeit. Und ich sehe im wesentlichen nur zwei Methoden, um diese Variationen formal explizit zu machen.

Die erste Methode besteht darin, einen Grundbegriff einzuführen (den man als "tatsächliche individuelle Bewertung" bezeichnen könnte) und dann die Informationsbasis durch eine Implikation der Bewertungskonstanz zu kennzeichnen.

Die beiden o.g. Beispiele intrapersoneller Vergleichbarkeit nehmen damit folgende Gestalt an:

(1) Ordinalität der individuellen Bewertungsfunktionen

$$\forall \xi \in E : r_i(\xi) = r'_i(\xi) \Rightarrow \exists T \in T_m^+ : \forall \xi \in E : T \circ U_i(\xi) = U'_i(\xi) \quad 1$$

(2) Kardinalität der individuellen Bewertungsfunktionen

$$\forall \xi \in E : r_i(\xi) = r'_i(\xi) \Rightarrow \exists T \in T_1^+ : \forall \xi \in E : T \circ U_i(\xi) = U'_i(\xi) \quad 2$$

---

<sup>1</sup>  $T_m^+$  := Menge der positiv-monotonen Transformationen

<sup>2</sup>  $T_1^+$  := Menge der positiv-linearen Transformationen

Die Verschärfung zur Äquivalenz wäre natürlich inadäquat, denn i.a. erfaßt z.B. eine ordinale Bewertungsfunktion nicht alle Merkmale subjektiver Bewertung. Wenn jemand der Ansicht ist, daß eine kardinale Nutzenfunktion alle Informationen über die individuelle Bewertung der Zustände enthält (er könnte das mit einem methodologischen Behaviorismus begründen), so behauptet er die Umkehrbarkeit von (2).

Die zweite Methode besteht darin, die Vergleichbarkeitsannahmen in der Form von Aggregationspostulaten explizit zu machen.<sup>1</sup> Dies hat den Vorteil, daß man auf den Grundbegriff der "tatsächlichen individuellen Bewertung" verzichten kann.

Die intrapersonelle Vergleichbarkeit im ordinalen Fall nimmt dann die Form des folgenden Postulates an:

$$(1') \forall i \in K \exists \tau \in T_m^+ \forall \xi \in \Xi : \tau \circ u_i(\xi) = u_i'(\xi) \quad f(U) = f(U'),$$

und im kardinalen Fall:

$$(2') \forall i \in K \exists \tau \in T_1^+ \forall \xi \in \Xi : \tau \circ u_i(\xi) = u_i'(\xi) \quad f(U) = f(U')^2$$

Die kollektive Präferenzrelation darf sich für die Informationsbasis ordinaler individueller Bewertungsfunktionen nicht ändern, wenn diese positiv-monotonen Transformationen unterworfen sind. Im kardinalen Fall wird das für lineare positiv-monotone Transformationen gefordert.

---

<sup>1</sup> Das ist der Weg, den Hammond (1976a und b), d'Asprement/Gevers (1977), Sen (1970a und b, 1976, 1977a), Deschamps/Gevers (1978), Kern (1980a), Roberts (1980) beschreiten.

Sozialethische Aggregationsregeln  $f$ , die jedem  $n$ -Tupel individueller Wohlfahrtsfunktionen genau eine kollektive Präferenzordnung über die Menge der Alternativen zuordnen, werden in der collective-choice-Literatur "social welfare functionals" genannt, vgl. Sen (1970a) 8\*.2. und Sen (1977a) S. 1540.

<sup>2</sup> Kardinale Bewertungsfunktionen erlauben intrapersonelle Vergleiche auf folgende Weise:

Die Präferenz einer Person für  $\xi_1$  gegenüber  $\xi_2$  ist größer als die Präferenz für  $\xi_3$  gegenüber  $\xi_4$  genau dann, wenn sie eine 50:50-Lotterie zwischen  $\xi_1$  und  $\xi_4$  einer 50:50-Lotterie zwischen  $\xi_2$  und  $\xi_3$  vorzieht. Diese Möglichkeit intrapersoneller

(1') ist intuitiv deshalb überzeugend, weil jegliche positiv-monotone Transformation einer ordinalen Bewertungsfunktion sozusagen bedeutungslos ist. Das wird aber direkter durch (1) ausgedrückt. Wir kennen bloß eine notwendige Bedingung für unveränderte individuelle Bewertungen. Ist diese notwendige Bedingung nicht erfüllt, wissen wir, daß sich die entsprechenden individuellen Bewertungen geändert haben und das rechtfertigt u. U. auch eine Veränderung der kollektiven Präferenzen. Ist diese (notwendige) Bedingung jedoch erfüllt, wissen wir nicht, ob sich die individuellen Bewertungen geändert haben und können daher auch eine Veränderung der kollektiven Präferenzen nicht rechtfertigen. Es wäre jedoch in dieser Interpretation der formalen Charakterisierung unterschiedlicher Informationsvoraussetzungen falsch, zu sagen: "Wenn die jeweilige notwendige Bedingung erfüllt ist, sind auch die tatsächlichen individuellen Bewertungen unverändert."

Die erste Methode der formalen Charakterisierung der sozialetischen Informationsbasis ist in einem gewissen Sinne die explizite Fassung der intuitiven Begründung für die Aggregationspostulate der zweiten Methode.

Im folgenden fassen wir einige wichtige Variationen der sozialetischen Informationsbasis mithilfe der zweiten Methode als Aggregationspostulate an jede SWFL  $f$  zusammen, wobei  $f(\langle u_i \rangle) = f(U)$ ,  $\langle u_i \rangle_{i=1,2,\dots,n}$  mit  $n = \#K$ .

---

Bewertungsvergleichbarkeit beruht auf der Hypothese, daß es für eine rationale Person eine Bewertungsfunktion über die Alternativen gibt, die es erlaubt, den Wert einer Lotterie als die Summe der mit ihren Wahrscheinlichkeiten gewichteten Bewertungen der wechselseitig unvereinbaren Konsequenzen zu bestimmen.

Vgl. dazu detailliert: P. Suppes/M. Winet, *An axiomatization of utility based on the notion of utility differences*, Management science 1 (1955) 259-270.

<sup>1</sup> Bisweilen wird von "individuellen Wohlfahrtsfunktionen" gesprochen in der Bedeutung von interpersonell vollständig vergleichbaren kardinalen Nutzenfunktionen, die normiert worden sind: interpersonell gleiche Differenzen der numerischen Werte individueller Wohlfahrtsfunktionen repräsentieren auch gleiche faktische Wohlfahrtsdifferenzen. Vgl. z.B. R.C. Jeffrey, *On interpersonal utility theory*, Journal of philosophy 68 (1971), S. 649ff. Diesem Sprachgebrauch schließen wir uns nicht an. In §9 wird der Terminus näher erläutert.

1. Die individuellen Bewertungsfunktionen sind vollständig kardinal interpersonell vergleichbar bezüglich einer Aggregationsregel  $f^1$  genau dann, wenn gilt:

$$\exists \tau \in T_1^+ : \forall i \in K : \forall \xi \in \Xi : \tau \circ u_i(\xi) = u_i'(\xi) \Leftrightarrow f(U) = f(U')$$

2. Die individuellen Bewertungsfunktionen sind vollständig kardinal interpersonell unvergleichbar bezüglich einer Aggregationsregel  $f$  genau dann, wenn gilt:

$$\forall i \in K : \exists \tau_1, \tau_2, \dots, \tau_n \in T_1^+ : \forall \xi \in \Xi : \tau \circ u_i(\xi) = u_i'(\xi) \Leftrightarrow f(U) = f(U')$$

3. Die individuellen Bewertungsfunktionen sind ausschließlich kardinal interpersonell niveauevergleichbar bezüglich einer Aggregationsregel  $f$  genau dann, wenn gilt:

$$\exists a \in \mathbb{R}^+ : \forall i \in K : \exists b_1, b_2, \dots, b_n \in \mathbb{R}^+ : \forall \xi \in \Xi : a + b_i \cdot u_i(\xi) = u_i'(\xi) \Leftrightarrow f(U) = f(U')$$

4. Die individuellen Bewertungsfunktionen sind ausschließlich kardinal interpersonell einheitenvergleichbar bezüglich einer Aggregationsregel  $f$  genau dann, wenn gilt:

$$\exists b \in \mathbb{R}^+ : \forall i \in K : \exists a_1, a_2, \dots, a_n \in \mathbb{R}^+ : \forall \xi \in \Xi : a_i + b \cdot u_i(\xi) = u_i'(\xi) \Leftrightarrow f(U) = f(U')^2$$

5. Die individuellen Bewertungsfunktionen sind vollständig ordinal interpersonell vergleichbar bezüglich einer Aggregationsregel  $f$  genau dann, wenn gilt:

$$\exists \tau \in T_m^+ : \forall i \in K : \forall \xi \in \Xi : \tau \circ u_i(\xi) = u_i'(\xi) \Leftrightarrow f(U) = f(U')$$

6. Die individuellen Bewertungsfunktionen sind vollständig ordinal interpersonell unvergleichbar bezüglich einer Aggregationsregel  $f$  genau dann, wenn gilt:

$$\forall i \in K : \exists \tau_1, \tau_2, \dots, \tau_n \in T_1^+ : \forall \xi \in \Xi : \tau \circ u_i(\xi) = u_i'(\xi) \Leftrightarrow f(U) = f(U')^3$$

Jeweils  $\#K = n$ .

Diese Charakterisierungen unterschiedlicher Vergleichbarkeit sind auf die jeweilige Aggregationsregel relativiert.

---

<sup>1</sup> Sen führt die Vergleichbarkeitsbegriffe bzgl. einer speziellen Aggregationsregel nämlich der Nutzensummenregel ein, vgl. Sen (1970a) 7\*.

Die dort diskutierten Begriffe partieller Vergleichbarkeit lassen sich mutatis mutandis in unsere Darstellung übertragen.

<sup>2 3</sup> s. nächste Seite



Interpersonell vollständig kardinal vergleichbare Bewertungsfunktionen teilen sehr viel mehr über die individuellen Bewertungen mit als interpersonell unvergleichbare. Nur mit interpersonell vergleichbaren Bewertungsfunktionen können bspw. Intensitäten individueller Präferenzen dargestellt werden. Dennoch hängt es noch weitgehend vom Charakter der Präferenzen ab, wie Bewertungsniveaus und -differenzen zu interpretieren sind. Wenn man etwa  $u_i(\xi) > u_j(\xi')$  als "i geht es bei  $\xi$  besser als j bei  $\xi'$ " interpretiert, so ist das nur unter der Prämisse sinnvoll, daß sich die individuellen Präferenzen nach den persönlichen Interessen richten, was im Rahmen des 'revealed-preference'-Konzeptes keineswegs selbstverständlich ist. Unabhängig vom spezifischen Charakter der zugrundegelegten Präferenzen lassen sich bei interpersoneller kardinaler Einheitenvergleichbarkeit Aussagen wie z.B. "Person i präferiert die Alternative  $\xi$  intensiver gegenüber  $\xi'$  als j" durch  $u_i(\xi) - u_i(\xi') > u_j(\xi) - u_j(\xi')$  ausdrücken<sup>1</sup>, während dies mit der von Neumann-Morgensternschen Kardinalisierung von Nutzenfunktionen<sup>2</sup> allein nicht möglich ist.

---

<sup>2</sup> Die erste Methode hätte die volle kardinale Vergleichbarkeit durch " $\forall i, j \in K : r_i = r_j \Rightarrow \forall \xi \in E : u_i(\xi) = u_j(\xi)$ " ausgedrückt.

<sup>3</sup> Natürlich lassen sich auch alle Übergänge zwischen reiner Niveau- und reiner Differenzvergleichbarkeit denken: Für  $\#\{b_i\} < n$  in (3) wäre z.B. zusätzlich eine partielle Differenzvergleichbarkeit (eine Differenzvergleichbarkeit in einer Teilmenge von K) gegeben; ebenso wäre für  $\#\{a_i\} < n$  in (4) eine partielle Niveauvergleichbarkeit gegeben.

<sup>1</sup> Selbst Aussagen wie z.B. "i präferiert  $\xi_1$  gegenüber  $\xi_2$  intensiver als j  $\xi_3$  gegenüber  $\xi_4$  präferiert" lassen sich formulieren:  $u_i(\xi_1) - u_i(\xi_2) > u_j(\xi_3) - u_j(\xi_4)$ .

<sup>2</sup> Interpersonell vollständig vergleichbare kardinale Nutzenfunktionen werden auch "absolute Nutzenfunktionen" genannt.

Die sozialetische Informationsbasis spielt in zweierlei Hinsicht eine große Rolle.

Einmal machen unterschiedliche normative Theorien der Sozialetik verschiedene Annahmen über die subjektiven Bewertungen, so daß die Informationsbasis mit darüber entscheidet, welche Theorie adäquat ist. Wenn es sich z.B. als unmöglich herausstellen würde, die Zunahme bzw. Abnahme des persönlichen Wohlergehens durch die Wahl einer bestimmten Alternative interpersonell vergleichbar zu bestimmen, so würde das den utilitaristischen Theorien ihre Informationsbasis entziehen. Dann wäre es nämlich i.a. nicht entscheidbar, welche von zwei Alternativen utilitaristisch besser ist. Ein Sonderfall ist davon allerdings ausgenommen: Wenn es allen Personen bei  $\xi$  besser als bei  $\xi'$  geht, ist  $\xi$  utilitaristisch besser als  $\xi'$ . Bei interpersonell nicht vergleichbaren ordinalen subjektiven Bewertungsfunktionen als Informationsbasis ergibt sich so als einzige utilitaristisch gerechtfertigte Entscheidungsregel die Pareto-Regel: Entscheide dich bei zwei Alternativen  $\xi$  und  $\xi'$  für  $\xi$ , wenn  $\xi$  von allen individuellen Bewertungsfunktionen höher bewertet wird.

Rothkirch subsumiert Aggregationstheorien individueller Präferenzrelationen allgemein unter die Sozialetik des 'ordinalen Utilitarismus', so daß in dieser Interpretation Arrows Unmöglichkeitstheorem das Scheitern des ordinalen Utilitarismus aufzeigt.<sup>1</sup> Entscheidend ist jedoch nicht die intrapersonelle, sondern die interpersonelle Vergleichbarkeit: Auch bei unvergleichbaren kardinalen Bewertungsfunktionen kann nur die Pareto-Regel utilitaristisch begründet werden.<sup>2</sup>

---

<sup>1</sup> Vgl. C.v. Rothkirch, *Rationales Handeln im öffentlichen Interesse: Theorien kollektiver Entscheidungen*, (Habil.-Schrift), Frankfurt/M. 1981, Kap. 5 u. 6.

<sup>2</sup> Darüberhinaus ist es fraglich, ob der Ansatz der Aggregation individueller Präferenzrelationen sinnvoll in toto dem Utili-

Völlige Unvergleichbarkeit der subjektiven Bewertungsfunktionen schränkt die Möglichkeiten normativer Theorienbildung sehr stark ein, so stark, daß schon die äußerst schwachen Arrow-Bedingungen jede Aggregationsregel logisch ausschließen.<sup>1</sup>

Zum zweiten kann es sinnvoll sein, sozialetische Kriterien zu axiomatisieren, d.h. eine Reihe von Metakriterien anzugeben, die zusammen notwendig und hinreichend für das betreffende sozialetische Kriterium sind. Eine solche Axiomatisierung hat zwei Funktionen. Sie kann der Begründung dienen, wenn diese Metakriterien für sich eine stärkere Evidenz beanspruchen können als das Kriterium selbst. Eine Axiomatisierung ist aber auch schon dann sinnvoll, wenn damit die Unterschiede konkurrierender sozialetischer Kriterien schärfer herausgearbeitet werden, so daß sich die Diskussion, welches von beiden vorzuziehen sei, neben konkreten Einzelfallbeurteilungen auch auf die theoretische Differenz stützen kann. Sehr gründlich ist das bisher für die Konkurrenz von (sozialwahltheoretisch interpretiertem) Utilitarismus und Rawlsschem Differenzprinzip geleistet worden.<sup>2</sup>

---

tarismus zugeordnet werden kann. Utilitarismus und Theorien normativ gebotener CCR (s. FN S.59) sind teleologische Theorien, die zumeist von einem methodischen Individualismus begleitet sind (darauf gehen wir in §11 noch ein), aber es ist nicht sinnvoll, alle Theorien dieser Art als utilitaristische zu bezeichnen.

<sup>1</sup> Vgl. J.K. Arrow, *Social choice and individual values*, New York/London/Sidney 1963, Kap.5

<sup>2</sup> Vgl. A.K. Sen, *Informational bases of alternative welfare approaches*, *Journal of public economics* 3 (1974) 387-403; ders., *Rawls versus Bentham, An axiomatic examination of the pure distribution problem*, in: Daniels (o.J.) 283-292; ders., *Welfare inequalities and Rawlsian axiomatics*, *Theory and decision* 7 (1976) 243-262; C. d'Asprement/L. Gevers, *Leximin and utilitarian rules: A joint characterization*, *Journal of economic theory* 19 (1978) 143-163; Sen (1977a), S.1545-1550; W. Gaertner, *Rawlsianism, utilitarianism and profiles of extended orderings*, Bielefeld 1980. S.a. Strasnick (1976, 1979) und Wittmann (1979)

Da wir im zweiten Teil der Arbeit volle interpersonelle kardinale Vergleichbarkeit der individuellen Wohlfahrtsfunktionen benötigen, müssen wir uns hier mit der Frage auseinandersetzen, ob eine so reichliche Informationsbasis gerechtfertigt werden kann. Ist es sinnvoll, einen absoluten Nutzen oder (in Übereinstimmung mit der später vorgeschlagenen Begriffsdifferenzierung müßten wir besser sagen:) ein absolutes Wohlfahrtsmaß vorauszusetzen?

Dabei spaltet sich das Problem in zwei zu unterscheidende Fragestellungen auf: Existiert ein absoluter Nutzen? und Wie ist der absolute Nutzen meßbar?

Die Annahme der Existenz eines absoluten Nutzens erscheint weniger problematisch, wenn wir davon ausgehen könnten, daß sich die subjektiven Bewertungen rationaler Personen durch jeweils beschränkte Nutzenintervalle repräsentieren lassen.<sup>1</sup>

Die Annahme ausschließlich beschränkter Nutzenintervalle hat intuitiv sehr überzeugende Argumente für sich. Denn wäre der Nutzenbereich einer Person  $i$  nach oben unbeschränkt, so hätte das folgende paradoxe Konsequenz:

Sei  $\epsilon$  ein irgendwie geartetes für  $i$  besonders positives Ereignis, etwa, daß seine seit langem verehrte Freundin nun doch seinen Heiratsantrag erhört, und  $\epsilon'$  ein für  $i$  äußerst unangenehmes Ereignis, etwa aufgrund einer Verwechslung unschuldig und ohne Aussicht auf Aufklärung zu lebenslänglicher Haft verurteilt zu werden, so muß es, wäre  $i$ 's Nutzenbereich nach oben unbeschränkt, dennoch ein Ereignis  $\epsilon_0$  geben, so daß  $i$  aufgrund rationaler Entscheidung die 50:50-Lotterie ( $0,5 \cdot \epsilon_0$  &  $0,5 \cdot \epsilon'$ )

---

<sup>1</sup> Isbell vertritt diese Ansicht, die er u.a. mit dem Petersburger Paradoxon begründet, welches nur durch die Annahme ausschließlich beschränkter Nutzenfunktionen zu umgehen sei, vgl. J.R. Isbell, *Absolute games*, in: A.W. Tucker/R.D. Luce, *Contributions to the theory of games, IV*, Princeton/N.J. 1959. Jeffrey hält dem jedoch eine - sozusagen pragmatische - Alternative zur Auflösung des Paradoxons entgegen: Niemand kann glaubwürdig ein St. Petersburg Spiel anbieten, da niemand unbegrenzt hohe Auszahlungen garantie-

dem lang ersehnten sicheren Ereignis  $\epsilon$  vorziehen würde.<sup>1</sup> Falls nur eine Ereigniskombination  $\epsilon, \epsilon'$  gefunden werden kann, für die eine solche Präferenz undenkbar ist, so müssen wir von einer nach oben beschränkten Nutzenfunktion  $U_i$  ausgehen. Analoge Argumentation liefert uns die Beschränkung von  $Bi(U_i)$  nach unten. Es liegt dann nahe, die individuellen beschränkten Nutzenfunktionen auf ein Intervall, z.B.  $[0,1]$  zu normieren.

Sind die individuellen Nutzenfunktionen kardinal, hätten wir mit diesem Verfahren die Existenz eines absoluten Nutzenmaßes garantiert. Die Operationalisierbarkeit dieses Nutzenmaßes ist natürlich damit noch in keiner Weise geklärt.

Aber zunächst stellt sich die Frage, ob die so gewonnene vollständige interpersonelle Nutzenvergleichbarkeit adäquat ist, oder ob sie nicht mit anderen Vergleichsverfahren in Konflikt gerät. Es ist ja durchaus denkbar, daß diese Methode des Vergleichs subjektiver Bewertungen mit intuitiv überzeugenden und u.U. intersubjektiv durchgängigen interpersonellen Vergleichen individueller Nutzenwerte nicht nur das eine oder andere Mal, sondern regelmäßig kollidiert. Die Annahme der Beschränktheit der individuellen Nutzenfunktionen impliziert also nicht unbedingt eine Lösung des Vergleichbarkeitsproblems.<sup>2</sup>

---

ren kann. Dennoch bestreitet Jeffrey die Plausibilität der Beschränktheitsannahme für 'Wünschbarkeitsfunktionen' nicht. Vgl. R. Jeffrey, *Logik der Entscheidungen*, Wien/München 1967, S.188-190.

<sup>1</sup> Denn:

(1) Beschränktheit:  $\sim \exists r \in R : \forall \epsilon \in Z_i(U_i) : U \leq r$

(2) Angenommen  $\exists \epsilon, \epsilon' \in P(\bar{E}) : \sim \exists \epsilon_0 \in P(\bar{E}) : 0,5U_i(\epsilon_0) + 0,5U_i(\epsilon') > U_i(\epsilon)$

$\Rightarrow \exists \epsilon, \epsilon' \in P(\bar{E}) : \sim \exists \epsilon_0 \in P(\bar{E}) : U_i(\epsilon_0) > 2U_i(\epsilon) - U_i(\epsilon')$

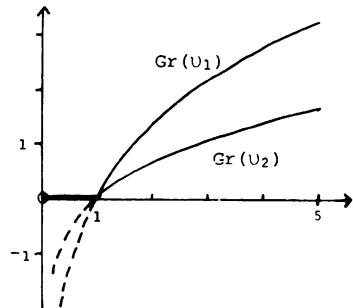
$\Rightarrow \forall \epsilon \in Z_i(U_i) : U \leq 2U_i(\epsilon) \Rightarrow (1).$

<sup>2</sup> Owen simplifiziert daher das Problem absoluten Nutzens zu weitgehend, wenn er meint es sei auf die Frage reduzierbar, ob der Nutzenbereich einer Person beschränkt ist oder nicht. Vgl. G. Owen, *Spieltheorie*, Berlin/Heidelberg/New York 1971, S.139.

So steht diese Methode z.B. in einem gewissen Spannungsverhältnis zu der intuitiven Überzeugung, daß Menschen in unterschiedlicher Weise zur Zufriedenheit befähigt sind.<sup>1</sup> Wer dennoch an dieser Art des interpersonellen Nutzenvergleichs festhält, könnte dies jedoch im Rahmen einer normativen Theorie tun. So führt der Utilitarismus in einem gewissen Sinne zu einer 'Übervorteilung' der 'weniger Glücksfähigen', wenn wir eine klassische Nutzeninterpretation zugrunde legen. Da sich jeder so verhalten soll, daß sich der Gesamtnutzen maximiert, gelangen in einer utilitaristisch geprägten Gesellschaft Güter tendenziell in den Besitz derjenigen, bei welchen sich eine stärkere Zunahme der individuellen Wohlfahrtsniveaus hervorrufen. Für einen auf individuellen, in dieser Weise normierten Bernoulli-Nutzenfunktionen basierenden Utilitarismus würde das nicht gelten.

Das sei an einem einfachen Beispiel erläutert:

Man stelle sich eine Welt vor, die aus zwei Personen A und B und einem Gut, materiell oder immateriell, aber kontinuierlich zerlegbar, besteht. Die utilitaristische Handlungsanleitung fordert eine Maximierung des Gesamtnutzens:  $U(x_1, x_2) = U_1(x_1) + U_2(x_2)$  soll maximal sein. Ist von diesem Gut nur eine bestimmte Menge  $x_0$  vorhanden, so ist eine Verteilung  $\langle x, x_0 - x \rangle$ ,  $x \in [0, x_0]$  so zu wählen, daß  $U_1(x) + U_2(x_0 - x)$  maximal wird. In unserer Welt ist A 'glücksfähiger' als B: bei



$$U_1(x) = \begin{cases} 0, & \text{für } x < 1 \\ 2 \ln x, & \text{für } x \geq 1 \end{cases}$$

$$U_2(x) = \begin{cases} 0, & \text{für } x < 1 \\ \ln x, & \text{für } x \geq 1 \end{cases}$$

$$U_1'(x) = \frac{dU_1(x)}{dx} = \frac{2}{x}$$

$$U_2'(x) = \frac{dU_2(x)}{dx} = \frac{1}{x};$$

<sup>1</sup> Auf das Verhältnis Nutzenfunktion - individuelle Wohlfahrt gehen wir in §11 noch detaillierter ein.

derselben Ausstattung ist das Nutzenniveau von A mindestens so groß wie das von B, und dieselbe zusätzliche Gütermenge bringt A bei gleicher Anfangsausstattung mehr zusätzlichen Nutzen als B.<sup>1</sup>

Die Tatsache, daß die erste Ableitung der Nutzenfunktion von A größer ist als die der Nutzenfunktion von B allein, würde diese Interpretation keinesfalls rechtfertigen. Das zeigt sich für  $x < 1$ , wo zwar weiterhin  $u_1'(x) > u_2'(x)$ , aber nun  $u_1(x) < u_2(x)$ . Gerade das zweite Merkmal,  $u_1'(x) > u_2'(x)$  für  $x \in (1, x_0]$ , ist jedoch entscheidend dafür, daß A den größeren Anteil in einer utilitaristischen Gesellschaft bekommt als B.

Die Zunahme des Gesamtnutzens infolge des zusätzlich aufgeteilten Gutes  $dx$  ist bei der Verteilung  $\langle x, x_0 - x \rangle$  maximal, bei der  $u_1'(x) + u_2'(x_0 - x)$  maximal ist, was in unserem Fall für  $u_1'(x) = u_2'(x_0 - x)$  bzw. bei  $x = \frac{2}{3}x_0$  zutrifft.<sup>2</sup>

---

<sup>1</sup> Ob "A ist glücksfähiger als B in I" eine sinnvolle Interpretation für " $\forall \vec{x} \in I : u_1(\vec{x}) > u_2(\vec{x}), I \subseteq \mathbb{R}^k$ " bei einer Welt mit  $k$  Gütern ist, hängt von der empirischen Fundierung der Nutzentheorie ab. Die moderne Nutzentheorie von v. Neumann und Morgenstern rechtfertigt diese Interpretation jedenfalls in keiner Weise. Konstitutiv dafür wären Vergleichbarkeitsannahmen der Art "die Alternative  $x$  bringt im Vergleich zu  $y$  der Person A mehr zusätzlich Glück als der Person B" etc.

Die moderne Nutzentheorie hat gezeigt, daß sich die Präferenzen einer Person, soweit sie bestimmte (Rationalitäts-) Bedingungen erfüllen, durch eine reellwertige Funktion darstellen lassen. In analoger Weise kann die Vergleichbarkeit kardinaler individueller Bewertungsfunktionen durch entsprechende interpersonelle Ordnungsrelationen ("ist für Person A günstiger als für Person B") sichergestellt werden. Und so wie die Interpretation der von Neumann/Morgensternschen Nutzenfunktionen von dem Charakter der zugrundegelegten Präferenzen abhängt, so hängt die Interpretation numerischer Unterschiede interpersonell vergleichbarer Bewertungsfunktionen von den zugrundegelegten singulären Vergleichen ab.

<sup>2</sup>  $U(x_0, x) := u_1(x) + u_2(x_0 - x) ; 0 < x < x_0$   
 $\frac{\partial}{\partial x} U(x_0, x) = \frac{\partial}{\partial x} u_1(x) + \frac{\partial}{\partial x} u_2(x) = \frac{2}{x} - \frac{1}{x_0 - x} = 0 \Leftrightarrow x = \frac{2}{3} x_0$   
das ist gerade der Fall bei  $\frac{\partial}{\partial x} u_1(x) = \frac{\partial}{\partial x} u_2(x)$ .

$u_1$  und  $u_2$  sind dabei typische Nutzenfunktionen: konvex und monoton steigend. Für lineare Funktionen  $u_1 = 2x$  und  $u_2 = x$  stünde B überhaupt kein Güteranteil aufgrund utilitaristischer Ethik zu. Modifizieren wir die Funktionen  $u_1$  und  $u_2$  ein wenig,  $u_1(x) = 2 \cdot \ln x$  und  $u_2(x) = \ln x$  für  $x \geq 0$ , so ist A für  $x \in (0, 1)$  schlechter gestellt als B. An der utilitaristischen Verteilung  $\langle \frac{2}{3}x_0, \frac{1}{3}x_0 \rangle$  ändert das nichts.

Das Nutzenniveau von A ist für  $x < 1$  niedriger als das von B bei der gleichen Güterausstattung. Bei der utilitaristisch gebotenen Verteilung geht es A besser als B, wenn die Gesamtgütermenge  $x_0 > \frac{3}{4}$ .<sup>1</sup> Für  $x_0 \in (\frac{3}{4}, \frac{3}{2})$  wird derjenige durch die utilitaristische Verteilung bevorzugt, dem es bei Gleichverteilung der Güter schlechter gegangen wäre. Für alle anderen Größen der Gesamtgütermenge bleibt derjenige, der mit der gleichen Güterausstattung ein höheres Nutzenniveau verbindet, auch in der utilitaristischen Verteilung der Bessergestellte. Interessant ist es noch, die Veränderung gegenüber der Gleichverteilung in Abhängigkeit von  $x_0$  zu betrachten. Für  $x_0 > 3$  verschlechtert sich die Situation des schon bei Gleichverteilung Benachteiligten in der utilitaristisch gebotenen Verteilung. Für  $x_0 < \frac{3}{2}$  verbessert sich die Situation des bei Gleichverteilung Schlechtergestellten. Aber nur für  $x_0 > \frac{3}{4}$  reicht diese Verbesserung aus, um ihn über das Nutzenniveau des bei Gleichverteilung Bessergestellten zu bringen.

Ganz andere sozialetische Konsequenzen würden sich ergeben, wenn die individuellen Nutzenfunktionen beschränkt und normiert wären, z.B. auf  $[0, 1]$ .

---

<sup>1</sup>  $2 \ln(\frac{2}{3}x_0) - \ln(\frac{1}{3}x_0) > 0$   
 $2 \ln 2 + \ln \frac{1}{3}x_0 > 0$   
 $\ln \frac{1}{3}x_0 > \ln 2^{-2}$   
 $x_0 > \frac{3}{4}$



$$U_1(x) = a_1 + b_1 U_1(x) ; a_1 \in \mathbb{R} , b_1 \in \mathbb{R}^+$$

$$\min \{U_1(x)\} = 0 = \min \{U_1(x)\} \Rightarrow a_1 = 0$$

$$\max \{U_1(x)\} = 1 = b_1 \cdot 2 \cdot \ln x_0 \Rightarrow b_1 = \frac{1}{2 \ln x_0} .$$

Analog ergibt sich für B:  $a_2 = 0$  und  $b_2 = \frac{1}{\ln x_0}$  .

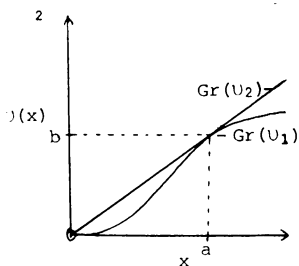
In diesem Beispiel hat das die Konsequenz:

$$\forall x \in [0, x_0] : U_1(x) = U_2(x) .$$

Die utilitaristisch gebotene Verteilung ist jetzt  $\langle 0,5x_0, 0,5x_0 \rangle$ .

Es wäre übrigens ein Irrtum, zu glauben, Gleichverteilung sei genau dann utilitaristisch geboten, wenn die individuellen Nutzenfunktionen gleich sind. Es gilt nicht einmal eine Richtung: interpersonelle Gleichheit ist weder eine notwendige noch eine hinreichende Bedingung dafür, daß Gleichverteilung utilitaristisch geboten ist. Sind alle individuellen Nutzenfunktionen interpersonell gleich, konkav und nach oben unbeschränkt, so wäre es sogar utilitaristisch geboten, einer Person alles und den übrigen nichts zu geben.<sup>1</sup> Interpersonelle Gleichheit ist aber auch keine notwendige Bedingung, wie man sich anhand geeigneter Nutzenfunktionen rasch klar machen kann.<sup>2</sup>

<sup>1</sup> Man kann davon ausgehen, daß konkave unbeschränkte Nutzenfunktionen empirisch nicht vorkommen. Daher ist es fraglich, ob solche intuitiv nicht überzeugende normative Empfehlungen einer sozialetischen Theorie in kontrafaktischen Einzelfällen geeignet sind, diese Theorie zu 'widerlegen'. Angenommen, eine Art des Utilitarismus würde für die empirisch gegebenen subjektiven Bewertungsfunktionen durchweg mit unseren moralischen Basisurteilen übereinstimmende Ergebnisse haben, für manche empirisch auszuschießende Interaktionsstrukturen jedoch inakzeptable Empfehlungen geben, könnte diese Art des Utilitarismus dann nicht dennoch als die angemessene sozialetische Theorie akzeptiert werden?



In der Zwei-Personen-Welt der nebenstehenden Abbildung z.B. wäre bei einer Gesamtgütermenge  $x_0 = 2a$  Gleichverteilung utilitaristisch geboten, mit  $U(x, x_0) = 2b$  als maximalem Gesamtnutzen.

Nun kann eine Methode interpersoneller Nutzenvergleiche nicht zureichend dadurch allgemein begründet werden, daß sie die intuitive Plausibilität einer speziellen Theorie normativer Sozialethik erhöht. Immerhin besteht aber die Möglichkeit, die Methode interpersoneller Nutzenvergleiche jeweils relativ zu einer Theorie der Sozialethik zu wählen und zu rechtfertigen. Im Rahmen des Utilitarismus wären damit verschiedene normative Theorien möglich, die sich ausschließlich in der gewählten Vergleichsmethode unterscheiden: Die Wahl einer Vergleichsmethode wäre damit Teil der normativen Theorie selbst und könnte mit moralischen Argumenten verteidigt werden. Der Status der Vergleichsmethode wäre in dieser Interpretation ein normativer. Natürlich kann man auch den Versuch machen, eine Vergleichsmethode als allgemeines Metakriterium für beliebige teleologische Theorien normativer Sozialethik zu bestimmen. So könnte die Methode der Normierung beschränkter Nutzenfunktionen (nur die Beschränktheit ist eine empirische Annahme) auf ein interpersonell invariantes Intervall<sup>1</sup> mit Gleichheitsprinzipien allgemeiner Art und ohne direkten Bezug auf konkrete sozialetische Aggregationsregeln gerechtfertigt werden. Auch in diesem Fall bliebe die Wahl der Vergleichsmethode ein Gegenstand normativ-ethischer Beurteilung.

Anders ist es, wenn die sozialetische Informationsbasis als empirisches Problem behandelt wird, was intuitiv naheliegend ist.

Auch wenn die Existenz eines absoluten Nutzens bejaht wird, ist in diesem Fall die Frage der Meßbarkeit noch in keiner Weise geklärt. Unbestreitbar ist man sich intersubjektiv häufig einig in der festen Überzeugung, ein bestimmtes Ereignis  $e$  habe für die eine Person  $A$

---

<sup>1</sup> I.a. *offenes* Intervall, da nicht vorausgesetzt werden kann, daß der Wertebereich einer individuellen Nutzenfunktion ein maximales Element enthält.

weit gravierendere Konsequenzen als für B, m.a.W.:  $\epsilon$  verursacht eine weit größere Nutzenniveauveränderung bei A als bei B. Die Abwägung, welches als überwiegendes Interesse in einem Konflikt gelten kann, ist für alltägliche moralische Entscheidungen von großer Bedeutung, und es steht daher zu vermuten, daß auch ethische Theorien ohne Berücksichtigung solcher Differenzierung inadäquat bleiben müssen. Ein Mindestmaß an interpersoneller Vergleichbarkeit subjektiver Bewertungen erscheint daher zumindest für teleologische Theorien unverzichtbar. Die modernen Wohlfahrts- und Sozialwahltheorien haben sich eines Großteils ihrer normativen Relevanz durch eine radikale Informationsbeschränkung beraubt. Ein Teil der sozialwahltheoretischen 'Aporien' ist Folge des Verzichts auf interpersonelle Vergleichbarkeit subjektiver Bewertungen.

So ist z.B. die Anonymitätsbedingung in Arrows bekanntem Möglichkeitstheorem<sup>1</sup> eine naheliegende normative Konsequenz dieser Informationsbeschränkung. Gleichbehandlung impliziert Unterschiedliches, je nachdem, wieviel von den Gleichzubehandelnden als bekannt vorausgesetzt wird. Sind nur individuelle Präferenzrelationen bekannt, muß etwa Franzens Präferenz für Monika als seine zukünftige Angetraute gleich viel gelten für die Entscheidung dieser Frage, wie die Präferenz seiner Großtante für Ursula. Liberalitätsbedingungen, die sicherstellen sollen, daß wenigstens einige Entscheidungen von denen getroffen werden, die sie (in erster Linie) betreffen,<sup>2</sup> sind ein recht schwacher Ersatz für die oft schwierige, aber für viele moralische Fragen unverzichtbare Abwägung konkurrierender Interessen - und eine solche Abwägung heißt, subjektive Bewertungen interpersonell vergleichen.<sup>3</sup>

---

<sup>1</sup> Vgl. Arrow (1963), S.46-60

<sup>2</sup> Vgl. A.K. Sen, *Collective choice and social welfare*, San Francisco 1970, Kap.6 und 6\*, S.78-88

<sup>3</sup> Vgl. dazu etwa P.J.Hammond, *Why ethical measures of inequality need interpersonal comparisons*, *Theory and decision* 7 (1976)263-274 und K.Roberts, *Interpersonal comparability and social choice theory*, *Review of economic studies* 48 (1980) 421-439

Mit diesen Bemerkungen ist das Meßbarkeitsproblem natürlich in keiner Weise gelöst.<sup>1</sup> Sie sollten auf die Relevanz interpersoneller Vergleiche subjektiver Bewertungen für ethische Theorien hinweisen, die eine rasche Kapitulation bei Meßproblemen verbietet, zumal die Möglichkeit interpersoneller Nutzenvergleiche von Wohlfahrts- und Sozialwahltheoretikern aufgrund methodologischer Überzeugungen verneint wird, die als überholt gelten können.<sup>2</sup> Der Begriff 'Nutzen' wird nicht erst dadurch auch für wissenschaftliche Aussagensysteme sinnvoll, daß er mithilfe unmittelbar beobachtbarer Größen definierbar ist.<sup>3</sup>

---

<sup>1</sup> Vgl. z.B. L. Robbins, *An essay on the nature and significance of economic science*, London 1932. Methodologische Kritik dieser Art an der Konzeption interpersonell vergleichbarer Nutzenfunktionen führte zur Neuen Wohlfahrtsökonomie in den 30er Jahren, die sich auf ordinale Nutzenfunktionen beschränkte. Vgl.a. F. Alt, *Über die Meßbarkeit des Nutzens*, Zeitschrift für Nationalökonomie 7 (1936), englisch erneut veröffentlicht in: Chipman u.a. (1971), *On the measurability of utility*, 424-431

<sup>2</sup> Die 'ordinalistische' Schule konnte sich zunächst u.a. auf Rudolf Carnap, *Der logische Aufbau der Welt* (1928) stützen. Carnap revidierte jedoch die Reduzierbarkeitsthese schon wenige Jahre darauf, vgl. R. Carnap, *Testability and meaning*, *Philosophy of science* 3 (1936) 420-471 und 4 (1937) 2-40; ders., *The methodological character of theoretical concepts*, in: *Minnesota studies in the philosophy of science*, Bd.1, Minneapolis 1954, 38-76; vgl. dazu a. C.G. Hempel, *Aspects of scientific explanation*, New York 1965; W. Stegmüller, *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Bd.2, *Theorie und Erfahrung*, Berlin/Heidelberg/New York 1970, bes. S.213-254; W.K. Essler, *Wissenschaftstheorie*, Bd.1, *Definition und Reduktion*, München 1982

<sup>3</sup> Zum 'psycho-physikalischen' Aspekt der Meßbarkeitsfrage vgl. die Überlegungen und Forschungsergebnisse von L.L. Thurstone, bes. *The measurement of values*, *Psychological review* 61 (1954) 47-58 u. ders., *The rational origin of measuring subjective values*, in der Aufsatzsammlung zu diesem Thema Thurstone, *The measurement of values*, Chicago 1959, 182-194; aber auch R.D. Bock/ L.V. Jones, *The measurement and prediction of judgement and choice*, San Francisco 1968; vgl. dazu auch W. Spohn, *Grundlagen der Entscheidungstheorie*, Kronberg/Ts. 1978, S.196-200

Ökonomische Gerechtigkeits- und Fairneßtheorien machen seit einigen Jahren Gebrauch von sog. 'erweiterten individuellen Ordnungen', die Patrick Suppes 1966 mit seinem 'grading principle of justice' in die Diskussion eingeführt hat.<sup>1</sup> Da sich das Problem intersubjektiver Vergleichbarkeit mithilfe dieses Begriffs wesentlich abschwächen läßt, wollen wir auf diese Möglichkeit zum Schluß des Abschnittes kurz eingehen.

Üblicherweise wird zunächst der Grundbegriff einer (gesellschaftlichen) Position  $p \in P$  eingeführt; eine individuelle Position, bzw. der Sachverhalt, daß sich eine Person in einer Position  $p$  befindet, wird mit  $\langle p, i \rangle$  symbolisiert; eine gesellschaftliche Situation  $s$  ist damit eine Funktion von  $K$  in  $P$ .

In den meist behandelten Fällen ist  $p$  ein Güterbündel und  $\langle p, i \rangle$  die Verfügbarkeit des Güterbündels für  $i$ . Eine gesellschaftliche Situation ist damit nichts anderes als eine bestimmte Güterallokation.

Wir wollen nun annehmen, daß jede Person eine vollständige Präferenzordnung  $\hat{R}$  über  $P$ , eine vollständige subjektive Präferenzordnung  $\hat{R}_i$  über  $P \times \{i\}$ , ein vollständige erweiterte subjektive Präferenzordnung  $\tilde{R}_i$  über  $P \times K$  und eine Präferenzrelation  $R_i$  über  $S$ , die Menge der gesellschaftlichen Situationen, hat.

---

<sup>1</sup> Vgl. Patrick Suppes, *Some formal models of grading principles*, Synthese 16 (1966) 284-306. Vgl.a. D. Foley, *Resource allocation and the public sector*, Yale economic essays 7 (1967); Sen (1970a), S.146-160; E. Pazner/D. Schmeidler, *A difficulty in the concept of fairness*, Review of economic studies 41 (1974); H.R. Varian, *Equity, envy and efficiency*, Journal of economic theory 9 (1974), 63-91; ders., *Distributive justice, welfare economics and the theory of fairness*, Philosophy of public affairs 4 (1974/75) 223-247; E. Pazner, *Recent thinking on economic justice*, Journal of peace science 2 (1976) 143-153; ders., *Pitfalls in the theory of fairness*, Journal of economic theory 14 (1977) 458-465; L.G. Svensson, *Social justice and fair distributions*, Lund 1977; Pazner/Schmeidler, *Egalitarian equivalent allocations: A new concept of economic equity*, Quarterly journal of economics 92 (1978) 671-687; Pazner, *Equity, nonfeasible alternatives and social choice: A reconsideration of the concept of social welfare*, in: Laffont (1979) 161-173

Betrachten wir einige mögliche Postulate intra- und interpersoneller Präferenzen-Interdependenz:

- (1)  $\forall i \in K \forall p, p' \in P : \langle p, p' \rangle \in \dot{R}_i \Leftrightarrow \langle \langle p, i \rangle, \langle p', i \rangle \rangle \in \hat{R}_i$
- (2)  $\forall i \in K \forall p, p' \in P : \langle \langle p, i \rangle, \langle p', i \rangle \rangle \in \hat{R}_i \Leftrightarrow \langle \langle p, i \rangle, \langle p', i \rangle \rangle \in \tilde{R}_i$
- (3)  $\forall i, j \in K : \forall p, p' \in P : \langle \langle p, i \rangle, \langle p', i \rangle \rangle \in \hat{R}_i \Rightarrow \langle \langle p, i \rangle, \langle p', i \rangle \rangle \in \tilde{R}_j$
- (4)  $\exists i, j, k \in K \exists p, p' \in P : \langle \langle p, i \rangle, \langle p', j \rangle \rangle \in \tilde{R}_k \Rightarrow \forall l \in K : \langle \langle p, i \rangle, \langle p', j \rangle \rangle \in \tilde{R}_l$
- (5)  $\forall i \in K \forall p, p' \in P : \langle \langle p, j \rangle, \langle p', k \rangle \rangle \in \tilde{R}_i \Leftrightarrow \langle \langle p, i \rangle, \langle p', i \rangle \rangle \in \hat{R}_i$
- (6)  $\forall i \in K : \langle s, s' \rangle \in R_i \Leftrightarrow \langle \langle s(i), i \rangle, \langle s'(i), i \rangle \rangle \in \hat{R}_i$
- (7)  $\forall i \in K \forall s, s' \in S : (\exists f \in \text{Bij}(K, K) \forall j \in K : \langle \langle s(j), j \rangle, \langle s'(f(j)), f(j) \rangle \rangle \in \tilde{R}_i \Rightarrow \langle s, s' \rangle \in R_i) \wedge (\exists f \in \text{Bij}(K, K) \forall j \in K : \langle \langle s(j), j \rangle, \langle s'(f(j)), f(j) \rangle \rangle \in \tilde{R}_i \wedge \exists k \in K : \langle \langle s'(k), s(k) \rangle \rangle \in \tilde{R}_k \Rightarrow \langle s, s' \rangle \in R_i \wedge \langle \langle s', s \rangle \rangle \in R_i)$

Ob (1) und (2) intuitiv selbstverständliche Postulate sind oder nicht, hängt von der Interpretation der drei Präferenzbegriffe ab. Eine individuelle Präferenzrelation  $\dot{R}_i$  über  $P$  sollte so interpretiert werden, daß (1) nicht analytisch wahr ist. Und daher fragt es sich, ob  $\dot{R}_i$  überhaupt einer eigenständigen Interpretation fähig ist: Kann ich gesellschaftliche 'Positionen' bzw. Güterausstattungen per se in eine Rangordnung bringen, oder beurteile ich dabei nicht entweder, was ich persönlich bevorzuge (womit (1) analytisch wäre), oder welcher Wertschätzung sich die Positionen in der Gesellschaft i.a. erfreuen, womit ich eine statistische Aussage über die Verteilung der unterschiedlichen logisch möglichen  $\hat{R}_i$  machen würde? Eine Alternative wäre,  $\dot{R}_i$  objektiv und  $\hat{R}_i$  subjektiv zu interpretieren.  $\dot{R}_i$  würde dann die Ansichten von  $i$  darüber widerspiegeln, welche Güterausstattung er für objektiv besser hält<sup>1</sup>, unabhängig davon, welche er

<sup>1</sup>  $\dot{R}_i$  wäre damit Ergebnis der *außermoralischen Werturteile* von  $i$ . Für die *moralischen Werturteile* ist  $R_i$  reserviert.

persönlich - beschaffen wie er nun einmal ist - tatsächlich vorzieht, wie es seine Präferenzordnung  $\hat{R}_i$  zeigt. Damit ist es aber zweckmäßig, auch die erweiterte individuelle Präferenzordnung  $\tilde{R}_i$  subjektiv zu interpretieren, wodurch wenigstens (2) analytisch wäre.

Postulat (3) hingegen garantiert ein gewisses Maß subjektiver interpersoneller Vergleichbarkeit, denn wenn (3) erfüllt ist, stimmen die Personen darin überein, welche Position  $p$  für welche Personen günstiger als eine andere Position  $p'$  ist. Man könnte auch sagen, wenn (3) erfüllt ist, gibt es keinen Streit über die jeweiligen persönlichen Interessen, es besteht Konsens in der Analyse der Interessenstruktur.

Postulat (4) läßt zwar zu, daß sich die Wertschätzung der Positionen von Person zu Person unterscheidet, verlangt aber interpersonell gleiche erweiterte Präferenzordnungen  $\tilde{R}_i$ . (3) ist eine notwendige Bedingung für (4), aber (4) verlangt über (3) hinaus vollständigen interpersonellen Konsens über die Anordnung individueller Positionen: Für beliebige Personen  $i, j \in K$  und Positionen  $p, p' \in P$  besteht Einigkeit darin, ob es  $i$  in  $p$  mindestens so gut wie  $j$  in  $p'$  geht oder umgekehrt:

$\forall k \in K : \langle \langle p, i \rangle, \langle p', j \rangle \rangle \in R_k \vee \langle \langle p', i \rangle, \langle p, j \rangle \rangle \in R_k$  ; und wenn einer meint, es ginge einer Person  $i$  in  $p$  besser als einer anderen Person  $j$  in  $p'$ , so sind alle übrigen derselben Ansicht.

(3) und (4) garantieren in unterschiedlichem Ausmaß subjektive (ordinale) interpersonelle Vergleichbarkeit. 'Subjektiv' deshalb, weil Konsens Irrtum nicht ausschließt. Beziehen sich normative Aggregationsregeln auf tatsächliche und nicht auf subjektive, u.U. nur vermeintliche, individuelle Interessen, so geben (3) und (4) keine zuverlässige sozialethische Informationsbasis ab.

Eine subjektivistische Sozialethik<sup>1</sup> dagegen könnte sich mit der durch (4) garantierten ordinalen interpersonellen Vergleichbarkeit vollauf zufrieden geben, denn die mangelnde Kardinalität hat bei einer hinreichend großen Personengruppe und einer hohen Zahl von gesellschaftlichen Positionen in der interpersonellen Positionenvergleichbarkeit einen guten Ersatz gefunden: Die interpersonell gleichen erweiterten Ordnungen legen so viele 'Teilungsstriche' auf der Bewertungsskala fest, wie starke Präferenzen in  $\tilde{R}_i$  enthalten sind. Nimmt man daher an, daß es eine minimale Unterscheidbarkeitsgrenze  $k > 0$  gibt, und ist die Bewertungsskala beschränkt, sind auch positiv-monotone Transformationen ('Verzerrungen' der Skala), die mit  $\tilde{R}_i$  übereinstimmen, nur begrenzt möglich.<sup>2</sup>

Gilt (5), orientiert sich jede Person bei der Beurteilung der gesellschaftlichen Positionen ausschließlich an ihren eigenen Bewertungen nach dem Motto: was mir gut tut, wird auch anderen gut tun. Für die (subjektive) interpersonelle Vergleichbarkeit wäre diese Art intrapersoneller Präferenzeninterdependenz in einem Fall völlig unproblematisch, nämlich dann, wenn die Präferenzordnungen  $\hat{R}_i$  interpersonell gleich wären. Sind die persönlichen Vorlieben und Bewertungen gleich (oder sehr ähnlich), ist es nicht nötig, die Eigenheiten anderer Personen zu berücksichtigen, um interpersonelle Vergleiche anzustellen.

---

<sup>1</sup> Bisweilen auch als individualistischer Ansatz der collective-choice-Theorien gepriesen, vgl. E. Schlicht, *Die Theorie der kollektiven Entscheidung und der individualistische Ansatz*, Leviathan 2 (1974) 263-280. Wir gehen darauf in §11 noch genauer ein.

<sup>2</sup> Vgl. die Präferenzschwellentheorie von Georgescu-Roegen und Armstrong zur Begründung kardinaler Nutzenfunktionen. Die Kardinalisierung der erweiterten Ordnungen nach der Armstrong-Methode führt, wenn (4) erfüllt ist, zur subjektiven vollen interpersonellen kardinalen Vergleichbarkeit individueller Präferenzen, und zwar auch dann, wenn mit der Präferenzschwellentheorie allein interpersonelle Vergleiche nach der Goodman-Markowitz-Methode nicht



Zwischen (5) und (6) gibt es einen Zusammenhang: Wenn eine Person im interpersonellen Vergleich ausschließlich ihre eigenen Präferenzen im Sinne von (5) berücksichtigt, ist es naheliegend, auch die Präferenzen über die gesellschaftlichen Situationen  $R_i$  ausschließlich an den eigenen Interessen auszurichten.

(7) fordert bestimmte Präferenzen  $R_i$  über die Menge der gesellschaftlichen Situationen  $S$  in Abhängigkeit von den individuellen erweiterten Ordnungen  $\tilde{R}_i$ . Welche Abhängigkeit das genau ist, läßt sich wortsprachlich nur sehr umständlich formulieren. (7) fordert u.a., daß  $\langle s, s' \rangle \in R_i$ , wenn  $\forall i \in K : \langle \langle s(j), j \rangle, \langle s'(j), j \rangle \rangle \in \tilde{R}_i$ , bzw. daß man eine gesellschaftliche Situation dann vorziehen sollte, wenn man selbst der Ansicht ist, daß es allen Personen in ihr besser geht. Dieser Sonderfall kann als subjektivistische Interpretation der Pareto-Regel interpretiert werden, womit (7) als eine Art erweitertes Pareto-Prinzip erscheint. Tatsächlich kann (7) jedoch als eine interessante Konzeption ethischer Präferenzen angesehen werden,<sup>1</sup> die weit über die Bedeutung der Pareto-Regel hinausgeht. Insbesondere enthält (7) eine Anonymitäts- oder Symmetrieforderung: Wenn sich von  $s$  zu  $s'$  zwar individuelle Positionen ändern, i jedoch der Ansicht ist, daß dabei insgesamt nur interpersonelle Vertauschungen stattgefunden haben, so ergibt sich aufgrund von (7) auch keine unterschiedliche Bewertung von  $s$  und  $s'$ . (7) verlagert die zur Festlegung ethischer Präferenzen notwendigen interpersonellen Vergleiche in die intrapersonelle Beziehung zweier Arten individueller Präferenzrelationen.

---

gerechtfertigt werden könnten, vgl. S.N. Georgescu-Roegen, *The pure theory of consumer's behavior*, Quarterly journal of economy 1 (1936) 545-593; W.E. Armstrong, *The determinateness of the utility function* Economic journal 49 (1939) 453-467; ders., *Utility and the theory of welfare*, Oxford economic papers, new series 3 (1951) 259-271; L.A. Goodman/H. Markowitz, *Social welfare functions based on individual rankings*, American journal of sociology 18 (1952) 257-262.

<sup>1</sup> Sie entspricht Suppes' grading principle of justice, vgl. Suppes (1966).

## 7, DIE INTERDEPENDENZPROBLEMATIK NORMATIVER AGGREGATIONSTHEORIEN

*"Game theory is capable of providing a formal analysis only of those situations in which egoistic values are to be maximized."*

Wolff (1962)

In diesem Abschnitt geht es um die Rolle der Interdependenz individueller Präferenzen für (normative) Aggregationstheorien.

Mit Präferenzeninterdependenz wird dabei das Phänomen bezeichnet, daß Personen ihre Präferenzen von den Präferenzen anderer Personen abhängig machen. Normative Aggregationstheorien, die sich einen uniformen Präferenzbegriff in Anlehnung an die moderne Nutzentheorie (nicht aufgrund methodologischer Erwägungen) zu eigen machen, geraten in folgendes Dilemma: Entweder sie setzen voraus, daß Präferenzeninterdependenz keine wesentliche Rolle für das menschliche Verhalten spielt, dann stellen sie zugleich die ethische (zumindest individualethische) Relevanz ihres Ansatzes in Frage, oder sie aggregieren kategorial verschiedene Präferenzen in gleicher Weise, was inadäquat, ja im schlimmsten Falle sogar zirkulär ist.

Dies ist ein wesentlicher Grund für eine Differenzierung des Präferenzbegriffes im Rahmen normativer Aggregationstheorien.

Was mit Präferenzeninterdependenz gemeint ist, läßt sich am besten an einer alltäglichen Entscheidungssituation erläutern: Nehmen wir an, A und B seien befreundet und auf Reisen. Es stellt sich für sie eines Tages die Frage, gemeinsam das Nationalmuseum in Neapel zu besuchen ( $p_1$ ),

oder an einen Strand der nahegelegenen costa amalfitana zu gehen ( $p_2$ ). Es ergeben sich für das Paar zwei kollektive Entscheidungsmöglichkeiten:  $p_1 > p_2$  oder  $p_2 > p_1$ , da wir annehmen wollen, daß die Liaison so jung ist, daß wir die Alternativenmenge P nicht um  $p_3$  (A an den Strand, B ins Museum) und  $p_4$  (B an den Strand, A ins Museum) erweitern müssen. Ist die Liebe groß, kommen sie vermutlich in das nämliche Dilemma, das wir mithilfe einiger Begriffsdifferenzierungen für die Sozialethik, insbesondere die der kollektiven Präferenzen, ausschließen möchten: die beiden wissen nicht, was sie wollen. A würde mit B lieber ins Museum gehen, wenn B lieber ins Museum will und umgekehrt: A würde lieber an den Strand gehen, wenn B lieber an den Strand will. Symmetrisches gilt für B. Eine solche Situation läßt sich formal folgendermaßen beschreiben:

- (1)  $p_1 \underset{2}{>} p_2 \Rightarrow p_1 \underset{1}{>} p_2$
- (2)  $p_2 \underset{2}{>} p_1 \Rightarrow p_2 \underset{1}{>} p_1$
- (3)  $p_1 \underset{1}{>} p_2 \Rightarrow p_1 \underset{2}{>} p_2$
- (4)  $p_2 \underset{1}{>} p_1 \Rightarrow p_2 \underset{2}{>} p_1$

In einem solchen Fall sprechen wir von zirkulär bestimmten Präferenzen.

Verallgemeinert auf eine Gruppe von Personen K ist  $\{R_i | i \in K_0\}$ ,  $K_0 \subset K$  eine Menge zirkulär bestimmter Präferenzrelationen gdw.

$$\begin{aligned} & \exists \Xi_0 \subset \Xi : \exists \phi \in \text{Bij.}(\{1, 2, 3, \dots, \#K_0\}, K_0) : \\ & \exists \psi \in \text{Bij.}(\{1, 2, 3, \dots, \#K_0\}, \Xi_0 \times \Xi_0) : \\ & | (\forall v \in \{1, 2, 3, \dots, \#K_0 - 1\} : \psi(v) \in R_{\phi(v)} \Rightarrow \psi(v+1) \in R_{\phi(v+1)} \wedge \\ & \wedge (\psi(\#K_0) \in R_{\phi(\#K_0)} \Rightarrow \psi(1) \in R_{\phi(1)}) \end{aligned}$$

Am Ende des Urlaubs, wenn die Liebe schon merklich abgekühlt ist, und möglicherweise schon Zeiten gegenseitiger Aversion vorkommen, mag sich in Rom eine ähnliche Situation mit ganz anderen Konsequenzen ergeben: Wieder geht es um die Frage, ob sie heute nach Ostia fahren ( $p_1$ ) oder die Sixtinische Kapelle besuchen ( $p_2$ ) sollen. Jetzt aber

will A gerade dann nach Ostia, wenn es B bevorzugt, die Sixtinische Kapelle zu besuchen. Andererseits hat A ein besonderes Interesse an der Sixtinischen Kapelle gerade dann, wenn sich B in Ostia sonnen möchte.

Sind B's Präferenzen in symmetrischer Weise von A's Präferenzen abhängig, ergibt sich folgende Präferenzeninterdependenz:

$$(5) p_{1,2} > p_2 \Rightarrow p_{2,1} > p_1$$

$$(6) p_{2,2} > p_1 \Rightarrow p_{1,1} > p_2$$

$$(7) p_{1,1} > p_2 \Rightarrow p_{2,2} > p_1$$

$$(8) p_{2,1} > p_1 \Rightarrow p_{1,2} > p_2$$

Je nachdem, zu welcher anfänglichen wechselseitigen Einschätzung der Präferenzen es kommt, bleibt es bei stabilen, aber entgegengesetzten Präferenzen.

Hat sich dagegen die Abhängigkeit der Präferenzen von B seit Urlaubsbeginn nicht geändert, gilt also (3), (4), (5) und (6), handelt es sich um eine logisch widersprüchliche Präferenzstruktur.

In diesem Fall sprechen wir von kontradiktorisch bestimmten Präferenzen.

$\{R_i | i \in K_0\}$  ist eine Menge kontradiktorisch bestimmter Präferenzrelationen,  $\Leftrightarrow$

$$\exists \Xi_0 \subset \Xi : \exists \phi \in \text{Bij.}(\{1, 2, 3, \dots, \#K_0\}, K_0) :$$

$$\exists \psi \in \text{Bij.}(\{1, 2, 3, \dots, \#K_0\}, \Xi_0 \times \Xi_0) :$$

$$[(\forall v \in \{1, 2, 3, \dots, \#K_0 - 1\} : \psi(v) \in R_{\phi(v)} \Rightarrow \psi(v+1) \in R_{\phi(v+1)}) \wedge \wedge (\psi(\#K_0) \in R_{\phi(\#K_0)} \Rightarrow \sim \psi(1) \in R_{\phi(1)})]$$

Präferenzeninterdependenz<sup>1</sup> kann unterschiedliche Ursachen haben: Zuneigung, Abneigung, Verantwortungsgefühl, mora-

---

<sup>1</sup> Gegen das "general possibility theorem for social welfare functions" von Arrow, das aufzuzeigen scheint, daß es keine vernünft-

lische Überzeugungen etc.. Unabhängig von den möglichen Ursachen und Gründen stellt sich die Frage, welche Konsequenzen wechselseitige Präferenzeninterdependenz hat. Eine Hypothese liegt nahe: Wenn die Personen ihre subjektiven Bewertungen der Zustände von den subjektiven Bewertungen (dieser oder anderer Zustände) anderer Personen abhängig machen, gibt es jeweils einen Gleichgewichtspunkt, in dem diese wechselseitige Abhängigkeit nicht mehr zu einer Veränderung der Bewertungen führt. Diese Hypothese ist jedoch - wie wir gleich zeigen - falsch, und entsprechend dringlich stellt sich die Frage, auf was sich normative Aggregationsregeln eigentlich beziehen. Das soll im Folgenden genauer untersucht werden.

tige demokratische Methode der Aggregation individueller Präferenzen gibt (vgl. K.J. Arrow, *Social choice and individual values*, New York/London/Sidney 1963, Kap.5), hat Gordon Tullock argumentiert, daß die Annahme eines einfachen und wahrscheinlichen Typs der 'Interdependenz' individueller Präferenzfunktionen in größeren Gruppen ausreicht, um mit hoher Wahrscheinlichkeit eine kollektive Wahl zu erhalten, die nur unwesentlich von der kollektiv besten abweicht. Bei vernünftigen empirischen Annahmen sollte damit die allgemeine Irrelevanz des Unmöglichkeitstheorems aufgezeigt werden. Vgl. G. Tullock, *The general irrelevance of the general impossibility theorem*, *Quarterly journal of economics* 81 (1967) 256-270.

In unserem Zusammenhang ist es wichtig, die dort angesprochene Interdependenz individueller Präferenzen sorgfältig von der in diesem Abschnitt diskutierten zu unterscheiden. So wird z.B. die 'single-peakedness' von Präferenzen, wie sie u.a. von Newing und Black untersucht wurde (vgl. D. Black, *The theory of committees and elections*, Cambridge 1958), als eine Art von Präferenzeninterdependenz bezeichnet, aber das entspricht nicht der hier verwendeten Bedeutung. Die Idee ist, daß man die Zustandsmenge durch einen mehrdimensionalen cartesischen Raum, in dem die Achsen bestimmte Merkmale repräsentieren, darstellen kann, und dann jedes Individuum "some optimal point in the resulting dimensional space" hat, so "that the individual's degree of satisfaction falls off as we move away from his optima in any direction". Tullock (1967), S.267.

Die von Tullock diskutierten und die darüberhinaus möglichen Abschwächungen brauchen hier nicht dargestellt werden, denn es liegt auf der Hand, daß Bedingungen dieser Art an die Gestalt der individuellen Präferenzrelationen keine Interdependenzannahmen in unserem Sinne sind; sie sind mit Arrow besser als Ähnlichkeitsannahmen zu bezeichnen. Vgl. Arrow (1963), S.74-80.

INTERPERSONELLE NUTZENINTERDEPENDENZ

Zustände  $\xi \in \Xi$  sind für uns mögliche Welten. Zwei mögliche Welten  $\xi, \xi'$  sind schon dann verschieden, wenn in  $\xi$  die Person A die eine, und in  $\xi'$  eine andere Präferenz hat.  $\xi$  ist ein n-Tupel von Propositionen. Sei  $\xi = \langle z, R_1 \rangle$ , d.h. in  $\xi$  besteht die Präferenzrelation  $R_1$  und die übrigen Merkmale von  $\xi$  sind in  $z$  zusammengefaßt. Sei  $\xi' = \langle z, R_1' \rangle$ , dann gilt:  $R_1 \neq R_1' \Rightarrow \xi \neq \xi'$ .

Das o.g. Beispiel von Präferenzeninterdependenz (S. ) läßt sich damit ohne Verwendung der Implikation in folgender Weise formulieren:

- (1)  $\langle \langle p_1, p_1 \rangle_2 \langle p_2 \rangle, \langle p_2, p_1 \rangle_2 \langle p_2 \rangle \rangle \in R_1$
- (2)  $\langle \langle p_2, p_2 \rangle_1 \langle p_1 \rangle, \langle p_1, p_2 \rangle_2 \langle p_1 \rangle \rangle \in R_1$
- (3)  $\langle \langle p_1, p_1 \rangle_1 \langle p_2 \rangle, \langle p_2, p_1 \rangle_1 \langle p_2 \rangle \rangle \in R_2$
- (4)  $\langle \langle p_2, p_2 \rangle_1 \langle p_1 \rangle, \langle p_1, p_2 \rangle_1 \langle p_1 \rangle \rangle \in R_2$

Hier ist von Propositionen die Rede. Die tatsächliche Situation könnte jedoch mit dieser Redeweise nur sehr grob wiedergegeben werden, denn man kann davon ausgehen, daß nicht für beliebige  $\xi_1 \in P_1, \xi_2 \in P_2$  gilt

$$\xi_1 \succ_2 \xi_2 \Rightarrow \xi_1 \succ_1 \xi_2 \text{ etc..}$$

Präferenzeninterdependenz sollte daher zunächst mit Zustands- und nicht mit propositionalen Präferenzen erläutert werden.

Stellen wir uns daher vor, die Welt bestünde nur aus den zwei möglichen Zuständen 'beide am Strand' und 'beide im Museum'.

Wenn die Bedingung für eine kardinale Metrisierung der Präferenzrelationen (Marschakbedingungen) erfüllt sind,<sup>1</sup> können wir interpersonelle Präferenzeninterdependenz als Nutzeninterdependenz behandeln.

<sup>1</sup> Vgl. H. Bühlmann/H. Loeffel/E. Nievergelt, *Entscheidungs- und Spieltheorie*, Berlin/Heidelberg/New York 1975, S. 107-115

Allerdings beziehen sich die den bedingten Präferenzrelationen entsprechenden bedingten Nutzenfunktionen<sup>1</sup> nicht auf Elemente aus  $\Xi$ . Es ist zweckmäßig, sich die Elemente der Menge der möglichen Welten  $\Xi$  aufgespalten zu denken in einen Anteil, der die Nutzenwerte angibt, und in einen Anteil, der alle Informationen, bis auf die über Präferenzrelationen bzw. Nutzenwerte, enthält. Im folgenden kennzeichnen die üblichen Zustandssymbole  $z_1, z_2 \dots$  ausschließlich diesen zweiten Anteil.  $Z$  ist die Menge der Zustände in diesem Sinne.  $Z$  nennen wir auch die Menge der Anaxzustände ('Zustände ohne Bewertungen'). Zustände im weiteren Sinne, also u.a. unter Einschluß subjektiver Bewertungen nennen wir weiterhin  $\xi$ .  $\Xi$  ist die entsprechende Alternativenmenge. Die Zustandsmenge  $\Xi$  kann damit als eine Menge von geordneten Paaren aufgefaßt werden, an deren erster Stelle das Symbol für den jeweiligen Zustand aus  $Z$  steht und an zweiter Stelle ein  $n$ -Tupel,  $n := \#K$ , von Nutzenwerten dieses Zustandes.

$$\begin{aligned} \Xi = \{ & \langle z_1, \langle v_1(z_1), v_2(z_1), v_3(z_1), \dots, v_n(z_1) \rangle \rangle \\ & \langle z_2, \langle v_1(z_2), v_2(z_2), v_3(z_2), \dots, v_n(z_2) \rangle \rangle \\ & \langle z_3, \langle v_1(z_3), v_2(z_3), v_3(z_3), \dots, v_n(z_3) \rangle \rangle \\ & \dots \\ & \langle z_k, \langle v_1(z_k), v_2(z_k), v_3(z_k), \dots, v_n(z_k) \rangle \rangle \} \end{aligned}$$

$$Z = \{ z_1, z_2, \dots, z_k \}, \#Z=k$$

---

<sup>1</sup> Nicht zu verwechseln mit dem "bedingten Nutzen" z.B. bei H. Schneeweiß, *Probability and utility - Dual concepts in decision theory*, in: Menges (1974) 113-144, der die Analogie zum Begriff der bedingten Wahrscheinlichkeit den "bedingten Nutzen" einer Konsequenz  $c$  aufgrund von  $c_0$  als  $u(c|c_0) := u(c+c_0) - u(c_0)$  definiert. Im Fall von Güterausstattungen bietet sich dafür eine natürliche Interpretation an: Der Nutzen eines zusätzlichen Güterbündels  $c$  ist der Nutzen der ursprünglichen und der zusätzlichen Güterausstattung zusammengekommen abzüglich des Nutzens der ursprünglichen Güterausstattung. Auf der Menge der Konsequenzen allgemein aber scheint eine additive Verknüpfung "+" keine vergleichbare natürliche Interpretation zu haben. Vgl. P.C. Fishburn, *Additivity in utility theory with denumerable product sets*, *Economica* 34 (1966) 500-503.

In §2 war der Definitionsbereich der individuellen Bewertungsfunktionen  $v_1$  die Menge der möglichen Welten  $\Xi$ . Dies kann auch so bleiben: Es gibt unbedingte subjektive Bewertungsfunktionen  $v_1$  über  $\Xi$  und bedingte subjektive Bewertungsfunktionen  $v_1$  über  $Z$ .

Die Kardinalisierbarkeit der  $v_1$  setzt zwar entsprechende Präferenzen über Lotterien von  $Z$  voraus, aber da Lotterien von  $Z$  nicht selbst zu  $Z$  gehören sollen, kann  $Z$  finit bleiben.<sup>1</sup>

In unserer Zwei-Personen-Zwei-Zustände-Welt - die beiden Zustände sind jetzt Elemente von  $Z$ , nicht von  $\Xi$  - gibt es  $2 \cdot 2^2 = 8$  mögliche Zustände  $\xi$  von  $\Xi$ .

$$\begin{aligned} \Xi = \{ & \langle z_1, z_1 \succ_1 z_2, z_1 \succ_2 z_2 \rangle, \langle z_1, z_1 \succ_1 z_2, z_1 \prec_2 z_2 \rangle, \\ & \langle z_1, z_1 \prec_1 z_2, z_1 \succ_2 z_2 \rangle, \langle z_1, z_1 \prec_1 z_2, z_1 \prec_2 z_2 \rangle, \\ & \langle z_2, z_1 \succ_1 z_2, z_1 \succ_2 z_2 \rangle, \langle z_2, z_1 \succ_1 z_2, z_1 \prec_2 z_2 \rangle, \\ & \langle z_2, z_1 \prec_1 z_2, z_1 \succ_2 z_2 \rangle, \langle z_2, z_1 \prec_1 z_2, z_1 \prec_2 z_2 \rangle \} \end{aligned}$$

$$\begin{aligned} \Xi_1 = \{ & \langle z_1, z_1 \succ_2 z_2 \rangle, \langle z_1, z_1 \prec_2 z_2 \rangle, \\ & \langle z_2, z_1 \succ_2 z_2 \rangle, \langle z_2, z_1 \prec_2 z_2 \rangle \} \end{aligned}$$

$$\begin{aligned} \Xi_2 = \{ & \langle z_1, z_1 \succ_1 z_2 \rangle, \langle z_1, z_1 \prec_1 z_2 \rangle, \\ & \langle z_2, z_1 \succ_1 z_2 \rangle, \langle z_2, z_1 \prec_1 z_2 \rangle \} \end{aligned}$$

Die bedingten Präferenzrelationen  $v_1$  der Personen haben als Bezugsmenge die interpersonell gleiche Menge  $Z$ ,

<sup>1</sup> Die kardinale Metrisierbarkeit setzt voraus, daß es neben  $v_1$  über  $Z$  auch eine subjektive Bewertungsfunktion  $u_1$  über die Lotterien von  $Z$  mit bestimmten Eigenschaften gibt, und daß der Zusammenhang von  $v_1$  und  $u_1$  durch die Bayessche Regel bestimmt ist.



ihre unbedingten Präferenzrelationen  $v_1$  jedoch gehen über eine Menge möglicher Welten  $E_1$ , die sich von  $E$  geringfügig unterscheiden: die eigenen Präferenzen (bzw. Bewertungsfunktionen) sind in  $\xi \in E_1$  nicht enthalten. Auch wenn mehrere individuelle Bewertungsfunktionen angenommen werden, dürfen die eigenen Bewertungen der gleichen  $z$  und der gleichen Ordnung nicht in der Bezugsmenge enthalten sein.

Das bisher eingeführte entscheidungstheoretische Begriffssystem muß also vorläufig folgendermaßen erweitert werden:

- (1) Die Menge von Anaxzuständen  $Z$ .  $\#Z \in \mathbb{N}$ .
- (2) Jede Person hat eine (bedingte) Bewertungsfunktion  $v_1$  mit dem Definitionsbereich  $Z \times N^1$  sei die Menge der möglichen individuellen Bewertungsfunktionen  $v_1$  von  $Z$  in  $\mathbb{R}$ .
- (3) Die nicht bedingte Bewertungsfunktion  $v_1$  hat als Definitionsbereich  $E_1 = Z \times N_1$ , wobei  $N_1 = N^1 \times N^2 \times \dots \times N^{l-1} \times N^{l+1} \times \dots \times N^{\#K}$
- (4)  $E = Z \times \prod_{i=1}^{\#K} N_i$

Im Zwei-Personen-Fall ergibt sich

$$E_1 = Z \times N_2; E_2 = Z \times N_1; v_1: E_1 \ni \xi \rightarrow x \in \mathbb{R}; v_2: E_2 \ni \xi \rightarrow x \in \mathbb{R};$$

$$v_1: Z \ni z \rightarrow x \in \mathbb{R}; v_2: Z \ni z \rightarrow x \in \mathbb{R}.$$

Bevor wir die sozialetischen Folgerungen aus der Möglichkeit der Bewertungsinterdependenz ziehen, soll das Interdependenzproblem in idealisierter Form etwas genauer analysiert werden.

Wir befinden uns in einer Welt mit einem einzigen Anaxzustand  $z$  und unendlich vielen möglichen Zuständen  $\xi \in E$ ;  
 $E = Z \times N_1 \times N_2$ .

Nehmen wir einmal an, die individuellen Präferenzordnungen seien durch kardinale reellwertige normierte subjektive Bewertungsfunktionen repräsentiert.<sup>1</sup> Nehmen wir

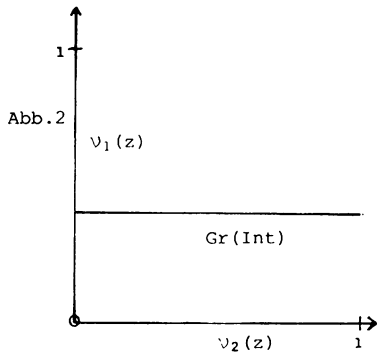
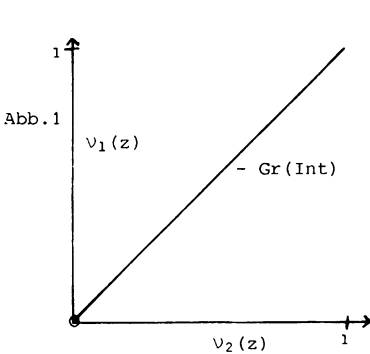
---

<sup>1</sup> Das summum bonum und das summum malum jeder Person habe jeweils die gleichen Nutzenwerte, z.B. 1 und 0.

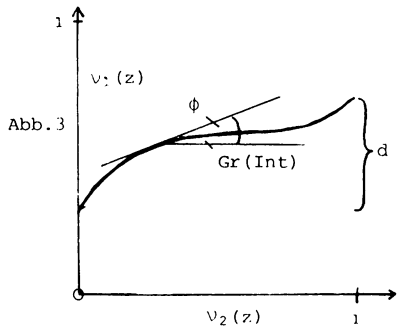
weiter an, daß die jeweilige Einschätzung der Bewertungen des anderen eine Rolle für die eigene Bewertung des Zustandes spielt.

Im folgenden sei vorausgesetzt, daß sich weder A noch B in ihrer Einschätzung des bedingten Nutzens für den anderen täuschen, so daß wir die bedingten Nutzenbewertungen  $v_1(z)$ ,  $i=1,2$  in einen direkten funktionalen Zusammenhang (Int) bringen können:  $\text{Int} \circ v_2 = v_1$ .

Können wir diese Voraussetzung nicht machen, ergibt sich die Nutzenbewertung der einen Person in Abhängigkeit von der Einschätzung der Nutzenbewertung der anderen Person. Entsprechend ist dann die Abszisse zu interpretieren.



In Abb. 1 ist die Nutzenbewertung von A völlig von der Nutzenbewertung von B abhängig. In Abb. 2 dagegen repräsentiert der Graph eine Nutzeninterdependenz, in der der Nutzen des Zustandes  $z$  für A vollständig unabhängig ist von  $v_2(z)$ . Abb. 3 zeigt eine Interdependenzkurve, in



der der Nutzen eines Zustands  $z$  für die Person A sich um  $d$  verändert, bei einer maximalen Veränderung des Nutzens von  $z$  für Person B. Den Betrag der ersten differentiellen Ableitung der Nutzeninterdependenzkurve nach  $v_2(z)$  können wir als Maß für die Stärke der Korrelation, die natürlich über den Definitionsbereich der Variablen nicht konstant zu sein braucht, definieren:

$$\text{korr}_{12}(v_2(z)) = \frac{dv_1(z)}{dv_2(z)} (v_2(z))$$

Als Maß der Gesamtkorrelation bietet sich  $d$  von Abb.3 an.

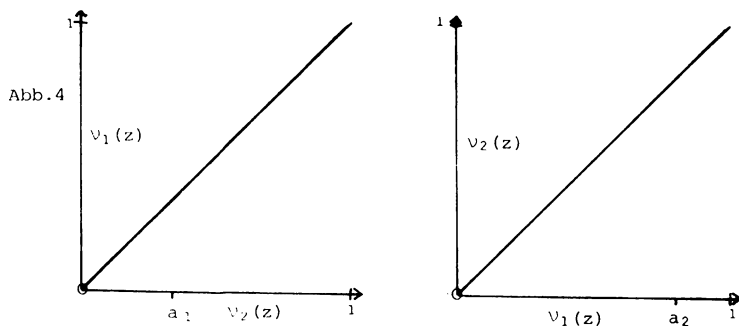
$$\text{Korr}_{12} = \max\{|\text{Int}(v_2(z)) - \text{Int}(v_2'(z))|\} \}_{v_2(z), v_2'(z) \in [0,1]}$$

Wechselseitige Nutzeninterdependenz ist in den unterschiedlichsten Formen denkbar.

(1) In einem besonders einfachen Fall haben beide Personen als Interdependenzfunktion die identische Abbildung. (Abb.4)

Nehmen wir an, Person A schätzt den Nutzen des Zustandes  $z$  für die Person B mit  $a_1$ , womit sich sein bedingter Nutzen zu  $v_1(z) = a_1$  ergibt. Umgekehrt schätzt Person B  $v_1(z)$  zu  $a_2$ , womit sich dessen bedingter Nutzen zu  $v_2(z) = a_2$  ergibt.

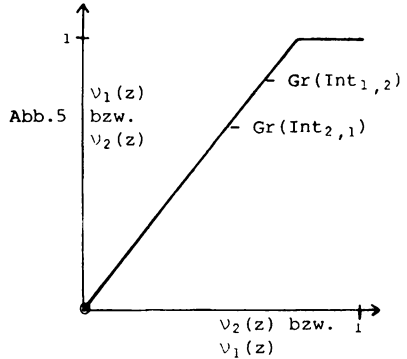
Für diese Interdependenz-situation führt jedoch auch eine Offenlegung der Nutzenbewertungen zu keinem befriedigenden Ergebnis: Die individuellen Nutzenbewertungen der beiden Personen springen bei jeder erneuten Offenlegung von  $a_1$  zu  $a_2$  bzw. von  $a_2$  zu  $a_1$ .



Die erste gegenseitige Information (Offenlegung) führt zu den Nutzenbewertungen  $v_1'(z)=a_2$  und  $v_2'(z)=a_1$ , die zweite Offenlegung zu  $v_1''(z)=a_1$  und  $v_2''(z)=a_2$ , die dritte wieder zu  $v_1'''(z)=a_2$  und  $v_2'''(z)=a_1$ .<sup>1</sup>

(2) Ein ähnlich einfacher Fall hat ganz andere Eigenschaften:

Der Graph der Abbildung 5 repräsentiert die gleiche Nutzeninterdependenz beider Personen. Eine solche Interdependenz hat eine eigenartige Konsequenz: Wie auch immer die Ausgangssituation geartet ist, eine endliche Anzahl von Offenlegungen führt schließlich zu einer maximalen Bewertung des Zustandes durch beide



Personen. Dazu ist es nicht nötig, daß es sich um gleiche individuelle Nutzeninterdependenzen handelt. Eine hinreichende Bedingung<sup>2</sup> für dieses Ergebnis wäre:

$$\exists \epsilon > 0 [\forall x \in [0, 1] \setminus \{x | f_1(x) = 1\} : f_1(f_2(f_1(x))) - f_1(x) \geq \epsilon \wedge \forall x \in [0, 1] \setminus \{x | f_2(x) = 1\} : f_2(f_1(f_2(x))) - f_2(x) \geq \epsilon];$$

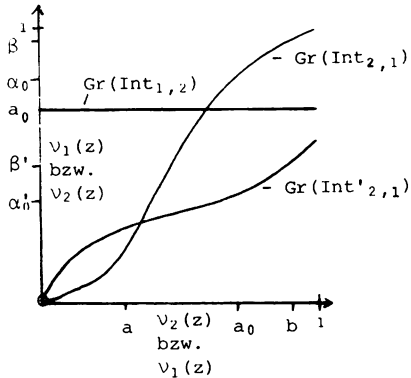
wobei  $f_1 = \text{Int}_{1,2}$  und  $f_2 = \text{Int}_{1,2}$ .

(3) Hat eine der beiden Personen eine konstante Interdependenzfunktion, so liegt - unabhängig davon, wie die

<sup>1</sup> Die Nutzenbewertung bei einer gegebenen Interdependenzstruktur nach einer Offenlegung kennzeichnen wir durch  $v_1'(z)$ , nach zwei Offenlegungen durch  $v_1''(z)$ , nach  $n$  Offenlegungen durch  $v_1^{(n)}(z)$ .

<sup>2</sup> Eine hinreichende und notwendige Bedingung zu formulieren, würde einen unangemessen großen mathematischen Aufwand erfordern.

Interdependenzfunktion der anderen Person beschaffen ist - jeweils nach der ersten Offenlegung eine stabile Nutzenbewertung vor: die Nutzenbewertungen beider Personen ändern sich bei erneuten Offenlegungen nicht mehr.



$$v_1(z) = a_0, \quad v_2(z) = \beta,$$

$$v_2'(z) = \beta', \quad \text{wobei}$$

$$v_2 = Int_{2f} \circ v_1 \quad \text{und}$$

$$v_2' = Int_{2f}' \circ v_1$$

$$v_1'(z) = v_1''(z) = \dots, \quad \text{unabhängig von } Int_{21} \text{ und } Int_{21}'$$

$$v_2'(z) = Int_{21}(v_1(z)) = Int_{21}(a_0) = \alpha_0$$

$$v_2''(z) = Int_{21}(v_1'(z)) = Int_{21}(a_0) = \alpha_0$$

....

$$v_2'''(z) = Int_{21}'(v_1(z)) = Int_{21}'(a_0) = \alpha_0'$$

$$v_2''''(z) = Int_{21}'(v_1'(z)) = Int_{21}'(a_0) = \alpha_0'$$

...

Wer diese bisher genannten Beispiele als harmlos für die normative Theorie der subjektiven Bewertungsaggregation empfunden hat, wird sich vielleicht von einem vierten Fall von Interdependenz überzeugen lassen, daß eine normative Aggregationstheorie ohne eine Klärung der Frage, auf welche Art von Präferenzen bzw. (subjektiven) Bewertungen sich die Theorie bezieht, nicht auskommt.<sup>1</sup>

Leider bringt dieser Fall ein bißchen Rechnerei mit sich, daher sei vorneweg das Ergebnis gesagt: Bestimmte Interdependenzstrukturen führen zu interpersonell und intrapersonell divergenten Bewertungen eines Zustandes.

<sup>1</sup> F. Schick hat sich schon Anfang der 70er Jahre mit ähnlichen

(4) Ist  $a_1$  die ursprüngliche (bedingte) subjektive Bewertung dieses Zustandes  $z$  seitens der Person A und  $a_2$  die ursprüngliche Bewertung dieses Zustandes seitens der Person B, so ergeben sich bei iterierten Offenlegungen der Bewertungen folgende Veränderungen.

Für die Person A:  $a_1, f_1(a_2), f_1(f_2(a_1)), f_1(f_2(f_1(a_2))), \dots$   
mit dem  $n$ -ten Glied  $v_1^{(n)}$ :

$f_1(f_2(f_1 \dots (f_1(a_2)) \dots))$ , für  $n \geq 2N$   
und  $\underbrace{\hspace{10em}}_{(n-2) \text{ Hintereinanderausführungen}}$

$f_1(f_2(f_1 \dots (f_2(a_1)) \dots))$ , für  $n \geq 2N+1$   
 $\underbrace{\hspace{10em}}_{(n-2) \text{ Hintereinanderausführungen}}$

Für die Person B:  $a_2, f_2(a_1), f_2(f_1(a_2)), f_2(f_1(f_2(a_1))), \dots$   
mit dem  $n$ -ten Glied  $v_2^{(n)}$ :

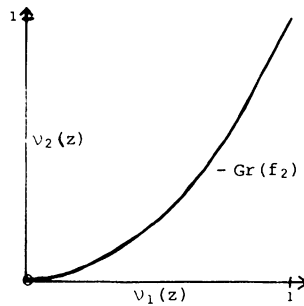
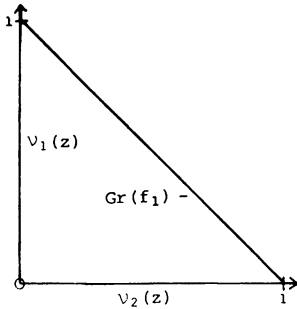
$f_2(f_1(f_2 \dots (f_2(a_1)) \dots))$ , für  $n \geq 2N$   
und  $\underbrace{\hspace{10em}}_{(n-2) \text{ Hintereinanderausführungen}}$

$f_2(f_1(f_2 \dots (f_1(a_2)) \dots))$ , für  $n \geq 2N+1$ ,

wobei  $f_1 = \text{Int}_{12}$  und  $f_2 = \text{Int}_{21}$ .

Nun sei  $\text{Int}_{12}(x) = f_1(x) = 1-x$  und  $\text{Int}_{21}(x) = f_2(x) = x^2$ .

Dann gilt:  $f_2(f_1(x)) = (1-x)^2 = 1-2x+x^2$  und  $f_1(f_2(x)) = 1-x^2$ .




---

Schlußfolgerungen für die Aggregationstheorien des social-choice-Ansatzes mit der interpersonellen Präferenzeninterdependenz befaßt. Schick beschreitet allerdings einen anderen Weg aus der Aporie als wir (vgl. §9). Er schlägt vor, bei der Aggregation nur solche individuellen Präferenzen zu berücksichtigen, die im Gleichgewicht sind. Das kann jedoch zu inadäquaten Ergebnissen führen (vgl. dazu den Anhang "Präferenzeninterdependenz"). Vgl. bes. Schick (1972), aber auch Schick (1971) und Schick (1978).

Es gilt:

$$(1) \quad x < f_2(f_1(x)) \wedge x \in [0, 1] \Leftrightarrow x \in [0, \frac{3-\sqrt{5}}{2}]$$

$$(2) \quad x > f_2(f_1(x)) \wedge x \in [0, 1] \Leftrightarrow x \in (\frac{3-\sqrt{5}}{2}, 1]$$

$$(3) \quad x < f_1(f_2(x)) \wedge x \in [0, 1] \Leftrightarrow x \in [0, \frac{\sqrt{5}-1}{2}]$$

$$(4) \quad x > f_1(f_2(x)) \wedge x \in [0, 1] \Leftrightarrow x \in (\frac{\sqrt{5}-1}{2}, 0]$$

Für  $a_1 > \frac{3-\sqrt{5}}{2}$  konvergiert die Teilfolge von  $v_1(n)$ ,  $n \in \mathbb{N}$  die aus dem 5., 9., 13., usw. Glied besteht, gegen 0, während die Teilfolge, die aus dem 3., 7., 11., 15., usw. Glied besteht, gegen 1 konvergiert. Die Bewertungen der Person A divergieren also allein aufgrund wechselseitiger Information und nehmen zunehmend extremere Werte an. Analoges gilt für die zweite Person. Diese intrapersonelle Divergenz bleibt für alle Kombinationen von Anfangsbewertungen  $a_1, a_2$  bestehen. Die Bewertungen streben nicht nur keinen Gleichgewichtspunkt an, sondern wechselseitige Information verschlimmert sozusagen die Situation: die betroffenen Personen schwanken in ihren Bewertungen immer heftiger, um am Ende den Zustand hintereinander, einmal als den erstrebenswertesten und dann wieder als den denkbar schlechtesten einzuschätzen.<sup>1</sup>

## ZWEI INTERDEPENDENZPROBLEME

Vor dem Hintergrund dieser vier Beispiele können wir uns die beiden Interdependenzprobleme normativer Aggregationstheorien klarmachen.

Normative Aggregationstheorien beziehen sich auf Präferenzen

---

<sup>1</sup> mathematische Darstellung im Anhang: "Zur Interdependenzproblematik"

bzw. auf subjektive Bewertungen. Es stellt sich die Frage, auf welche? Darf sich eine normative Aggregationstheorie auf interdependente Präferenzen beziehen?

Es ist wichtig zu sehen, daß es bei wechselseitiger Interdependenz der Präferenzen nicht sinnvoll ist, zu antworten auf die tatsächlichen. Denn die tatsächlichen Präferenzen sind abhängig von der interpersonellen Nutzenabschätzung - und damit hängen die 'tatsächlichen' Präferenzen vom Stand der (iterierten) Informationen ab.

Zwei Beispiele können das verdeutlichen:

- (a) Einige der Schwarzen in den USA ziehen Separation der Integration vor, weil sie der Überzeugung sind, daß die Weißen immer gegen Integration sein werden. Manche 'progressive' Weiße unterstützen diese Bewegung unter den Schwarzen. Eines Tages könnte es so weit gekommen sein, daß alle Schwarzen Separation vorziehen, weil sie wissen, daß die Weißen Separation wollen, und alle Weißen ebenfalls Separation vorziehen, weil sie den Eindruck haben, daß die Schwarzen selbst das wollen, und zugleich ist nicht ausgeschlossen, daß alle diese Separation beklagen.<sup>1</sup>

Pareto-Inklusivität der Aggregationsregeln würde also Separation für kollektiv besser als Integration bestimmen.

- (b) Ein altes Ehepaar teilt seine Frühstücksemmeln jeweils in eine obere und eine untere Hälfte. Sie ißt immer die oberen und er die unteren Hälften. Am Morgen ihres Goldenen Hochzeitstages hat die alte Dame einen besonderen Wunsch: "Heute hätte ich zur Feier des Tages ausnahmsweise einmal lieber die untere Hälfte. Denk' nicht, daß es mir während der vielen gemeinsamen Jahre schwergefallen wäre, sie an dich abzutreten - nein, im Gegenteil, das habe ich gern getan, weil ich wußte, daß du sie lieber magst - und ab morgen wollen wir es auch wieder so halten wie bisher." 'Tragik' dieser (vielleicht nicht ohne Grund gern er-

---

<sup>1</sup> Dieses Beispiel übernehme ich von Frederic Schick; vgl. Schick (1972), S.61.



zählten) Geschichte ist es, daß auch er 50 Jahre lang nur deshalb auf die obere Hälfte verzichtet hat, weil er die ganze Zeit geglaubt hat, daß seine Frau die obere Semmelhälfte bevorzugt.

Man könnte sagen, daß die beiden sich eben über die tatsächlichen Präferenzen der anderen geirrt haben. Aber haben sie das wirklich? Sie hatte ja tatsächlich eine Präferenz für die obere und er tatsächlich eine für die untere Hälfte, und insoweit Gründe und Motive nicht interessieren, kann man nicht leugnen:

- (I) Sie hatten beide korrekte Vermutungen über die Präferenzen des anderen.
- (II) Sie hatten tatsächlich - bis zu ihrem Goldenen Hochzeitstag - diese Präferenzen: er bevorzugte tatsächlich die obere und sie bevorzugte tatsächlich die untere Hälfte.
- (III) Ihre kollektive Präferenz war bezüglich ihrer tatsächlichen Präferenzen pareto-optimal.

Dieses Beispiel spricht weniger gegen das Pareto-Prinzip als dagegen, dieses Prinzip und andere Aggregationsregeln auf interdependente Präferenzen anzuwenden.

Da es jedoch als empirisch gegeben gelten kann, daß die tatsächlichen - später werden wir sagen die 'resultierenden' - Präferenzen interdependent sind, benötigen wir noch zumindest eine weitere präferenzielle Basis für subjektive Zustandsbewertungen, die es uns in dem letzten Beispiel erlauben würde zu sagen, die Dame hätte zwar während ihrer Ehe unverändert eine Präferenz (dieser Art) für die untere Hälfte gehabt, zusätzlich jedoch - in der irrigen Annahme, daß ihr Ehemann ebenfalls eine Präferenz<sup>o</sup> derselben Art<sup>1</sup> für die untere Hälfte hat - eine resultierende Präferenz für die obere Hälfte und vice versa.

---

<sup>1</sup> Um diese Präferenzen von den anderen zu unterscheiden, schreiben wir "Präferenz<sup>o</sup>".

Geändert hat sich bei dieser Interpretation eine ganze Menge:

1. In der ursprünglichen Interpretation hatten die beiden Ehepartner keine irrigen Annahmen. In der neuen Interpretation hatten sie über all die Jahre hinweg falsche Vermutungen über die Präferenzen<sup>o</sup> des Partners. Nur über die resultierenden Präferenzen des anderen waren sie richtig informiert.
2. Sobald jedoch die Präferenzen<sup>o</sup> aufgedeckt sind, ändern sich auch die resultierenden interdependenten Präferenzen. Die Interdependenz besteht jedoch nicht zwischen den Präferenzen der gleichen Art, sondern zwischen den resultierenden Präferenzen und den Präferenzen<sup>o</sup>. Die resultierenden Präferenzen von A sind abhängig von den Präferenzen<sup>o</sup> von B (und wohl auch von seinen eigenen Präferenzen<sup>o</sup>).
3. Aggregationstheorien sollten sich offensichtlich nicht auf resultierende sondern auf Präferenzen<sup>o</sup> beziehen: ja die individuellen resultierenden Präferenzen der Betroffenen selbst richten sich in diesem Beispiel nach dem Pareto-Kriterium, angewendet auf die Präferenzen<sup>o</sup>.<sup>1</sup>

Die gleichen Veränderungen ergeben sich für die Interpretation des ersten Beispiels; allerdings kommt dort eine Komplikation durch den statistischen Charakter der wechselseitigen Einschätzungen hinzu.

Wir wollen es hier mit diesen Bemerkungen zum ersten Interdependenzproblem vorerst bewenden lassen. Soviel sollte schon jetzt klar geworden sein: Es ist nicht sinnvoll, (resultierende) interdependente individuelle Präferenzen (bzw. die entsprechenden subjektiven Bewertungen) zum Gegenstand der sozialetischen Aggregation zu machen. Erstens deshalb, weil Interdependenz eine gewisse

---

<sup>1</sup> Wir werden diese Eigenschaft individueller resultierender Präferenzen später als eine spezielle Rationalitätsbedingung (PIRPIW) einführen, vgl. §17.

Willkürlichkeit der aggregierten Präferenzen (bzw. Bewertungen) nach sich ziehen kann (wie die vier 'mathematischen' Beispiele deutlich gemacht haben), und zum zweiten, weil sich die resultierenden interdependenten Präferenzen als Grundlage für die Bestimmung des kollektiv Besseren oder (je nach Interpretation auch individuell) moralisch Gebotenen nicht so recht eignen: Schon die Anwendung des Pareto-Kriteriums in einem besonders einfachen Fall hat sich als offensichtlich inadäquat herausgestellt.

Das zweite Interdependenzproblem ist bei genauer Betrachtung eine Spezialisierung des ersten. Denn subjektivistische sozialetische Theorien allgemein und normative Aggregationstheorien im besonderen können selbst als Theorien normativ gebotener Präferenzeninterdependenz interpretiert werden.

Wird eine Aggregationstheorie praktisch relevant, führt das zu einer speziellen Art von Präferenzeninterdependenz, die von den Aporien, wie sie sich in den o.g. Beispielen zeigen, nicht verschont bleiben.

Zur Konkretisierung genügt ein Beispiel. Wir wählen eine moderne Form des Utilitarismus, den ethischen Bayesianismus. Der ethische Bayesianismus, wie er von Harsanyi entwickelt wurde,<sup>1</sup> ist im Gegensatz zu fast allen anderen Formen des Utilitarismus eine subjektivistische sozialetische Theorie.

---

<sup>1</sup> Vgl. bes. J.C. Harsanyi, *Cardinal welfare, individualistic ethics, and interpersonal comparison*, *Journal of political economics* 63 (1955) 309-321, abgedruckt in: Harsanyi, *Essays ...* (1976). Vgl. a. Harsanyi 1958, 1975c, 1977a u. b, 1978a, 1979b, 1980. Stegmüller schlägt für diese Form des Utilitarismus die Bezeichnung "ethischer Bayesianismus" oder "unpersönlicher Bayesianismus" vor. s. W. Stegmüller, *On the interrelations between ethics and other fields of philosophy and science*, *Erkenntnis* 11 (1977) 55-80.

Grundlage der Aggregation sind die tatsächlichen Präferenzen der Personen. Durch eine spezielle Methode impliziter Bewertungsvergleiche, die hier nicht weiter interessiert, werden die moralisch gebotenen Präferenzen ermittelt.

Für jedes Individuum ergibt sich aufgrund seiner Präferenzrelation  $R_1$  über  $Z$  (und Marschak-konsistenter Präferenzen über die Menge der Lotterien von  $Z$ ) eine kardinale Nutzenfunktion  $v_1$ .

Der ethische Bayesianismus legt die je individuell moralisch gebotenen Präferenzen aufgrund dieser subjektiven kardinalen Bewertungsfunktionen (die hier interpersonell vollständig vergleichbar sind) durch folgende normative Interdependenzempfehlung (\*) fest:

$$\langle z_1, z_2 \rangle \in R_1 \Leftrightarrow \sum_{\kappa=1}^{\#K} v_{\kappa}(z_1) > \sum_{\kappa=1}^{\#K} v_{\kappa}(z_2).$$

Je nach anfänglicher wechselseitiger Einschätzung der subjektiven Bewertungen ergeben sich damit in einer Gesellschaft von Utilitaristen ohne  $\alpha\kappa\rho\alpha\sigma\iota\alpha$ <sup>1</sup> entsprechende

Präferenzen  $R_1$ , die zwar nicht objektiv, aber je nach subjektiver Einschätzung (\*) erfüllen.

Die individuellen subjektiven Bewertungsfolgen, die sich bei dieser Art von Präferenzeninterdependenz ergeben, haben zwar einen interpersonell gleichen Grenzwert, aber dieser Grenzwert hängt ausschließlich von den anfänglichen Einschätzungen ab - was sie als moralisch gebotene Bewertung inadäquat macht.

---

<sup>1</sup>  $\alpha\kappa\rho\alpha\sigma\iota\alpha$  hier eher im Gegensatz zu  $\sigma\omega\phi\rho\sigma\acute{\upsilon}\nu\eta$  als zu  $\acute{\epsilon}\gamma\rho\acute{\alpha}\tau\epsilon\iota\alpha$ .  $\alpha\kappa\rho\alpha\sigma\iota\alpha$  liegt dann vor, wenn jemand davon überzeugt ist, für Handlung  $\sigma$  sprächen bessere Gründe als für Handlung  $\sigma'$  und sich dennoch - nicht aus Versehen oder mangelnder Information - wohlüberlegt und bewußt für  $\sigma'$  entscheidet. Einige Überlegungen dazu bei D. Davidson, *How is weakness of the will possible*, in Feinberg (1970) und bei D. Walsh, *'Akrasia' reconsidered*, *Ethics* 85 (1974/75) 151-158. Vgl.a. die Diskussion im 7. Buch der Nikomachischen Ethik. Dazu A. Rorty, *Akrasia and pleasure*, in: Oksenberg (1980).

Die je individuell moralisch gebotene Bewertungsfunktion ist bestimmt durch die Bewertungsfunktionen aller (einschließlich der eigenen). Insofern spielt die individuelle Bewertungsfunktion eine doppelte Rolle: einmal als Bestandteil der (subjektivistischen) außermoralischen empirischen Grundlage für die Bestimmung des moralisch Gebotenen, und zum zweiten als Gegenstand moralischer Beurteilung.

Für einen hedonistischen Utilitarismus besteht die außermoralische Grundlage der moralischen Präferenzen in der - so könnte man es nennen - Luststruktur der Gesellschaft: Ist von jedem Zustand  $z$  und  $z'$  aus  $Z$  bekannt, welche Lustveränderung für jede einzelne Person der Übergang von  $z$  zu  $z'$  bedeuten würde, so kann bestimmt werden, welche Präferenzen die Personen über  $Z$  haben sollten.<sup>1</sup> Die Relation ' $z$  bringt  $i$  mehr Lust als  $z'$ ' können wir uns ebenfalls als kardinalisierbar zu vollständig vergleichbaren, normierten numerischen Bewertungsfunktionen  $w_i$  vorstellen, so daß sich die moralisch gebotenen Präferenzen in der für alle gleichen Bewertungsfunktion (\*)

$$m(z) = \frac{1}{\#K} \sum_{i=1}^{\#K} w_i(z) \text{ zusammenfassen lassen.}^2$$

---

<sup>1</sup> Man könnte mit Recht einwenden, utilitaristischen Theorien ginge es um Handlungen oder in der regel-utilitaristischen Version um Handlungsregeln, nicht um Zustandspräferenzen. Harsanyis Utilitarismus-Version sucht den Anschluß zur rationalen Entscheidungstheorie und will daher der ethischen Theorie die Bestimmung der (moralischen) Bewertungsfunktionen überlassen, während es den Kriterien rationalen Verhaltens überlassen bleibt, danach die moralisch gebotene Handlung zu bestimmen. Dieser Rekonstruktionsansatz läßt sich auch auf andere Formen des Utilitarismus übertragen. Vgl. etwa Mill (1863), Lyons (1965), Smart (1961 u. 1973), Bohnen (1964), Braybrooke (1967). Einige wichtige Texte des Utilitarismus in: Höffe (1975b).

<sup>2</sup> Seit Moore (1903) wird i.a. von Utilitaristen nicht der Anspruch erhoben, hiermit eine Definition oder Begriffsexplikation vorzunehmen, sondern (\*) ist als Theorie aufzufassen; d.h. es ist eine Frage der (normativen) Richtigkeit und nicht des tatsächlichen oder des empfehlenswerten *Sprachgebrauchs*, ob (\*) gültig ist oder nicht.

Ähnlich ließe sich die eudämonistische Form des Utilitarismus rekonstruieren. Etwas schwieriger wäre es vermutlich, die Mooresche Version eines 'idealen' Utilitarismus in diese Form zu bringen.<sup>1</sup> Aber eines ist den nicht-subjektivistischen Versionen des Utilitarismus gemeinsam: Sie kennen, im Gegensatz zum ethischen Bayesianismus eine klare Trennung der außermoralischen Basis und der moralischen Beurteilung selbst.<sup>2</sup>

Die Konsequenz ist, daß der ethische Bayesianismus für den Fall vollständiger Information aller und völligen Mangels an ἀκρᾶσια-Problemen, also für den Idealfall, keine Handlungsempfehlungen mehr gibt, oder anders ausgedrückt, keine ethische Präferenzen bestimmen kann. Gerade dieser Idealfall aber müßte es erlauben, frei von allen 'pragmatischen' Beschränkungen aufgrund mangelnden Wissens und Charakterschwäche den Gehalt einer normativen Theorie festzustellen. Die Präferenzeninterdependenz des ethischen Bayesianismus führt also zu einer ähnlichen Ratlosigkeit wie sie unser Liebespaar am Anfang ihrer Reise in Neapel erlebt hat.

Es gilt allgemein, daß moralische Überzeugungen die individuellen Präferenzordnungen beeinflussen - unabhängig davon, auf welche ethische Theorie sich diese Überzeugungen berufen - und damit erheben die resultierenden individuellen Präferenzen ihrerseits u.U. den Anspruch normativer Richtigkeit. Auf normativen Urteilen gründende oder von ihnen beeinflusste Präferenzen können daher nicht

---

<sup>1</sup> Zur Theorie des idealen Utilitarismus vgl. Moore (1903), *Principia ethica*; zur Theorie des intrinsisch Wertvollen bes. Kap.6; in der dt. Ausgabe (1970) bes. S.254-283.

<sup>2</sup> Harsanyi spricht allerdings von "subjektiven Präferenzen" und von "ethischen Präferenzen", so daß der Eindruck entsteht, die vermißte Differenzierung sei in Wirklichkeit sehr wohl erfolgt. Daß dieser Eindruck irreführend ist, läßt sich besser im Kontext des nachfolgenden Abschnitts zeigen, so daß wir diese Frage hier zurückstellen.

(außermoralische) Grundlage eines normativen Prinzips der Interessenaggregation sein. Solche Präferenzen repräsentieren - und sei es nur partiell - normative Urteile, die (im Gegensatz zu Interessen) objektive Gültigkeit beanspruchen. Auf diese Weise kommen kollektive Präferenzen, die auf normativen Regeln der Präferenzenaggregation beruhen, mit individuellen moralischen Überzeugungen in Konflikt:

Die beiden Ebenen

- Eigeninteresse versus moralische Verpflichtung

und

- individuelle moralische Überzeugung versus kollektive moralische Überzeugung

können von subjektivistischen Aggregationstheorien mit einem uniformen Präferenzbegriff nicht auseinandergelassen werden.

Das ist schon während dieses Abschnittes angeklungen. Da dieses Defizit jedoch die zweite Säule ist, auf die sich unser Vorschlag einer Aufspaltung des Präferenzbegriffes stützt, müssen wir uns damit eingehender befassen.

## 8. PRAKTISCHER VERSUS THEORETISCHER KONSENSUS: INTERESSEN- UND URTEILSAGGREGATION

Normative teleologische Theorien der Sozialethik, speziell Theorien des fairen Interessenausgleichs und der distributiven Gerechtigkeit bestimmen (zumindest implizit) ethisch gebotene Präferenzen. Wessen Präferenzen gemeint sind, ist je nach Theorie verschieden und bleibt in vielen Fällen offen.

Ökonomische Wohlfahrtstheorien wenden sich traditionell eher an planerisch tätige Institutionen, Bürokratie und Politiker. Die Institutionen selbst sind nicht Gegenstand der Interessenaggregation; insofern wird bei dieser Interpretation wohlfahrtstheoretischer Aggregationstheorien eine zirkuläre oder sogar kontradiktorische Präferenzenbestimmtheit vermieden. Ob diese Idealisierung der Institutionen als interessenfrei und gemeinwohlorientiert angemessen ist, ist dabei nicht wesentlich.

Die modernen Nachfahren des wohlfahrtstheoretischen Ansatzes in der Sozialwahltheorie sind jedoch i.a. nicht mehr als bloße Handlungsanleitung öffentlicher Entscheidungsträger gedacht, sondern beanspruchen, allgemeine moralische Prinzipien, Gerechtigkeits- und Fairneßkriterien zu formulieren. Für sie stellt sich daher dringlicher als für traditionelle Wohlfahrtstheorien die Frage, in welcher Weise die persönlichen Interessen repräsentiert werden, so daß eine zirkuläre oder kontradiktorische Präferenzenbestimmtheit ausgeschlossen ist.

Die notwendige Begriffs differenzierung läßt sich mit einem Satz so begründen: Normative Aggregationstheorien der Sozialethik dienen einem praktischen, nicht einem theoretischen Konsensus.



Theorien normativer Interessenaggregation stellen - wenn sie allgemein akzeptiert sind - einen praktischen Konsensus her:

Die Personen sind sich bei jeder Interessenkonstellation einig, was als gerechter Ausgleich individueller Ansprüche gelten kann, und was in diesem Sinn 'gemeinsames Interesse' ist. Die Interessenkonstellation ist jeweils empirisch gegeben und die normative Aggregationstheorie begründet ein gemeinsames Urteil über das, was damit jeweils moralisch geboten ist. Um eine übereinstimmende Beurteilung der jeweiligen Interessenkonstellation zu erhalten, ist interpersonelle Vergleichbarkeit nötig - das erforderliche Ausmaß ist theorienabhängig.

Gibt es jedoch unterschiedliche Ansichten darüber, welche Aggregationstheorie angemessen ist, welche Kriterien eines fairen Interessenausgleichs anzuwenden sind etc., und gibt es dementsprechend auch bei übereinstimmender empirischer Analyse der Interessenkonstellation interpersonell unterschiedliche ethische Präferenzen, so kann es nicht Aufgabe einer normativen Theorie der Aggregation individueller Interessen - 'subjektiver' Präferenzen, 'subjektiver' Bewertungen<sup>1</sup> - sein, einen theoretischen Konsensus herzustellen.

Die beiden Fragestellungen: "Was ist moralisch richtig?" und "Was ist zu tun, wenn verschiedene Personen unterschiedliche Ansichten darüber haben, was moralisch richtig ist?" erfordern eine getrennte Behandlung.<sup>2</sup>

---

<sup>1</sup> Die übliche Gegenüberstellung 'subjektiver' Präferenzen und 'ethischer' Präferenzen in zahlreichen Beiträgen der Sozialwahltheorie ist nicht besonders glücklich, weil es auch *subjektive moralische* Präferenzen gibt. Im nächsten Abschnitt (§9) sprechen wir stattdessen von individuellen (subjektiven) Wohlfahrtspräferenzen und (subjektiven) moralischen Präferenzen - subjektive moralische Präferenzen können genauso irren wie subjektive Wohlfahrtspräferenzen.

<sup>2</sup> Eine 'pluralistische' Ethik akzeptiert normative Prinzipien auch zur Aggregation individueller moralischer Überzeugungen. Jedoch haben die Handlungsempfehlungen einer solchen Ethik die seltsame Eigenschaft, keinen Anspruch normativer Gültigkeit zu erheben, sondern nur relativ zu individuellen Moralurteilen akzeptiert zu sein, die von den Personen wechselseitig für falsch gehalten

Über Richtigkeit oder Unrichtigkeit einer Theorie - sei es eine normative oder eine empirische - entscheiden nicht kollektive Präferenzen, sondern die Qualität ihrer Rechtfertigung. Wenn unterschiedliche Meinungen zu dieser Frage bestehen, kann u.U. abgestimmt werden. Dieses Verfahren stellt aber keine Begründungsmethode dar: Die Richtigkeit einer Theorie kann nicht unter Berufung auf ein Abstimmungsergebnis aufgezeigt werden.<sup>1</sup>

Es ist jedoch nicht anzunehmen, daß die individuellen resultierenden Präferenzen die persönlichen Interessen der betreffenden Person repräsentieren. Ebenso können sie ausschließlich oder überwiegend moralisch motiviert sein. Die beiden Sätze : (1) "Ich halte die gesellschaftliche Situation  $s$  für (moralisch) besser als  $s'$  " und (2) "Mir geht es in  $s$  besser als in  $s'$  " bringen völlig unterschiedliche Sachverhalte zum Ausdruck.<sup>2</sup> Sowohl aufgrund von Sätzen der ersten Art, wie aufgrund von Sätzen der zweiten Art, ergeben sich individuelle Präferenzordnungen  $R_i$ . Die 'tatsächlichen' resultierenden (overall-) Präferenzen erlauben jedoch i.a. nicht, entweder im Sinne von (1) oder von (2) interpretiert zu werden.<sup>3</sup>

Daher konfundiert die normative Sozialwahltheorie

---

werden. In jedem Fall sind aber normative Regeln zur Lösung sozialer Interessenkonflikte streng zu unterscheiden von Verhaltensregeln im Fall widersprüchlicher normativer Urteile.

<sup>1</sup> Das gilt für normative ebenso wie für empirische Theorien: Ob sich die Bewegung von Körpern im Schwerfeld der Erde mit dem Newtonschen Gravitationsgesetz erklären läßt, kann nur durch Beobachtung und Experiment geprüft werden, nicht durch Umfragen.

<sup>2</sup> Andere Formulierungen wären:

- |  |  |
|--|--|
| (1) "Ich halte $s$ für besser als $s'$ ",<br>"s ist besser als $s'$ " etc. | (2) "Ich persönlich bevorzuge $s$ gegenüber $s'$ ",<br>"Meinen Interessen entspricht $s$ besser als $s'$ "<br>etc. |
|--|--|

Ganz eindeutig läßt sich der Unterschied alltagssprachlich kaum festlegen.

<sup>3</sup> Entscheidungstheoretiker legen sogar einen besonderen Wert darauf, daß individuelle Präferenzen bzw. die entsprechenden Nutzenfunktionen *beide* Arten von Urteilen, also moralische Überzeugungen und persönliche Interessen zum Ausdruck bringen.

notwendigerweise diese beiden Ebenen der Aggregationsproblematik, wenn sie sich auf die Aggregation resultierender individueller Präferenzen bzw. 'abstrakter' Nutzenfunktionen beschränkt.<sup>1</sup>

Nun äußert sich in der ökonomischen und speziell der sozialwahltheoretischen Literatur an manchen Stellen durchaus ein Unbehagen an einer mangelnden Differenzierung der Aggregationsproblematik. Sen spricht sogar ausdrücklich davon, daß es unterschiedliche Typen der Aggregation gebe, je nachdem

"wether the exercise is one of aggregating the conflicting interests of different people, or one of aggregating the conflicting judgements of different people as to what should be done."<sup>2</sup>

Jedoch ist auffallend, daß das bisher keine entsprechenden systematischen Konsequenzen gehabt hat.

#### PERSÖNLICHE UND MORALISCHE PRÄFERENZEN IM ETHISCHEN BAYESIANISMUS

Der schon im vorausgegangenen Abschnitt erwähnte ethische Bayesianismus hat den Anspruch, die Beziehung 'ethischer' und 'subjektiver' Präferenzen geklärt zu haben ohne vom uniformen Präferenzbegriff der modernen Nutzentheorie abgehen zu müssen. Dieser Anspruch ist daher erst noch zu prüfen, bevor wir im nächsten Abschnitt eine Differenzierung des Präferenzbegriffes vorschlagen.

---

<sup>1</sup> Auf Regeln der 'Urteilsaggregation' und Regeln der 'Interessenaggregation' sind unterschiedliche Kriterien anzuwenden. Vgl. dazu den Anhang "Die Aggregation individueller Urteile zu einem Gruppenurteil" S.

<sup>2</sup> A.K. Sen, *Personal utilities and public judgements: Or what is wrong with welfare economics?*, The economic journal 89 (1979), S.544. Vgl. a. J. Graaf, *Equity and efficiency as components of the general welfare*, South African journal of economics 45 (1977) 362-375 Vgl. a. A.K. Sen, *Behavior and the concept of preference*, *Economica* 40 (1973),

Nach Harsanyi hat jedes Individuum zwei Klassen von Präferenzen: persönliche und moralische. Die persönlichen Präferenzen beruhen

"mainly on his own personal interests and on the interests of his closest associates"<sup>1</sup>

und werden definiert als die jeweiligen tatsächlichen ("actual") individuellen Präferenzen. Die moralischen Präferenzen bzw. (bedeutungsgleich verwendet) die moralischen Werturteile werden definiert als die (hypothetischen) Präferenzen, die die einzelnen Personen in einer Situation hätten, in der sie nicht wüßten, wer sie sind, sondern sich vorstellen, mit gleicher Wahrscheinlichkeit jeden beliebigen Platz in der Gesellschaft einzunehmen.

Unter den bekannten Konsistenzvoraussetzungen lassen sich daher die individuellen persönlichen Präferenzen in der entsprechenden abstrakten Nutzenfunktion repräsentieren und die individuellen moralischen Präferenzen in der entsprechenden (individuellen) Wohlfahrtsfunktion. Da die Personen in der hypothetischen Entscheidungssituation zur Bestimmung ihrer moralischen Präferenzen mit der Wahrscheinlichkeit  $\frac{1}{n}$ ,  $n = \#K$ , nicht als jeweilige Person, sondern sozusagen 'mit Haut und Haar' in die Position jedes beliebigen anderen Mitgliedes der Gesellschaft versetzt werden und auch dessen persönliche Präferenzen, Charaktermerkmale, Eigenschaften übernehmen,<sup>2</sup> diese hypothetische Entscheidungssituation also interpersonell völlig symmetrisch ist, sind die individuellen moralischen Präferenzen und damit die individuellen

---

<sup>1</sup> J.C. Harsanyi, *Essays on ethics, social behavior and scientific explanation*, Dordrecht 1976, S.IX.

<sup>2</sup> Es ist fraglich, ob dies eine adäquate Explikation des Kriteriums der Verallgemeinerbarkeit ethischer Urteile ist. Vgl. J.L. Mackie, *Ethik*, Stuttgart 1981, Kap.4, der drei Stufen der Universalisierung unterscheidet, deren radikalste Harsanyis Kriterium der Universalisierbarkeit entspricht.

Wohlfahrtsfunktionen identisch. Diese Wohlfahrtsfunktion ist die Funktion des Durchschnittsnutzens.<sup>1</sup> Rationales moralisches Handeln maximiert damit den Erwartungswert des durchschnittlichen persönlichen Nutzens.<sup>2</sup>

Diese Theorie macht eine empirische Annahme, wobei allerdings nicht ganz klar ist, an welcher Stelle: Harsanyi 'definiert' die persönlichen Präferenzen als die tatsächlichen individuellen Präferenzen. Sollte "definiert" im strengen Wortsinn gemeint sein (d.h. "persönliche Präferenzen" und "tatsächliche individuelle Präferenzen" sind substituierbar), stellt sich die Frage, wozu dann der Begriff "persönliche Präferenzen" dienen soll. Ist die Gleichsetzung persönlicher Präferenzen mit tatsächlichen individuellen Präferenzen keine Definition, dann ist es eine empirische Hypothese. Um diese zu prüfen, müßten wir allerdings erfahren, wie wir die persönlichen Präferenzen feststellen können. Die Antwort, diese erhalten wir aufgrund des individuellen Wahlverhaltens, wäre natürlich eine *petitio principii*.

Nun meint Harsanyi, die persönlichen Präferenzen beruhen hauptsächlich auf den persönlichen Interessen und auf den Interessen derjenigen, die einem besonders nahestehen. Wir können also - was immer die Wortbedeutung von "persönlicher Präferenz" ist - die empirische Hypothese destillieren: Die (tatsächlichen) individuellen Präferenzen beruhen hauptsächlich auf den persönlichen Interessen und auf den Interessen derjenigen Personen,

---

<sup>1</sup> Dennoch ist Harsanyis Theorie kein Utilitarismus in der üblichen Bedeutung, vgl. J.C. Harsanyi, *Ethics in terms of hypothetical imperatives*, *Mind* 67 (1958) 305-316.

<sup>2</sup> Diese Theorie ist dargestellt in J.C. Harsanyi, *Cardinal welfare, individualistic ethics, and interpersonal comparison of utility*, *Journal of political economy* 62 (1955) 309-321. Vgl. außerdem Harsanyi (1977 a und b, 1978 a, 1979). Zur Kritik vgl. Diamond (1967).

die einem besonders nahestehen. Vieles deutet darauf hin, daß diese Hypothese empirisch falsch ist. Harsanyi hat selbst in einem anderen Zusammenhang in einer Kontroverse mit Sneed sehr überzeugende Argumente gegen die These der ausschließlich interessenbestimmten Motivation menschlicher Handlungen ins Feld geführt,<sup>1</sup> ohne aber auf die Implikationen für seine eigene Theorie<sup>2</sup> einzugehen.

"More importantly, the motivational assumptions of Sneed's model are much too crude and simple-minded. No doubt, self-interest (or more exactly, self-interest combined with a concern for the interests of one's family, friends and other associates) is a very important human motivation. But so are various altruistic and moral considerations."<sup>3</sup>

Als Beispiel wird die Veränderung der Haltung der weißen Mehrheit in den USA gegenüber der schwarzen Minderheit genannt,

"this was largely a result of changes in the moral attitudes of the white majority, rather than a result of changes in the people's self-interest".<sup>4</sup>

Wenn dem so ist, wenn die tatsächlichen individuellen Präferenzen in einem hohen Grad von den jeweiligen moralischen Überzeugungen abhängen, dann lassen sich die richtigen moralischen Präferenzen nicht durch eine gleichgewichtige Aggregation der tatsächlichen individuellen Präferenzen feststellen, denn das hieße ja, von neuem Urteilsaggregation und Interessenaggregation zu konfundieren.

Der ethische Bayesianismus steht vor dem Dilemma: Entweder sind die empirischen Annahmen richtig, und die

---

<sup>1</sup> Vgl. J.C. Harsanyi, *Sneed on Rawls' theory of social institutions: some comments*, Erkenntnis 13 (1978) 225-230.

<sup>2</sup> Die Harsanyi u.a. gegen Rawls verteidigt hat, vgl. J.C. Harsanyi, *Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory*, American political science review 59 (1975) 594-606.

<sup>3</sup> Harsanyi (1978), S.226.

<sup>4</sup> Harsanyi (1978), S.266.

tatsächlichen individuellen Präferenzen sind im wesentlichen Ausdruck persönlicher Interessen, dann spielen die Handlungsempfehlungen der aufgrund des ethischen Bayesianismus bestimmten individuellen Wohlfahrtsfunktionen nur eine theoretische Rolle (d.i. werden nach Voraussetzung nicht befolgt), oder die tatsächlichen individuellen Präferenzen können auch Ausdruck moralischer Überzeugungen sein, dann stellt der ethische Bayesianismus - wie wir an anderer Stelle gezeigt haben<sup>1</sup> - eine zirkuläre normative Theorie dar.

Die Differenzierung individueller Präferenzen in persönliche und ethische, die Einführung individueller Nutzenfunktionen und individueller Wohlfahrtsfunktionen, in der Art, wie sie die moderne Utilitarismusversion von Harsanyi vornimmt, ist daher keine Lösung des Problems.

---

<sup>1</sup> Vgl. §7

## 9. ZUR DIFFERENZIERUNG DES PRÄFERENZBEGRIFFS

*"A more serious objection to hedonism is simply that it is unrealistic as an account of responsible adult behavior. For instance my self-imposed activity, thinking and writing, is no fun. Very well: responsible adult behavior is an obscure and complicated matter"*

Quine (1973), S.32

Die beiden Abschnitte, die sich mit der Interdependenzproblematik (§7) und der Unterscheidung zweier Aggregationsebenen (§8) befaßt haben, sollten die - wie mir scheint zwingenden - Gründe für eine Aufspaltung des Präferenzbegriffes im Kontext subjektivistischer Sozialethik deutlich gemacht haben. In diesem Abschnitt schlagen wir vor, drei Grundtypen (subjektiver) Präferenzen zu unterscheiden. Im Anschluß daran werden einige in der sozialetischen Literatur vorkommende Charakterisierungen von Präferenzen, soweit sie für die Zwecke dieser Untersuchung dienlich erscheinen, diskutiert.

Die Frage, ob diese Unterscheidungen in der vorgeschlagenen Form sinnvoll sind, wird in diesem Abschnitt nur am Rande gestreift. Die Hoffnung ist, daß sie sich durch die Anwendung der Begriffe im zweiten Teil der Arbeit (vgl. bes. §16 und 17) positiv beantwortet.

Normativer Ethik geht es zu einem guten Teil um eine Theorie moralisch gebotener Präferenzen. Welche Präferenzen moralisch geboten sind, scheint in einem gewissen Ausmaß - in welchem haben wir nicht zu bestimmen versucht - von den Interessen der direkt oder indirekt betroffenen Personen abzuhängen. Diese Interessen lassen sich nicht an den jeweils vorliegenden individuellen Präferenzen ablesen, ebenso wenig natürlich an der sub-



jektiven Bewertung in Form einer Bernoulli-Nutzenfunktion, auch nicht, wenn diese durch zusätzliche Annahmen interpersonell vergleichbar gemacht wurde.

Welche Rolle das Mitgefühl, die "vicarious affects" für das moralische Verhalten haben, ist eine umstrittene Frage der Moralphilosophie. Für christliche Theologen, für Schopenhauer<sup>1</sup>, Spinoza<sup>2</sup>, Adam Smith<sup>3</sup>, oder Rescher<sup>4</sup> ist Moral ohne Mitfühlen, Nächstenliebe etc. nicht denkbar; für sie steht dementsprechend die Theorie des moralischen Gefühls im Zentrum der Ethik. Für Kant andererseits hat eine Handlung, die aus einem Gefühl des Mitleidens o.ä. hervorgeht, "so liebenswert sie auch ist, dennoch keinen wahren sittliche Wert"<sup>5</sup>.

Wir können hier keine Theorie der moralischen Gefühle und keine Theorie der persönlichen Interessen und ihrer Meßbarkeit entwickeln, wir werden diese Kontroverse daher nicht zu entscheiden versuchen. Was wir jedoch annehmen wollen, ist dreierlei:

- (1) Es ist weder analytisch noch empirisch wahr, daß das Verhalten jeder Person in jeder Situation ausschließlich den persönlichen Interessen dient.
- (2) Es ist weder analytisch noch empirisch wahr, daß die Intentionen menschlicher Handlungen ausschließlich auf die 'Maximierung' persönlicher Interessen gerichtet sind.
- (3) Jede individuelle Präferenzrelation erfüllt diejenigen Rationalitätsbedingungen, die notwendig sind, um ihre Repräsentation in Form einer kardinalen Be-

---

<sup>1</sup> Vgl. A. Schopenhauer, *Die beiden Grundprobleme in der Ethik* (1841) und *Parerga und Paralipomena* (1851). In Anlehnung an indische Denktraditionen gibt es bei Schopenhauer eine Befreiung vom Leiden nur dadurch, daß der Lebenswille verneint wird. In deutlicher Abkehr von indischer Metaphysik aber muß dem die Erkenntnis des "principium individuationis" vorausgehen. Der einzelne erkennt sich daraufhin im anderen wieder und daraus schöpft er ein Gefühl des Mitleidens, das das Fundament moralischen Verhaltens darstellt.

<sup>2</sup> Vgl. Spinoza (1677)

<sup>3</sup> Vgl. A. Smith, *The theory of moral sentiments* (1759)

<sup>4</sup> Vgl. N. Rescher, *Unselfishness*, Pittsburg 1975

<sup>5</sup> I. Kant, *Grundlegung zur Metaphysik der Sitten* (1786); Akademieausgabe, S.398; Reclam-Ausgabe (1974), S.35

wertungsfunktion<sup>1</sup> zu rechtfertigen.

Teleologische sozialetische Theorien sind Interdependenztheorien, die besagen, welche Beziehung zwischen persönlichen Interessen einerseits und moralischen Präferenzen andererseits besteht.

Subjektivistische Theorien halten sich dabei nicht lange bei der Bestimmung der persönlichen Interessen auf, sondern verlassen sich auf die Einschätzungen der Personen selbst.

Nun können wir uns die Einschätzungen der eigenen Interessen in entsprechenden individuellen Präferenzen über  $Z$  repräsentiert denken. Diese Präferenzen sind nicht notwendigerweise die handlungsbestimmenden 'resultierenden' Präferenzen der Personen. Es ist sogar 'Bedingung der Möglichkeit' praktischer Relevanz normativer Theorien, daß sich die resultierenden von den interessenorientierten Präferenzordnungen unterscheiden.

I.a. hat jede Person zusätzlich subjektive moralische Präferenzen, die weder mit den eigenen resultierenden noch mit den interessenorientierten Präferenzen übereinzustimmen brauchen.

Wenn wir in Interessenkonflikten anderer Personen die Position des unparteiischen Beobachters einnehmen können, so wäre es verwunderlich, wenn uns die Tatsache, daß eigene Interessen auf dem Spiel stehen, jeweils derart kopflos machen würde, daß wir alle dort zur Grundlage der moralischen Beurteilung gemachten Prinzipien plötzlich vergessen.

Ist das jedoch nicht der Fall, können wir davon ausgehen, daß eine Anwendung dieser Prinzipien auch in den Fällen, in denen eigene Interessen auf dem Spiel stehen, zu subjektiven moralischen Präferenzen führt, die i.a.

---

<sup>1</sup> Es ist nicht erforderlich, die Bedingungen anzuführen, die dabei von den Präferenzrelationen erfüllt sein müssen, da das jederzeit nachgelesen werden kann; vgl. Fn. S. 125

mit den persönlichen Interessen nicht übereinstimmen - und realistischweise wohl auch nur in seltenen Fällen mit den resultierenden handlungsbestimmenden Präferenzen.<sup>1</sup>

Wir erhalten damit drei numerische Bewertungsfunktionen:

- (1) subjektive Wohlfahrtsfunktionen  $\pi_i$ , die die Präferenzordnungen  $Rw_i$  repräsentieren,
- (2) subjektive moralische Bewertungsfunktionen  $\mu_i$ , die die Präferenzordnungen  $Rm_i$  repräsentieren,
- (3) resultierende subjektive Bewertungsfunktionen  $\nu_i$ , die die Präferenzordnungen  $Rres_i$  repräsentieren.

Da für jedes Alternativenpaar  $x, y$  aus  $Z$  und jede Person  $i$  aus  $K$  eine der folgenden drei Aussagen wahr ist:

- (1)  $x$  ist für  $i$  vorteilhafter als  $y$
- (2)  $x$  ist für  $i$  genauso vorteilhaft wie  $y$
- (3)  $y$  ist für  $i$  vorteilhafter als  $x$ ,

wollen wir auch annehmen, daß die subjektiven Wohlfahrtspräferenzen einer rationalen Person - "i bevorzugt  $x$  gegenüber  $y$  aufgrund persönlicher Interessen" besser: "i meint,  $x$  sei für  $i$  persönlich vorteilhafter als  $y$ " - konnex sind:

$\forall i \in K: \forall x, y \in Z : [ \langle x, y \rangle \in Rw_i \vee \langle y, x \rangle \in Rw_i ]$  bzw.  $\mathcal{D}(\pi_i) = Z$ .

Von den resultierenden Präferenzen nehmen wir an, daß sie in der Weise die individuellen Entscheidungen prägen, wie es dem revealed-preference-concept der modernen Entscheidungstheorie entspricht.<sup>1</sup> Auch die resultierenden Präferenzen seien daher konnex:  $\mathcal{D}(\nu_i) = Z$ .

Wie steht es mit den moralischen Präferenzen? Einige, aber nicht alle Theorien normativer Ethik bestimmen für

---

<sup>1</sup> Ob dieser enge Zusammenhang auch in Interaktionssituationen besteht, beschäftigt uns erst im zweiten Teil der Arbeit (vgl. §12).

jede Alternative  $x, y$  aus  $Z$  eine moralische Präferenz.<sup>1</sup> Man kann das dadurch zum Ausdruck bringen, daß man eine entsprechende Eingrenzung des Definitionsbereichs der moralischen Bewertungsfunktionen zuläßt<sup>2</sup> (womit allerdings auch die Konnexitätsforderung aufgegeben ist), oder man behandelt den Fall der expliziten moralischen Gleichwertigkeit und den Fall, daß eine Alternative keiner moralischen Bewertung unterliegt, gleichermaßen als moralische Indifferenz (damit wäre wiederum Konnexität der moralischen Präferenzen garantiert).<sup>3</sup> Aus Gründen der klareren Darstellungsweise, entscheiden wir uns hier für die zweite Möglichkeit:

$$\forall i \in K : \forall x, y \in Z : [ \langle x, y \rangle \in Rm_i \vee \langle y, x \rangle \in Rm_i ] \text{ bzw. } D(\pi_i) = Z.$$

Zur Bestimmung der Präferenzen  $Rm_i$  und  $Rw_i$  kann man von der Fähigkeit der meisten Menschen ausgehen,

<sup>1</sup> Warum geht die moralische Präferenzrelation nicht über  $\bar{E}$ , könnte man fragen, da ja subjektiven Präferenzen der betroffenen Personen eine Rolle bei der Festlegung moralischer Präferenzen spielen. Eine moralische Bewertungsfunktion über  $\bar{E}$ , die z.B. in folgender Weise aus der moralischen Bewertungsfunktion über  $Z$  hervorginge

$$\mu_i(z) = \sum_{i \in K} \pi_i(z)$$

$$\mu_i(\xi) = \sum_{i \in K} \pi_i(z), \text{ wobei } \xi = \langle z, \langle \pi_i(z) \rangle \rangle$$

würde auch eine implizite moralische Bewertung der Wohlfahrtsfunktion  $\pi_i$  enthalten.

Denn:

$$\mu_i(\xi) > \mu_i(\xi'), \text{ wobei}$$

$$\xi = \langle z, \langle \pi_i \rangle \rangle \text{ und } \xi' = \langle z, \langle \pi_i' \rangle \rangle$$

müßte wohl so interpretiert werden, daß die  $\langle \pi_i \rangle$  bei  $z$  moralisch besser sind als die  $\langle \pi_i' \rangle$  bei  $z$ .

<sup>2</sup> Wir sprechen hier von Theorien als ideale Anleitung subjektiver moralischer Präferenzen, in der Annahme, daß sich wesentliche Unterschiede auf der theoretischen Ebene auch in den moralischen Präferenzen widerspiegeln.

<sup>3</sup> Die meisten ethischen Theorien bestimmen nicht für beliebige Paare von Anax-Zuständen eine moralische Präferenz. Deontologische Theorien begründen i.a. nur eine rudimentäre moralische Bewertungsfunktion. Im Extremfall ist die moralische Präferenzrelation dergestalt, daß sich die Wertemenge der moralischen Bewertungsfunktion auf zwei Elemente beschränkt.

zwischen ihren persönlichen Interessen und dem, was sie tun sollten einerseits, aber auch dem, was sie tatsächlich zu tun geneigt sind andererseits zu unterscheiden, und durch Befragungen die Alternativen in die entsprechenden Präferenzordnungen zu bringen.<sup>1</sup>

Die subjektive Wohlfahrtsfunktion unter dem Aspekt des persönlichen Vorteils, vom Handlungsträger selbst gebildet, ist natürlich i.a. nur eine erste Annäherung an dessen tatsächliche individuelle Wohlfahrtsfunktion. Irrtümliche Annahmen darüber, welche Merkmale einzelner Alternativen seinen persönlichen Interessen in welchem Ausmaß entsprechen, und eine falsche Einschätzung der persönlichen Interessen selbst, können eine solche Präferenzordnung sogar zu einer ziemlich ungeeigneten Grundlage für die Bestimmung der tatsächlichen individuellen Wohlfahrtsfunktion machen. Ob das ein Defizit ist, hängt vom methodologischen Charakter der sozialetischen Theorie ab, die von diesen Bewertungsfunktionen Gebrauch macht.<sup>2</sup>

In § 6 haben wir uns eingehend mit der interpersonellen Vergleichbarkeit subjektiver Bewertungen befaßt, ohne dort schon verschieden Arten von Bewertungen zu unterscheiden.

Zweierlei hat dort die Möglichkeit interpersoneller Bewertungsvergleiche nahegelegt: Einmal das Argument für die Beschränktheit subjektiver Bewertungsfunktionen (und damit die Möglichkeit ihrer Normierung) und zum zweiten die Tatsache, daß interpersonelle Vergleiche im Einzelfall (als empirische Grundlage von Gerechtigkeitsempfehlungen) im 'moralischen Alltag' unumstritten sind und

---

<sup>1</sup> Zur Kardinalisierung müßten wir für alle drei Präferenzarten zulassen, daß es jeweils auch Präferenzen der gleichen Art zwischen Lotterien gibt. In §10 gehen wir auf dieses Problem kurz ein.

<sup>2</sup> Mit diesen metaethischen Fragen setzt sich §11 auseinander.

in Form intrapersoneller Präferenzeninterdependenz rekonstruiert werden können.

Auch hier können wir zwar davon ausgehen, daß alle drei Bewertungsfunktionen beschränkt sind. Aber eine Methode kardinaler interpersoneller Vergleiche von 'moralischen Niveaus'  $\mu_i(z)$  macht kaum Sinn, während interpersonelle Vergleiche von Nutzenniveaus  $\pi_i(z)$  alltäglich sind.

Die zweite Stütze interpersoneller Bewertungsvergleiche fällt daher sowohl bei moralischen wie bei resultierenden Präferenzen weg.

Allerdings ist interpersonelle kardinale Vergleichbarkeit dieser beiden Bewertungsfunktionen auch in keiner Weise erforderlich, denn Aggregationstheorien der Sozialethik sollten sich nicht auf resultierende und - a fortiori - nicht auf subjektiv moralische Präferenzen beziehen, wie die Überlegungen in §7 und in §8 gezeigt haben.

#### EINE ERSTE ANWENDUNG DES DIFFERENZIIERTEN PRÄFERENZBEGRIFFS : DAS LIBERALE PARADOXON

Um den Sinn dieser Differenzierung des Präferenzbegriffs schon hier deutlich zu machen - im zweiten Teil der Arbeit machen wir von ihr eingehend Gebrauch - wenden wir sie auf ein Paradoxon der Theorie kollektiver Wahl an. Dieses Paradoxon hat Sen entdeckt, er bezeichnet es als "the impossibility of a paretian liberal".<sup>1</sup>

---

<sup>1</sup> Vgl. A.K. Sen, *The impossibility of a paretian liberal*, Journal of political economy 78 (1970), dargelegt auch in Sen (1970a), 6\* als 'liberal paradox'.

Nach der Veröffentlichung des liberalen Paradoxons hat eine Diskussion darüber eingesetzt, wie es zu beheben sei - durch Einführung eines persönlichen Entscheidungsbereiches (Nozick (1973)), durch kardinale Nutzenvergleiche (Ng (1971)), andere Bedingungen (Blau (1975), Gibbard (1974)), Abschwächung der Liberalismusbedingung (Osborne (1975)) etc., die z.T. sehr interessante Ergebnisse enthalten. Sen (1976) stellt den Konflikt von Liberalismus und Einstimmigkeit in Auseinandersetzung mit diesen Kritiken und mit einigen Erweiterungen erneut dar. Seitdem ist die Diskussion zur Versöhnung von Liberalität und Pareto-Regel nicht abgeschlossen, vgl. Gaertner/Krüger (1978 und 1982).

Zunächst in formaler Darstellung:

Als eine minimale Bedingung des Liberalismus (L) für eine Gruppe K wird gefordert:

$$\forall i \in K \exists \langle z, z' \rangle \in Z \times Z: z \neq z' \wedge (z >_i z' \Rightarrow z > z') \wedge (z <_i z' \Rightarrow z < z') \quad (L)$$

Diese Liberalismusbedingung ist nicht vereinbar mit der Pareto-Regel P:

$$\forall \langle z, z' \rangle \in Z \times Z: [(\forall i \in K: z >_i z') \Rightarrow z > z'] \quad (P),$$

sofern wir verlangen, daß die Aggregation individueller Präferenzordnungen zu einer kollektiven azyklischen<sup>1</sup> Präferenzrelation für n-Tupel beliebiger (logisch zulässiger) konnexer individueller Präferenzordnungen möglich ist.<sup>2</sup> (A)

Beweis:<sup>3</sup>Für #K ≥ 2 ∧ #Z ≥ 2; Seien z<sub>1</sub>, z<sub>2</sub>, z<sub>3</sub>, z<sub>4</sub> ∈ Z; k, l ∈ K ∧ k ≠ l

1. Fall: <z<sub>1</sub>, z<sub>2</sub>> = <z<sub>3</sub>, z<sub>4</sub>>

$$\text{angenommen } z_1 >_k z_2 \wedge z_4 >_l z_3 \quad \stackrel{(L)}{\Rightarrow} z_1 > z_2 \wedge z_2 > z_1$$

⇒ mit L: sowohl z<sub>1</sub> > z<sub>2</sub>, wie z<sub>2</sub> > z<sub>1</sub> (Widerspruch)

2. Fall: {z<sub>1</sub>, z<sub>2</sub>} und {z<sub>3</sub>, z<sub>4</sub>} haben ein gemeinsames Element. o.B.d.A. sei z<sub>1</sub> = z<sub>3</sub>.

$$\text{angenommen } z_1 >_k z_2 \wedge z_4 >_l z_1 \wedge \forall i \in K: z_2 >_i z_4$$

⇒ mit L: z<sub>1</sub> > z<sub>2</sub> und z<sub>4</sub> > z<sub>1</sub> } ⇒ z<sub>1</sub> > z<sub>2</sub> > z<sub>4</sub> > z<sub>1</sub>,  
mit P: z<sub>2</sub> > z<sub>4</sub> } das widerspricht der Azyklität kollektiver Präferenzen.

<sup>1</sup> Eine binäre Relation > über Z ist azyklisch: ⇔ ∀ z<sub>1</sub>, z<sub>2</sub>, ..., z<sub>v</sub> ∈ Z. z<sub>1</sub> > z<sub>2</sub> > z<sub>2</sub> > z<sub>3</sub> ∧ ... ∧ z<sub>v-1</sub> > z<sub>v</sub> ⇒ z<sub>1</sub> > z<sub>v</sub>

<sup>2</sup> Bei Sen (1970c) die Bedingung U (unrestricted domain), vgl. 3\*, bei Arrow (1963) als Bedingung 1' in der späteren Exposition der Beweise für das "general possible theorem" (S.96ff)

<sup>3</sup> In diesem Beweis machen wir nicht von der minimalen Liberalismusbedingung L\*

$$\exists k, l \in K \exists \langle z_1, z_2 \rangle, \langle z_3, z_4 \rangle \in Z \times Z [ \{ \langle z_1, z_2 \rangle \neq \langle z_3, z_4 \rangle \wedge \{ z_1 >_k z_2 \Rightarrow z_1 > z_2 \} \wedge \{ z_1 <_k z_2 \Rightarrow z_1 < z_2 \} \wedge \{ z_3 >_l z_4 \Rightarrow z_3 > z_4 \} \wedge \{ z_3 <_l z_4 \Rightarrow z_3 < z_4 \} ]$$

Gebrauch, die Sen (1970c)6\* und Kern (1980), S.37 in ihren Beweisen verwenden (so daß sich bei ihnen der Satz als Korrolar ergibt), weil das erstens mathematisch unnötig ist, und L\* zweitens mit 'Liberalismus' kaum mehr etwas zu tun hat. Das ist schon daraus ersichtlich, daß eine weitere unwesentliche Abschwächung von L\* zu

3. Fall  $\#\{z_1, z_2, z_3, z_4\} = 4$

angenommen:  $z_1 >_k z_2 \wedge z_3 >_l z_4 \wedge \forall i \in K \{z_4 >_i z_1 \wedge z_2 >_i z_3\}$

$\Rightarrow$  mit L:  $z_1 > z_2$  und  $z_3 > z_4$   
 mit P:  $z_4 > z_1$  und  $z_2 > z_3$  }  $\Rightarrow z_1 > z_2 > z_3 > z_4 > z_1$ , das widerspricht der Azyklizität kollektiver Präferenzen

Nun ist die Liberalismusbedingung L eine sehr schwache Garantie für eine liberale kollektive Entscheidungsfindung, dennoch ist selbst die weiter abgeschwächte Bedingung L\*, die fordert, daß es in einer Personengruppe K wenigstens zwei Personen geben sollte, die jeweilils für wenigstens ein Alternativenpaar die kollektive Präferenz bestimmen, mit der Pareto-Regel (unter A) nicht vereinbar, wie der Gang des Beweises zeigt, in dem wir ja nur von zwei Individuen und vier Zuständen Gebrauch gemacht haben.

Auch die Bedingung L ist bei weitem zu schwach, um die individuelle Entscheidungsfreiheit zu garantieren, die für jede Verfassung selbstverständlich ist. Ich darf nicht nur frei entscheiden, auf welcher Seite ich gern einschlafen möchte, sondern auch, welche Farbe mein Auto hat, welche Zeitung ich lese, welchen Beruf ich anstrebe, wo ich wohne etc., und in fast allen diesen Fällen steht mehr als ein Alternativenpaar zur Wahl.

Es gibt in allen Gesellschaften einen persönlichen Entscheidungsbereich, in dem die Präferenzen anderer Personen irrelevant sind. Wenn eine Mehrheit sich wünschen sollte, daß Herr i die Bildzeitung liest, so entscheidet dennoch allein Herr i, ob er das tut oder nicht - in diesem Sinne ist seine Präferenz entscheidend für die kollektive Präferenz. Ist z der Zustand, in dem Herr i die Bildzeitung liest, und z' ein Zustand, der z in allem gleicht, nur daß i in z' die Bildzeitung nicht liest, so gilt:  $[z >_i z' \Rightarrow z > z'] \wedge [z <_i z' \Rightarrow z < z']$ .

L\*\*  $\exists k \in K \exists \langle z_1, z_2 \rangle \in Z \times Z: \{z_1 >_k z_2 \Rightarrow z_1 > z_2\} \wedge \{z_1 <_k z_2 \Rightarrow z_1 < z_2\}$

dazu führt, daß Liberalismus und Ein-Mann-Diktatur

D  $\exists k \in K \forall \langle z_1, z_2 \rangle \in Z \times Z: \{z_1 >_k z_2 \Rightarrow z_1 > z_2\}$

vereinbar wären.



Die Liberalismusbedingung ist also ein zunächst sehr überzeugendes normatives Minimal Kriterium für Aggregationsregeln kollektiver Wahl. L erscheint in jedem Fall als eine notwendige, aber bei weitem nicht hinreichende Bedingung für die normative Adäquanz einer Aggregationsregel. Das gleiche gilt für das Pareto-Kriterium: Wenn alle Personen aus K einen Zustand  $z$  einem Zustand  $z'$  individuell vorziehen, so sollte  $z$  auch kollektiv vorgezogen werden.

Sen illustriert das liberale Paradoxon mit folgendem Beispiel: A ist eine prude und B eine laszive Person. Es gibt eine Exemplar eines frivolen Buches "Lady Chatterly's lover". Drei Zustände stehen zur Entscheidung:

$z_1$  : A liest das Buch und B liest das Buch nicht.

$z_2$  : B liest das Buch und A liest das Buch nicht.

$z_3$  : Keiner liest das Buch.

A hat die Präferenzen  $z_3 > z_1 > z_2$ , denn am liebsten ist ihm natürlich, daß niemand das Buch liest; falls es jedoch einer liest, dann lieber er, denn er wird daran weniger Schaden nehmen als B.

B hat die Präferenzen  $z_1 > z_2 > z_3$ , denn eigentlich wäre es ihm am liebsten, wenn jeder das Buch liest, aber diese Alternative steht nicht zur Wahl. Wenn jedoch nur einer das Buch liest, dann meint er, sollte es A lesen - sozusagen aus pädagogischen Gründen.

Beide haben es lieber, daß die prude Person (A) als daß die laszive Person (B) das frivole Buch liest.

$z_1$  ist pareto-besser als  $z_2$ .

Da die Lektüre von Büchern jedoch in den persönlichen Entscheidungsbereich gehört, sollte A nicht dazu gezwungen werden:  $z_3$  ist  $z_1$  vorzuziehen. Ähnlich gilt für B, daß er selbst entscheiden sollte, ob er es lesen will:

$z_2$  ist  $z_3$  vorzuziehen.

Es ergibt sich also die zyklische kollektive Präferenzrelation :  $z_1 > z_2 > z_3 > z_1$ .

Die Situation ist verfahren, und es gibt eine Reihe von Möglichkeiten, sich mit ihr auseinanderzusetzen, die wir hier nicht diskutieren wollen.<sup>1</sup> Aber eine Möglichkeit besteht in der Anwendung unserer Begriffsdifferenzierung, und zu diesem Zweck haben wir das Beispiel gewählt.

(1) A fühlt sich besser, wenn er das Buch nicht lesen muß. Zumindest ist A davon überzeugt, daß es seinen Interessen eher entspricht, das Buch nicht zu lesen. (Selbstverständlich kann man auch falsche Vermutungen über seine eigenen Interessen haben)

(2) B hat ein Interesse daran, das Buch zu lesen.

(3) Über (1) und (2) gibt es zwischen A und B keine Meinungsverschiedenheiten. Beide sind von den starken (subjektiven) Wohlfahrtspräferenzen  $\langle z_2, z_1 \rangle \in PW_A$  und  $\langle z_2, z_1 \rangle \in PW_B$  überzeugt.

Die persönlichen Interessen von A, so dürfen wir annehmen, werden in  $z_2$  und  $z_3$  gleichermaßen (gut) berücksichtigt, und die persönlichen Interessen von B werden in  $z_1$  und  $z_3$  gleichermaßen (schlecht) berücksichtigt.

$$\pi_A(z_2) > \pi_A(z_1) = \pi_A(z_3). \quad (I)$$

$$\pi_B(z_3) = \pi_B(z_2) > \pi_B(z_1). \quad (II)$$

Wenden wir das Liberalitätskriterium L auf die Aggregation dieser subjektiven Wohlfahrtspräferenzen an, so ist  $z_2$  kollektiv besser als  $z_1$ , denn man sollte niemanden dazu zwingen, ein Buch zu lesen, das er nicht lesen will, ebensowenig sollte man niemandem verbieten, ein Buch zu lesen, falls das nicht mit dem Interesse einer anderen Person in Konflikt kommt, ebenfalls dieses Buch zu lesen - was hier nicht der Fall ist. Die Pareto-Regel verlangt, daß  $z_2$  kollektiv  $z_1$  vorgezogen wird. Werden beide normativen Aggregationsregeln L und P auf die subjektiven Wohlfahrtspräferenzen angewendet, so ist es nicht nur möglich, eine damit konsistente kollektive azyklische Präferenzordnung anzugeben, sondern es ist sogar unmög-

---

<sup>1</sup> Vgl. Fn S. 115

lich, aufgrund dieser Kriterien eine kollektive zyklische Präferenzordnung aufzustellen.

Zwar ließe sich aufgrund von L zusätzlich die kollektive Präferenz  $z_3 > z_1$  (II) und die kollektive Präferenz  $z_2 > z_3$  (I) rechtfertigen, aber damit ergibt sich die kollektive konnexe azyklische Präferenzordnung  $z_2 > z_3 > z_1$ ,<sup>1</sup> die auch ganz unserer moralischen Intention entspricht.

Wie kommt es dann überhaupt zu dem Paradoxon in diesem Beispiel? Neben persönlichen Interessen haben die beiden Personen A und B auch moralische Präferenzen. Zwar liest B das Buch lieber selber (was seine persönlichen Interessen angeht), aber zugleich ist er der Ansicht, A, der frühe Zeitgenosse, sollte das Buch lesen. Wie auch immer B dieses Urteil begründet, i.a. werden wir es als ein normatives Urteil interpretieren müssen, denn nur in Sonderfällen wird B überzeugend darlegen können, daß es in seinem persönlichen Interesse liegt, wenn A das Buch liest. Das könnte z.B. dann der Fall sein, wenn B mit A zusammenlebt und daher ein persönliches Interesse an einer toleranteren Lebenshaltung seines Wohngenossen hat.

A hat die moralischen Präferenzen:  $z_3 \succ z_1 \wedge z_1 \succ z_2$

B hat die moralischen Präferenzen:  $z_1 \succ z_2 \wedge z_2 \succ z_3$

Zum Vergleich die Wohlfahrtspräferenzen:

A :  $z_2 \succcurlyeq z_1 \wedge z_2 \succcurlyeq z_3$

B :  $z_2 \succcurlyeq z_1 \wedge z_1 \succcurlyeq z_3$

A und B haben unterschiedliche Ansichten darüber, was (moralisch) besser ist. Wir haben festgestellt, daß es grundsätzlich nicht zulässig ist, Aggregationsregeln auch auf unterschiedliche individuelle Überzeugungen anzuwenden - und das zeigt sich erneut an diesem Beispiel. Es ist eine sinnvolle Feststellung, ob die (ethische)

---

<sup>1</sup> Die kollektive Präferenzordnung muß nicht konnex sein. Entscheidend ist Azyklizität und die ist gegeben, denn zu  $z_2$  gibt es keinen kollektiv besseren Zustand.

München

kollektive Präferenzrelation sich in mindestens einer Präferenz nach den Interessen einer einzelnen Person richten sollte (L). Ebenso kann man sich überlegen, ob ein Zustand  $z$  einem anderen Zustand  $z'$  kollektiv vorgezogen werden sollte, wenn alle Personen ihre Interessen in  $z$  besser als in  $z'$  verwirklicht sehen. Aber es wäre inadäquat zu fordern, daß es für jede Person mindestens ein Alternativenpaar geben sollte, von dem allein diese Person entscheidet, welche Alternative moralisch besser ist. Ist es denn sichergestellt, daß sich jede Person in mindestens einer moralischen Präferenz nicht irrt? Und wenn ja, wie macht man diese subjektiven moralischen Präferenzen ausfindig, die nicht auf falschen moralischen Urteilen beruhen? Sofern die Aggregationsregeln der Theorie kollektiver Wahl auf  $n$ -Tupel individueller Wohlfahrtspräferenzrelationen angewendet werden, wird aufgrund gegebener persönlicher Interessen eine moralische besser-Relation über die Alternativen gebildet. Und so ist die sozialwahltheoretische Diskussion unterschiedlicher normativer Aggregationsregeln i.a. auch gemeint - unbeschadet der Tatsache, daß diese Interpretation in einem gewissen Spannungsverhältnis zur modernen ökonomischen Nutzentheorie steht.<sup>1</sup> Explizit gemacht werden kann sie wohl nur durch eine Differenzierung des Präferenzbegriffes. Natürlich sind dabei andere Differenzierungen als die von uns vorgeschlagenen denkbar. In dem *Senschen* Beispiel für das Auftreten des liberalen Paradoxons hat diese Begriffsdifferenzierung jedenfalls eine überraschend segensreiche Wirkung: Das Paradoxon verschwindet, wenn die Aggregationsregeln L und P nur auf die individuellen Wohlfahrtspräferenzen angewendet werden.

Es stellt sich nun die Frage: Gilt das allgemein? Und die Antwort lautet: In einem Sinn ja und in einem anderen Sinn nein. In einem formalen Sinn tritt das Paradoxon auch dann auf, wenn der Anwendungsbereich der Aggrega-

---

<sup>1</sup> Vgl. Harsanyi (1975c).

tionsregeln auf individuelle Wohlfahrtspräferenzen beschränkt wird, denn wir lassen ja beliebige (logisch zulässige) individuelle Präferenzordnungen auch bei den Wohlfahrtspräferenzen zu, und damit gilt Satz und Beweis (vgl. S. 151f) unverändert. Aber darüberhinaus lautet nun die These: Es wird sich kein intuitiv überzeugendes Beispiel mehr für dieses Paradoxon finden lassen (d.h. eine Kombination individueller Wohlfahrtspräferenzrelationen, die bei der Anwendung von P und L zu einer zyklischen kollektiven Präferenzordnung führt). Und in diesem 'materialen' Sinn ist das liberale Paradoxon infolge der vorgeschlagenen Begriffsdifferenzierung aufgelöst. Der Grund dafür liegt darin, daß sich der persönliche Entscheidungsbereich - der Bereich also, in dem die individuelle Wohlfahrtspräferenz für die kollektive Präferenz entscheidend sein soll - gerade auf solche Teilmengen  $Z_i \subset Z$  beschränkt, deren unterschiedliche Alternativen  $z \in Z_i$  zwar ein verschiedenes Wohlfahrtsniveau  $\pi_i(z)$  für  $i$  mit sich bringen, jedoch im übrigen keinen oder einen vernachlässigbar kleinen Unterschied für die Wohlfahrtsniveaus anderer Personen nach sich ziehen:

$$\forall j \in K \quad \forall z, z' \in Z_i : j \neq i \Rightarrow \pi_j(z) = \pi_j(z')$$

Das gleiche läßt sich nicht sinnvoll für den persönlichen Entscheidungsbereich bei resultierenden Präferenzen fordern, denn auch dann, wenn die persönlichen Interessen nicht tangiert sind, pflegen Präferenzen über das, was andere Personen zu tun oder zu lassen haben, nicht zu verschwinden; d.h. sind zwei Zustände indifferent bezüglich des persönlichen Wohlergehens, so hat das i.a. nicht zur Folge, daß diese Zustände auch bezüglich der individuellen resultierenden Präferenzen indifferent sind: Aus  $\pi_i(z) = \pi_i(z')$  folgt i.a. nicht  $v_i(z) = v_i(z')$ .

## ZUR INTERPRETATION INDIVIDUELLER WOHLFAHRTSFUNKTIONEN

Trotz dieses Erfolges in der Anwendung bleibt eine Frage offen: Wie läßt sich diese Begriffsdifferenzierung 'operationalisieren'? Welcher Test erlaubt es, diese drei Arten individueller Präferenzrelationen auseinanderzuhalten?

Bezüglich der resultierenden Präferenzen können wir uns auf das Standardkonzept der 'revealed preferences' stützen.

Woran man (subjektive) moralische Präferenzen einer Person erkennt, glauben wir i.a. ganz gut zu wissen, und es gibt dazu eine traditionsreiche metaethische Debatte, besonders im Umfeld des ordinary-language-approach in der analytischen Philosophie.

Eine Schwäche der Differenzierung des Präferenzbegriffs liegt jedoch darin, daß wir zur Interpretation der individuellen Wohlfahrtsfunktionen nicht viel mehr anbieten können als die These, es sei ein Merkmal rationaler Personen, zwischen ihren persönlichen Interessen, ihren moralischen Urteilen und ihren tatsächlichen (resultierenden) Präferenzen unterscheiden zu können. Aber wer ist sich schon wirklich sicher, seine persönlichen Interessen zu kennen? Diese vorgeschlagene Differenzierung des Präferenzbegriffs würde daher an Überzeugungskraft gewinnen, wenn sich die persönlichen Interessen bzw. die individuellen Wohlfahrtsfunktionen ohne Rekurs auf eine Form der Selbstbeurteilung bestimmen ließen.

Wenn die Unterschiede von Wohlfahrts- und resultierenden Präferenzen ausschließlich auf der Berücksichtigung fremder Interessen beruhen, müßten sich die Wohlfahrtspräferenzen einer Person als ihre resultierenden Präferenzen in einer speziellen Situation des Unwissens ergeben:

Stellen wir uns Personen eines Kollektivs K vor, denen jegliches Wissen über die persönlichen Wünsche der ande-

ren Personen aus K abgeht. Dieses Unwissen sei nicht nur distributiv sondern sogar kollektiv, d.h. jede Person aus K weiß nicht nur von keiner anderen Person, welche Präferenzen diese hat (und zwar weder bezüglich ihrer persönlichen Interessen noch bezüglich ihrer moralischen Beurteilung), sondern jede Person aus K weiß auch nicht, in welchem Ausmaß bestimmte Präferenzen in K verbreitet sind, ja die Personen können darüber nicht einmal allgemeine Wahrscheinlichkeitsannahmen machen, da ihnen das entsprechende Wissen fehlt.

Müßten sich nicht in einer solchen Situation die resultierenden Präferenzen den Wohlfahrtspräferenzen annähern?

Die Bedingung, in einer solchen Situation des Unwissens keine Wahrscheinlichkeitsannahmen machen zu können, hat eine Parallele in der Konstruktion der "original position" in Rawls' Gerechtigkeitstheorie.<sup>1</sup> Auch dort hängt das Ergebnis wesentlich davon ab, daß in der 'original position' keine Entscheidung unter Unsicherheit, sondern eine Entscheidung unter Risiko vorliegt. In der Debatte um diese Frage sind viele Argumente ausgetauscht worden, die auch für diesen Explikationsversuch relevant sind.<sup>2</sup> Dennoch wollen wir es bei diesem Vorschlag bewenden lassen. Ob er akzeptabel ist, hängt fast ausschließlich davon ab, ob man die Charakterisierung einer Situation als einer 'in der man mangels Wissen keine Wahrscheinlichkeitsannahmen machen kann' für zulässig hält. Wenn einen das 'Argument des unzureichenden Grundes' überzeugt, wenn man also glaubt, in einer Situation, in der man keinen Grund hat, unterschiedliche

---

<sup>1</sup> Vgl. J. Rawls, *A theory of justice*, New Haven 1971, bes. §4 und §20 - §26 und §28.

<sup>2</sup> Besonders Harsanyi (1975b), aber auch Hare (1973). Zu der bis heute andauernden Diskussion um den methodischen Status und die innere Konsistenz der rationalen Wahl einer Verfassung in in Rawls' original position vgl.a.: Dworkin (1973); Nowell-Smith (1973); Barry (1973) bes. Kap.2, 9 u.11; Keyt (1974); Lyons (1976); Buchanan (1977), §9,10,13,14; Tugendhat (1977); Wolff (1977) bes. VI-XII; Ellsworth (1979); Fauvre (1978); Gauthier (1978); Gray (1978); Lehning (1978); Essler (1979), §16; Brehmer (1980), S.64-82; Katzner (1980); Miller (1981); Potter (1981); Rothkirch (1981), Kap.3;

Wahrscheinlichkeiten anzunehmen, Grund zu haben, gleiche Wahrscheinlichkeiten anzunehmen, so muß man auch diesen Explikationsversuch zurückweisen.

Es ist nicht leicht, sich die vollen Konsequenzen einer solchen Informationsbeschränkung klar zu machen. Z.B. kann die moralische Beurteilung der Zustände unter dem Gesichtspunkt der Verteilungsgerechtigkeit für die resultierende Präferenzordnung keine Rolle mehr spielen. Würde sich die moralische Beurteilung ausschließlich in der Berücksichtigung (der Interessen und Wünsche) anderer Personen erschöpfen - wie viele Moralphilosophen angenommen haben, würde die (resultierende) individuelle Präferenzordnung in einer solchen Situation des Unwissens  $Rw_1$  nicht nur approximieren, sondern sogar mit dieser identisch sein.

Allerdings ist unsere Approximationsthese nicht einmal ethisch neutral. Wer auch in einer Ein-Personen-Welt normative Ethik für anwendbar hält, z.B. weil für ihn das  $\epsilon \cup \zeta \bar{\eta} \vee$  eine zentrale Kategorie der Moral ist, der müßte diese Approximationsthese ebenfalls zurückweisen.

#### ERWEITERUNGEN

In den meisten Fällen, in denen von Präferenzen und Präferenzenaggregation die Rede ist, genügt diese vorgeschlagene Differenzierung des Präferenzbegriffes. Aber manche Situationen sperren sich nach wie vor einer Analyse. Ein Beispiel:

"Schmidt und Meier diskutieren darüber, wie Schmidt sich in einem bestimmten Fall verhalten soll. Schmidt meint, er sollte X tun; doch Meier meint, Schmidt sollte Y tun. Beide nennen ihre Gründe, doch keiner überzeugt den anderen. Schmidt ist beunruhigt und fragt Meier, wie er sich denn nun verhalten sollte. An diesem Punkt sagt Meier vielleicht 'Ich bin immer noch der Meinung, du solltest Y tun'. Es ist aber auch denkbar, daß er sagt, 'Du solltest das tun, was du für richtig hältst' oder 'Tu das, was dein Gewissen dir



sagt'. Das legt den Schluß nahe, daß wir uns all unsere Überlegungen hätten sparen und einfach sagen können 'Handle stets so, wie du es für richtig hältst' oder 'Laß dich von deinem Gewissen leiten.'" <sup>1</sup>

Oder man denke an folgende Situation:

Ein Demokrat gibt sein Votum für die Alternative z ab. Er ist aus moralischen Gründen der Ansicht, z sei besser als jede andere Alternative - in diesem Sinne ist er der Ansicht, z sollte verwirklicht werden. Die Mehrheit im abstimmenden Gremium ist jedoch für eine andere Alternative z'. Als Demokrat ist er daher der Ansicht, z' sollte verwirklicht werden. Nach wie vor mag er aber überzeugt davon sein, daß z die bessere Alternative ist. Er präferiert damit eine Alternative z' gegenüber z, die er für moralisch schlechter hält als z. Dies ist nur dann eine ethische Antinomie, wenn uns nur eine Art subjektiver moralischer Präferenz zur Verfügung steht.

"Personally I would try to persuade p never to make a collective choice rule a matter of his morality."<sup>2</sup>

Andererseits mag es Demokraten geben, die sich sogar massiv dafür einsetzen, daß die Alternative, für die sich die Mehrheit entschieden hat, auch verwirklicht wird, und daß auch die in der Abstimmung Unterlegenen diese Entscheidung für z' 'mittragen'. Es wäre keineswegs paradox, eine solche Haltung sogar als moralisch geboten zu bezeichnen.<sup>3</sup>

Damit hätten wir im Rahmen des bisherigen Begriffsapparates einen logischen Widerspruch. Da die beschriebene 'Präferenzsituation' jedoch durchaus vorkommt, kann man dieses Beispiel als eine Aufforderung ansehen, die Differenzierung des Präferenzbegriffes zu erweitern.

### 1. Bedingte und unbedingte Präferenzen

In einem entscheidungstheoretischen Modell, in dem individuelle Präferenzen über Anax-Zustände gehen, können

---

<sup>1</sup> Frankena (1972), S.75f

<sup>2</sup> Stegmüller (1977), S.65

<sup>3</sup> Allerdings in den Grenzen, die eine subjektivistische Ethik eben hat. Vgl. dazu §11.

Bedingungen von Präferenzen nur Präferenzen sein.

Intuitiv (und alltagssprachlich) beziehen sich Präferenzen i.a. jedoch nicht auf Zustände, sondern auf Propositionen: "Ich habe Kirschkuchen lieber als Sahnetorte", "Ich präferiere Kirschkuchen gegenüber Sahnetorte" bzw. 'propositional' formuliert: "Ich ziehe es zu jedem Zeitpunkt  $t_0$  vor, daß ich in  $t_0$  Kirschkuchen esse, als daß ich in  $t_0$  Sahnetorte esse".

Propositionale Präferenzen sind in unserem Modell<sup>1</sup> Präferenzen über Teilmengen von  $\Xi$ . Welcher Zusammenhang besteht nun zwischen Präferenzen über  $\Xi$  und Präferenzen über  $\text{Pot}(\Xi)$  (bzw. über  $Z$  und über  $\text{Pot}(Z)$ )?

$p, p'$  sind Propositionen;  $p, p' \subset \Xi$ . Nehmen wir an, sowohl die propositionalen wie die Zustandspräferenzen seien repräsentiert durch eine kardinale Bewertungsfunktion  $\varnothing_1$  über  $\text{Pot}(\Xi)$  und  $\upsilon_1$  über  $\Xi$ .

Wir können auch weiterhin davon ausgehen, der Zusammenhang von  $\varnothing_1$  und  $\upsilon_1$  sei der in §2 geschilderte.<sup>2</sup> Aber damit ist noch nicht alles über diesen Zusammenhang ausgesagt.

Propositionale Präferenzen sind i.a. bedingte Präferenzen. Daß eine propositionale Präferenz bedingt ist, läßt sich durch die Angabe einer einzigen Bedingung belegen, bei der die Präferenz nicht bestehen bleibt; daß eine Präferenz nicht bedingt (oder unbedingt) ist, läßt sich nicht definitiv nachweisen, denn das hieße, alle logisch möglichen Bedingungen daraufhin zu testen, ob die Präferenz von ihnen unabhängig ist. Das schließt nicht aus, daß wir eine solche Hypothese für in hohem Grade bestätigt halten, falls mehrere Versuche, eine Bedingtheit festzustellen, fehlgeschlagen sind.

---

<sup>1</sup> Vgl. §2 S. 21

<sup>2</sup> Vgl. §2 S. 19

Ein Beispiel:

Jemand sagt, er bevorzuge es heute auswärts zu essen. Aber als er erfährt, daß die beiden nahegelegenen Gasthäuser heute Ruhetag haben, zieht er es vor zuhause zu bleiben. Man könnte sagen, seine Präferenz auswärts zu essen habe sich angesichts neuer Informationen geändert - oder aber: Seine Präferenz auswärts zu essen war u.a. davon abhängig, daß die beiden Gasthäuser heute keinen Ruhetag haben.

Auch moralische Präferenzen sind i.a. bedingte Präferenzen: "Ein höheres gesetzliches Mindesteinkommen ist aus Gründen der sozialen Gerechtigkeit dem jetzigen vorzuziehen" ist eine moralische Präferenz, die etwa nur insoweit bestehen mag, als ein höheres gesetzliches Mindesteinkommen nicht zu zahlreichen Konkursen und im Gefolge zu einem starken Ansteigen der Arbeitslosigkeit führen würde. Anders formuliert, es ist eine Bedingung dieser moralischen Präferenz, daß der genannte Zusammenhang nicht besteht.

Sen unterteilt die Werturteile in zwei Klassen: 'basic' und 'non-basic judgements'. 'Basic' sind Urteile, die unter allen faktischen Umständen gelten; 'non-basic' alle übrigen.<sup>1</sup>

Diese Unterscheidung läßt sich zwanglos in unsere Terminologie der bedingten moralischen Präferenzen übersetzen. Wobei Sen den Terminus "value judgement" in einer anderen Bedeutung verwendet als wir.<sup>2</sup> Einmal beschränkt er 'value judgement' offensichtlich auf den ethischen Bereich, und zum anderen bezeichnet Sen auch

---

<sup>1</sup> Vgl. Sen (1970a), S.59-64.

<sup>2</sup> Vgl. §11.

nicht-axiologische normative Urteile als "value judgement". Das 'value judgement' "men and women should be allowed to dress as they like" ist 'non-basic', da auf die Frage "Even if it turned out that mini-skirts cause cancer in the eyes of the beholder?" wohl "not in that case, of course" geantwortet würde.<sup>1</sup> Für uns ist "men and women should be allowed to dress as they like" kein Werturteil, sondern ein normatives, i.a. wohl auch ein normativ-ethisches Urteil.

Normative Urteile dieser Art, 'value judgements' in Sens Terminologie, lassen sich zumeist auf naheliegende Weise in die Behauptung einer moralischen Präferenz über Propositionen übersetzen.<sup>2</sup> Für oben zitiertes Beispiel: "Es ist besser, wenn es Männern und Frauen erlaubt ist, sich so zu kleiden, wie es ihnen gefällt, als wenn das nicht erlaubt ist."

---

<sup>1</sup> Sen (1970), S.60

<sup>2</sup> Die meisten normativen Urteile legen dabei eine einzige propositionale Präferenz fest und sind in diesem Sinne qualitativ:  $p > \sim p$   
Manche normative Urteile sind komparativ: "Die Wirtschaftspolitik sollte eine möglichst niedrige Inflationsrate anstreben"(\*) kann u.U. übersetzt werden in "Eine Wirtschaftspolitik ist um so besser, je niedriger die Inflationrate ist". Widerspricht hier derjenige, der (\*) behauptet hat, ist (\*) vermutlich 'mehrdimensional' zu interpretieren, d.h. die Güte einer Wirtschaftspolitik  $G_w$  ist für ihn dann eine Funktion mehrerer Variablen, z.B. Inflationsrate, Arbeitslosenquote, Außenhandelsbilanz, etc. und mit Inflationsrate  $= x_k$  ist (\*) zu interpretieren als

$$G_w(x, \dots, x_{k-1}, x_k + \Delta x_k, x_{k+1}, \dots, x_n) > G_w(x, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) \Leftrightarrow \Delta x_k < 0;$$

Ein komparatives normatives Urteil legt also in jedem Fall unendlich viele moralische Präferenzen fest. Lassen sich diese moralischen Präferenzen metrisieren, handelt es sich um ein quantitatives normatives Urteil. In diesem Fall können wir jeder Proposition sogar einen numerischen Wert zuordnen.

Vgl. dazu Pflanzagl (1959); und im Kontext einer allgemeinen mathematischen Darstellung der Theorie der reellwertigen ordnungstreuen Funktionen für endliche und unendliche Präferenzensysteme Fishburn (1972) Kap.7. Eine detailliertere Darstellung der präferenztheoretischen Grundlagen geben Luce/Suppes (1963) und Kirman (1981). Als allgemeine Einführung in die Metrisierungstheorie eignet sich Hempel (1952); Suppes/Zinnes (1963); Stegmüller, Bd.2 (1970) Kap.1 und ergänzend Bd.4 (1973), S.234-263.

Sen sieht die 'non-basicness' normativ-ethischer Urteile ausschließlich in ihrer Bedingtheit durch faktische Annahmen. Die getroffene Unterscheidung "lies at the root of the relevance of factual considerations in ethical debates". Die Bedingtheit moralischer Präferenzen kann aber auch in normativen Annahmen bestehen:

A zieht es (moralisch) vor, daß körperliche Züchtigung von Kindern auch für die eigenen Eltern verboten ist. Nun mag es sich herausstellen, daß diese subjektive moralische Präferenz von normativen Annahmen darüber abhängt, was Humanität ausmacht. So daß A diese moralische Präferenz aufgibt, sobald ihm jemand überzeugend darlegt, daß zur Humanität auch eine autoritäre Erziehung gehört.

## 2. Intrinsische Präferenzen

Eine Person, die sich angesichts einer vorgehaltenen Pistole zur Übergabe ihres Geldbeutels gezwungen sieht, würde vermutlich sehr erstaunt sein, von einem in moderner Nutzentheorie geschulten Polizisten zu hören, sie hätte ja nur gemäß ihrer (durch die Umstände bedingten) eigenen Präferenzen gehandelt. Sie wird vermutlich darauf bestehen, daß sie ihren Geldbeutel lieber behalten hätte. Bei hartnäckigem Nachfragen ließe sich vielleicht dennoch Einigkeit darüber herstellen, daß es bei Berücksichtigung der Gesamtsituation tatsächlich in ihrem Interesse lag, den Geldbeutel auszuhändigen, daß dieses Verhalten insofern also im Einklang mit ihren resultierenden Präferenzen gewesen ist. Aber es bleibt dennoch richtig: Sie hätte ihren Geldbeutel lieber behalten.

Wir können schlecht sagen: "Ja, wenn die Umstände anders gewesen wären, hätte sie eine Präferenz gehabt, ihren Geldbeutel zu behalten, aber in diesen Umständen hatte sie eben eine Präferenz ihn herauszugeben". Adäquater ist: "Sie hatte zwar den Wunsch, den Geldbeutel zu behalten, aber einen noch stärkeren Wunsch am Leben zu bleiben".

Solche Wünsche lassen sich als propositionale Präferenzen einer bestimmten Art beschreiben. Wir nennen diese propositionale Präferenzen "intrinsische Präferenzen".<sup>1</sup>

Es entspricht einer intrinsischen Präferenz, den Geldbeutel zu behalten; ebenso entspricht es einer intrinsischen Präferenz am Leben zu bleiben.

Eine Person  $i$  hat eine intrinsische Präferenz für die Proposition  $p$  gegenüber  $p'$  genau dann, wenn  $i$  jeden Weltzustand  $z$ . in dem  $p$  wahr und  $p'$  falsch ist, dem entsprechenden Weltzustand  $z'$  vorzieht, der sich von  $z$  ausschließlich darin unterscheidet, daß  $p$  falsch und  $p'$  wahr ist.

Diese Definition setzt allerdings voraus, daß Propositionen unabhängige Variablen sind. Diese Annahme ist nicht immer unproblematisch.<sup>2</sup> Die weitgehende Idealisierung hat zur Folge, daß die Interpretation ökonomischer und speziell wohlfahrtstheoretischer Modelle in vielen Fällen Schwierigkeiten bereitet.

Was beispielsweise ist unter einem abnehmenden

---

<sup>1</sup> Vgl. K. Baier, *Rationality and morality*, Erkenntnis 11 (1977) 197-223.

Intrinsische Präferenzen dürfen natürlich nicht mit *kategorischen* Präferenzen verwechselt werden: "I am taking all non-conditioned, actual preferences to be preferences come-what-way. That is, I am concerning myself only with *categorical* preferences. A person *categorically* prefers  $p$  to  $q$  ( $p$  und  $q$  sind Propositionen, d.A.), if and only if, for every proposition that is compatible given his information, not with  $p$  and  $q$ , he prefers  $p$ -and- $r$  to  $q$ -and- $r$ " (Schick (1972), S.59). Intrinsische Präferenzen sind *relativ* zu einer gegebenen Situation (zu bestimmten Werten der anderen Zustandsvariablen), während kategorische Präferenzen invariante intrinsische Präferenzen sind.

<sup>2</sup> Ein längerer Urlaub wäre mir lieber - 'ceteris paribus' natürlich und nicht, wenn der längere Urlaub beispielsweise dazu führen würde, daß mein Gehalt gekürzt wird. In diesem Fall hätte ich eine bedingte Präferenz für den kürzeren Urlaub. Dennoch bestünde eine Präferenz für einen längeren Urlaub fort, eben eine *intrinsische Präferenz*. Das Problem steckt dabei in der Interpretation der ceteris-paribus-Klausel, bzw. in der Prämisse der kausalen Unabhängigkeit der infrage stehenden Propositionen. Die im Urlaub

Grenznutzen des Einkommen verstehen? Die individuellen Nutzenfunktionen gehen zunächst über Zustände. Das Einkommen einer Person ist nur ein Merkmal von Zuständen. Es ist nicht anzunehmen, daß eine Person zwischen allen Zuständen, in denen sie ein gleiches Einkommen hat, indifferent ist. Das isolierte Merkmal 'Einkommen' hat andererseits keinen subjektiven Wert.

Nutzenfunktionen einer isolierten Größe (Geld, Güter, Freizeit etc.) sollte man nicht als Metrisierungen propositionaler Präferenzen interpretieren, sondern als Präferenzen über eine Menge von Zuständen, die sich nur in dem betreffenden Merkmal voneinander unterscheiden.

Da es sehr viele mögliche Mengen solcher Zustände gibt, die dabei zugrunde gelegt werden können, und die Gestalt der Bewertungsfunktion der betreffenden Größe i.a. davon abhängen wird, welche dieser Mengen man zugrunde legt, kann man sich das Ergebnis als eine Mittelung der Funktionen über die möglichen Präferenzklassen von Zuständen vorstellen, wobei sich eine solche Mittelung an den subjektiven Wahrscheinlichkeiten der Person orientieren wird.<sup>1</sup>

Die n-dimensionalen Güterräume der ökonomischen Theorie, in denen eine Nutzenfunktion mit n unabhängigen Variablen

---

zusätzlich verbrachte Zeit etwa kann ich nicht mehr herbeizaubern. Fällt sie also nicht unter 'ceteris paribus'? Angenommen, die Antwort lautet 'nein', muß man dann auch den eventuell entstehenden Zeitdruck ausschließen? Dies läßt sich ad infinitum fortsetzen. Andererseits ist der Ausweg, die intrinsische Präferenz als Präferenz über die von den infrage stehenden Propositionen bedingten Wahrscheinlichkeitsverteilungen zu interpretieren, versperrt, da bei dieser Interpretation die Bedeutung von intrinsischer und resultierender Präferenz ununterscheidbar würde.

<sup>1</sup> Hier stoßen wir wieder auf das Reduktionsproblem, das uns schon in §2 beschäftigt hat: Ob die Reduktionsregel notwendigerweise der Entscheidungsregel entsprechen muß, also ob das Akzeptieren des Bayesschen Prinzips als Entscheidungsregel auch verlangt, die Funktionswerte nach den subjektiven Wahrscheinlichkeiten der Zustände zu gewichten, brauchen wir hier nicht zu entscheiden.

definiert ist, geben die Menge der 'ceterorum parium' jeweils an. Wenn alle Zustandsmerkmale im Güterraum eine Dimension erhalten, so wäre dieses Modell nur adäquat, wenn jede Dimension einen intrinsischen (subjektiven) Wert repräsentierte. Handelt es sich jedoch ausschließlich um ökonomische Güter, kann der Nutzen von Güterallokationen als Funktion subjektiver intrinsischer Werte interpretiert werden, die sich mit den Gütern, die in den Dimensionen des Güterraumes repräsentiert sind, verwirklichen lassen und die selbst nicht unter die ceteris-paribus-Klausel fallen, da sie in den Dimensionen des Güterraumes nicht repräsentiert sind.

Auch die Art der Interaktionsstruktur könnte einen Einfluß auf die individuellen resultierenden Präferenzen haben, obwohl sich das nicht in entsprechenden Variablen der subjektiven Nutzenfunktion niederschlägt.

Damit ergeben sich für die spieltheoretische Analyse zwei Möglichkeiten: Die subjektiven Bewertungsfunktionen können als Repräsentation der resultierenden subjektiven Präferenzen interpretiert werden. In diesem Fall muß davon ausgegangen werden, daß die Art der Interaktionsstruktur i. a. einen Einfluß auf die Gestalt der Bewertungsfunktionen hat.

Oder sie werden als Repräsentation der intrinsischen Präferenzen der Personen interpretiert, dann hat die Art der Interaktionssituation zwar auf die Wahrscheinlichkeiten der Zustände aus den Referenzklassen Einfluß, aber die intrinsischen Präferenzen erster Ordnung anderer Beteiligter spielen für die Bewertungsfunktionen ex definitione keine Rolle.

Die spieltheoretische Analyse von Interaktionssituationen würde m.E. durch die Entscheidung für eine dieser beiden Interpretationsmöglichkeiten an Klarheit gewinnen, wobei ich die letztere für erfolgversprechender halte.<sup>1</sup>

---

<sup>1</sup> Im zweiten Teil der Arbeit werden wir einen für die Sozialethik besonders wichtigen Typ von Interaktionssituationen entsprechend analysieren.



### 3. Präferenzen höherer Ordnung

Kehren wir zurück zu unserem Paar aus §7, als es vor der Frage stand, gemeinsam ins Museum oder gemeinsam an den Strand zu gehen. Es könnte sein, daß sie eine intrinsische Präferenz hatte, (gemeinsam) das Museum zu besuchen, und er umgekehrt eine intrinsische Präferenz, sich (gemeinsam) am Strand zu sonnen.

Beide hatten aber wohl auch Präferenzen zweiter Ordnung: Jeder wünschte sich, daß die intrinsischen Präferenzen des anderen vorrangig befriedigt werden. Möglicherweise waren das sogar selbst wiederum intrinsische Präferenzen zweiter Ordnung, d.h. innerhalb der Zustandsmenge, deren Elemente sich nur darin unterscheiden, wessen intrinsische Präferenzen erster Ordnung befriedigt werden und die ansonsten alle wesentlichen Merkmale ihrer tatsächlichen Situation haben, würden sie beide jeweils den Zustand vorziehen, in dem die intrinsischen Präferenzen des anderen befriedigt werden.

Es könnte jedoch auch sein, daß ihre Präferenzen zweiter Ordnung bedingte Präferenzen waren. Vielleicht wollte z.B. in diesem Fall jeder nur deshalb auf die Wünsche des anderen eingehen, um das nächste Mal, wenn es um wichtigere Entscheidungen geht, ein Argument für die Berücksichtigung der eigenen (intrinsischen) Präferenzen erster Ordnung zu haben.

Präferenzen höherer Ordnung beziehen sich auf Präferenzen niedrigerer Ordnung: Präferenzen zweiter Ordnung auf solche erster Ordnung, Präferenzen  $n$ -ter Ordnung auf solche  $(n-1)$ -ter Ordnung. Es gibt intrinsische Präferenzen  $n$ -ter Ordnung, die sich auf bedingte Präferenzen  $(n-1)$ -ter Ordnung beziehen, es gibt bedingte Präferenzen, die sich auf intrinsische Präferenzen beziehen, es gibt intrinsische Präferenzen, die sich auf intrinsische Präferenzen beziehen, und es gibt bedingte Präferenzen, die sich auf bedingte Präferenzen beziehen.

Die ursprüngliche Dreiteilung ist mit dieser Erweiterung natürlich keineswegs obsolet geworden.

Wir haben jetzt intrinsische und bedingte Wohlfahrtspräferenzen, Wohlfahrtspräferenzen erster und höherer Ordnung.

Wir haben intrinsische und bedingte moralische Präferenzen. Unser Demokrat, der diese Erweiterungen mit motiviert hat, hat, so wollen wir annehmen, eine intrinsische moralische Präferenz für  $z$  und eine bedingte moralische Präferenz für  $z'$ . Seine resultierende Präferenz orientiert sich an der bedingten moralischen Präferenz für  $z'$ .

## 10. MORAL UND NUTZEN

*"There seems to be no natural way to introduce the concept of obligation into a complete decision theory, fully equipped with a numerical subjective probability function and the single rule of behaviour, always to maximize expected utility."*

Suppes (1973)

Scheinbar gibt es eine wesentlich elementarere Methode, Grundbegriffe der modernen Entscheidungstheorie so zu konzipieren, daß sie mit Grundbegriffen der Ethik nicht in Konflikt kommen. Man geht von zwei strikt getrennten Arten von Handlungsanleitungen aus: solchen, die auf generellen Verpflichtungsurteilen und solchen, die auf generellen Rationalitätsbehauptungen gründen.

Diese einfache Zweiteilung kann für sich in Anspruch nehmen, der alltäglichen Unterscheidung zwischen Situationen, in denen man zu etwas verpflichtet ist und Situationen, in denen man frei nach persönlichen Interessen entscheiden kann, zu entsprechen.

Der Begriff der Verpflichtung ließe sich dabei im Rahmen einer solchen Dichotomie z.B. auf folgende Weise axiomatisieren<sup>1</sup>: Wenn  $\Xi$  die Menge der Alternativen ist - hier in einem allgemeinen Sinn: vollständige Zustandsbeschreibungen umfassen gesellschaftliche und natürliche Merkmale -, so können wir eine Ordnungsrelation  $\succ_{\rho}$  auf  $\Xi$  definieren, die den Wahrscheinlichkeitsaxiomen genügt und die Zustände nach ihrer subjektiven Wahrscheinlichkeit ordnet, und eine zweite Ordnungsrelation  $\succ$ , die auf der Menge der Entscheidungen  $\Sigma^1$  die Präferenzordnung aufgrund

---

<sup>1</sup> Wir folgen dabei im wesentlichen P. Suppes, *The concept of obligation in the context of decision theory*, (1973). Um der Klarheit willen werden wir die Theorie jedoch in einer modifizierten Fassung darstellen.

moralischer Verpflichtung oder individueller Rationalität darstellt. Die Entscheidungen  $\sigma$  aus  $\Sigma^i$ , zu denen man moralisch verpflichtet ist, bilden die Klasse  $V_i$ .<sup>1</sup>

Da jedoch eine Entscheidung sinnvollerweise als Funktion über ganz  $\Xi$  zu definieren ist, müssen wir die Begriffe noch leicht modifizieren.

Zu einer Entscheidung ist man ja im allgemeinen nicht generell verpflichtet (unabhängig von der Wahrscheinlichkeitsverteilung über  $\Xi$ ), sondern nur im Fall des Eintretens eines bestimmten Ereignisses, oder bei Vorliegen einer subjektiven Wahrscheinlichkeit für dieses Ereignis, die größer als eine bestimmte - von Fall zu Fall verschiedene - Mindestwahrscheinlichkeit ist. Ich bin z.B. moralisch dazu verpflichtet, einen Mann, der auf einer Brücke steht und den Eindruck erweckt, Selbstmord begehen zu wollen, davon abzuhalten zu versuchen, (etwa indem ich ihn anspreche); ich bin aber nicht verpflichtet, jeden Passanten, der sich auf einer Brücke befindet, anzusprechen, um ihn von einem Selbstmord abzuhalten, obwohl natürlich die Wahrscheinlichkeit größer als 0 ist, daß einer sich mit diesem Gedanken trägt. Wenn wir Entscheidungen als Funktionen über  $\Xi$  definieren, so ist "Suche eine Person, die sich auf einer Brücke befindet und den Eindruck erweckt, Selbstmord begehen zu wollen, davon abzuhalten" eine Verpflichtung zu einer partiellen Entscheidung, insofern als sie nur im Falle eines bestimmten Ereignisses  $\epsilon$  eine Verhaltensvorschrift macht, für alle anderen Situationen jedoch keine Verhaltensempfehlung gibt. Eine Verpflichtung  $v$  legt die Funktion  $\sigma$  also nicht in ihrem gesamten Definitionsbereich  $\Xi$ , sondern nur in einem eingeschränkten Bereich  $\epsilon$  fest. Die Verpflichtung  $v$  ist also eine bestimmte partielle Entscheidung  $\sigma|_{\epsilon}$ . Ein Ereignis ist in diesem Modell eine

---

<sup>1</sup> Bei Suppes gibt es keine Spezifizierung von  $v$  auf  $i$ . Da jedoch anzunehmen ist, daß  $v$  nicht allgemein, sondern nur relativ zu einer bestimmten Person bestimmt werden kann, (ebenso wie die subjektive Wahrscheinlichkeit) wird das wohl stillschweigend vorausgesetzt.

Teilmenge  $\epsilon$  von  $\Xi$ , wobei  $\epsilon$  in unserem Fall alle Zustände  $\Xi$  enthält, in denen jemand auf einer Brücke den Anschein erweckt, Selbstmord begehen zu wollen.  $V$ , die Menge der Entscheidungen, zu denen man verpflichtet ist, enthält damit als Elemente partielle Entscheidungen. Ohne Zweifel ist diese Wortwahl nicht ganz befriedigend. Einerseits würde man alltagssprachlich für unser Beispiel nicht sagen, man sei zu der Entscheidung, besagte Person vom Selbstmord abzuhalten, zu suchen verpflichtet, sondern man würde sagen, man sei zu der entsprechenden Handlung verpflichtet. Andererseits ist die alltagssprachliche Beziehung von "Handlung" und "Entscheidung" ähnlich der von "Ereignis" und "Zustand". Ein Ereignis  $\epsilon$  ist eine Teilmenge  $\Xi^0$  von  $\Xi$ . Das kann so interpretiert werden, daß die propositionalen Beschreibungen der Zustände  $\xi$  aus  $\Xi^0$  das Ereignis jeweils enthalten. Da sich auf diese Weise jedem Ereignis genau eine Teilmenge von  $\Xi$  zuordnen läßt, können wir einen Grundbegriff einsparen, indem wir Ereignisse mit Teilmengen von  $\Xi$  identifizieren.

In der gleichen Weise könnte man Handlungen mit Teilmengen von  $\Sigma^1$  identifizieren.

Verpflichtungen wären dann nicht mehr partielle Entscheidungen sondern Mengen von Entscheidungen.

Die Begriffe der Theorie unterscheiden sich etwas von den in der Entscheidungstheorie üblichen. Da  $\Xi$  die Menge der Zustände ist, und die Menge der Entscheidungen  $\Sigma$  aus Funktionen mit dem Definitionsbereich  $\Xi$  besteht, könnte die Menge der Konsequenzen als Vereinigungsmenge der Wertebereiche dieser Funktionen definiert werden. Darauf wird jedoch verzichtet. Die Präferenzordnung  $>$  ist ausschließlich auf der Menge der Entscheidungen definiert,  $> \subset \Sigma \times \Sigma$ , die Ordnung nach subjektiver Wahrscheinlichkeit auf der Menge der Zustände.  $V$  ist die Menge der Verpflichtungen, deren Elemente partielle Entscheidungen sind.

Die Grundbegriffe der Theorie:

$\Xi$  : Menge der Zustände

$\Sigma$  : Menge der Entscheidungen  $\sigma$

(Def.)  $E := P(\Xi)$  : Die Menge der Ereignisse ist die Potenzmenge von  $\Xi$

$\succ$  : Präferenzordnung auf  $\Sigma^1$ ,  $\succ \subset \Sigma \times \Sigma$

$\succ_P$  : Ordnung subjektiver Wahrscheinlichkeit auf  $\Xi$ ,  
 $\succ_P \subset \Xi \times \Xi$

$V$  : Menge der Verpflichtungen (ist eine Teilmenge der Menge partieller Entscheidungen)

$v|_{\epsilon}$  : Verpflichtung bei Eintreten des Ereignisses  
 $\epsilon \in E$ ,  $v|_{\epsilon} \in V$

Def.: Maximaler Verpflichtungsbereich einer Entscheidung  $\sigma$  (MVB( $\sigma$ ))

$\epsilon_1 = \text{MVB}(v) \Leftrightarrow v|_{\epsilon_1} \in V \wedge \forall \epsilon_2 \in E:$

$[\epsilon_1 \subset \epsilon_2 \wedge \epsilon_1 \neq \epsilon_2 \Rightarrow \sim v|_{\epsilon_2} \in V]$

Die Axiome sollen vor allem darstellen, wie sich aus gegebenen Verpflichtungen neue ergeben, welche Präferenzordnung sich aus der Menge der Verpflichtungen ergibt, und in welcher Beziehung Verpflichtungen zu anderen Entscheidungen stehen.

Die Axiome der Theorie:

A1 :  $\epsilon_1 = \text{MVB}(v) \wedge \epsilon_2 = \text{MVB}(v') \wedge \sim(\epsilon_1 = \phi \wedge \epsilon_2 = \phi) \Rightarrow$   
 $\Rightarrow [v \succ v' \Leftrightarrow \epsilon_1 \succ_P \epsilon_2]$  <sup>2</sup>

A2 :  $\forall \epsilon \in E, \epsilon \succ_P \phi$

A3 :  $\sim \phi \succ_P \Xi$

A4 :  $\forall \epsilon_1, \epsilon_2, \epsilon_3 \in E,$

$\epsilon_1 \cap \epsilon_3 = \phi \wedge \epsilon_2 \cap \epsilon_3 = \phi \Rightarrow \epsilon_1 \succ_P \epsilon_2 \Leftrightarrow \epsilon_1 \cup \epsilon_3 \succ_P \epsilon_2 \cup \epsilon_3$

<sup>1</sup> Wir wollen annehmen, daß die Präferenzrelation  $\succ$  eine (konnexe) Ordnungsrelation ist.

<sup>2</sup>  $\sim(\sigma_1 = \phi \wedge \sigma_2 = \phi)$  deshalb, weil ansonsten für nicht-verpflichtende Ereignisse Indifferenz gelten würde.

$$A5 : \varepsilon_1 \subset \varepsilon_2 \wedge v|_{\varepsilon_2} \in V \Rightarrow v|_{\varepsilon_1} \in V$$

$$A6 : \varepsilon_1 \cap \varepsilon_2 = \emptyset \wedge v|_{\varepsilon_1} \in V \wedge v|_{\varepsilon_2} \in V \Rightarrow v|_{\varepsilon_1 \cup \varepsilon_2} \in V$$

$$A7 : \varepsilon_1 \cap \varepsilon_2 = \emptyset \wedge v|_{\varepsilon_1} \in V \wedge v|_{\varepsilon_2} \in V \Rightarrow v|_{\varepsilon_1 \cap \varepsilon_2} = v'|_{\varepsilon_1 \cap \varepsilon_2}$$

A1 bestimmt die Präferenzordnung auf der Menge der Verpflichtungen.

A2, A3, A4 entsprechen üblichen Glaubensaxiomen.

A5, A6, A7 können als Postulate der Verpflichtungskonsistenz interpretiert werden.

Die Beziehung von Verpflichtungen zu anderen Entscheidungen ist dadurch bestimmt, daß dieses Axiomensystem als "Einschränkung" der Entscheidungsregeln der rationalen Entscheidungstheorie gedacht ist. Entscheidungssituationen, in denen sich keine Verpflichtung ergibt, erfordern daher die Anwendung des üblichen Kriteriums der NEW-Maximierung.

A1 beinhaltet zusammen mit A2 eine lexikographische Entscheidungsregel. Für eine Situation ohne Verpflichtungsgehalt ist MVB gleich der leeren Menge. Mit A2 wird damit jegliche Verpflichtung einer Entscheidung unter anderen Kriterien, etwa dem des persönlichen Interesses, vorgezogen. Kriterien moralischer Entscheidung sind Kriterien rationaler Entscheidung vorgeordnet. Die Theorie legt zwei strikt getrennte Typen von Handlungsmaximen durch die formale Struktur des Axiomensystems fest. In der Wahl der jeweiligen Handlungsmaximen ist man noch völlig frei. Jedoch legt A1 eine Beziehung zwischen Maximen der Verpflichtung und solchen der Rationalität fest.

Die Handlungspräferenzen haben drei Eigenschaften: Die Präferenzrelationen bezüglich unterschiedlicher unvereinbarer Verpflichtungen werden allein von den Wahrscheinlichkeiten der bedingenden Ereignisse bestimmt, die Verpflichtungen sind anderen Handlungsmöglichkeiten lexikographisch vorgeordnet (d.h. erst wenn der Handelnde zu nichts mehr verpflichtet ist, stehen ihm andere Hand-

lungsalternativen offen) und die Präferenzrelationen bezüglich Handlungsalternativen, die allesamt keinen Verpflichtungscharakter haben, sind bestimmt vom Erwartungswert-Maximierungskriterium.

Die Interpretation der Verpflichtungsaxiomatik A1...A7 läßt Axiom A5 (wenn man zu einer Entscheidung  $\sigma$  bei Ereignis  $\varepsilon_2$  verpflichtet ist, so ist man dazu auch dann verpflichtet, wenn es sich um einen Spezialfall dieses Ereignisses  $\varepsilon_1$  handelt) und Axiom A6 (wenn zwei Ereignisse, die zu  $v$  bzw.  $v'$  verpflichten, nicht zugleich auftreten können, ist man zu beiden verpflichtet) als intuitiv unmittelbar einleuchtend erscheinen.<sup>1</sup>

Das Axiom A7 hingegen expliziert den Verpflichtungsbegriff auf eine Weise, die dem normalen, aber auch dem philosophischen Sprachgebrauch nicht entspricht. A7 schließt nämlich einen Konflikt zweier oder mehrerer Verpflichtungen logisch aus. Das System der Verpflichtungen ist im Einklang mit dieser Begriffsfestlegung so zu konstruieren, daß zusammentreffende Ereignisse in keinem Fall zu konkurrierenden Verhaltensempfehlungen führen. Prima-facie-Verpflichtungen, zwischen denen im Konfliktfall Vorrangregeln oder Willkür entscheiden muß, sind also in der Menge der Verpflichtungen nicht enthalten. Die Tragweite dieser Eigenschaft von Suppes' Verpflichtungsbegriff kann man sich damit klarmachen, daß es mit A7 den Fall nicht geben kann, in dem eine Person A eine (z.B. rechtliche) Verpflichtung  $v$  gegenüber einer Person B hat und eine weitere Verpflichtung  $v'$  gegenüber einer Person C, denen A - aus welchen Gründen auch immer - nicht gleichzeitig nachkommen kann. Mindestens eine dieser Verpflichtungen (im üblichen Sinne) ist mit A7 keine Verpflichtung im Sinne dieser Axiomatik.

---

<sup>1</sup> Das ist natürlich keine Adäquatheitsbedingung für diese Axiome, wie es die klassische Theorie der Axiomatik annahm.



Die Vorrangregel A1 für Verpflichtungen gilt nur für den Fall, in dem die die Verpflichtung bedingenden Ereignisse noch nicht eingetreten sind, oder deren Eintreten dem Handelnden noch nicht bekannt sind. Dann regelt die Wahrscheinlichkeitsabschätzung des Handelnden für diese Ereignisse, welche Verpflichtung Vorrang hat, und zwar unabhängig von einer etwaigen 'Schwere' der Verpflichtung. Es ist eine Schwäche dieser Axiomatik, daß sie, außer dem Grad der subjektiven Wahrscheinlichkeit, keine Möglichkeit der Gewichtung von Verpflichtungen kennt.

Während A7 nur eine gewisse Revision der gewohnten Bedeutung von "Verpflichtung" mit sich bringt, scheint mir A1 inakzeptabel, da es den normativen Theorien, die den Verpflichtungsbegriff dieser Axiomatik zugrundelegen, eine inhaltliche Beschränkung auferlegt, die sich wohl nur durch die Einführung eines zusätzlichen normativen Prädikats, mit all den damit zu erwartenden begrifflichen Schwierigkeiten, beheben ließe. Denn A1 induziert eine 'normative Gleichgewichtigkeit' unterschiedlicher Verpflichtungen. Es macht die Verpflichtung, z.B. das Leben eines Kindes (unter bestimmten Umständen) zu retten<sup>1</sup>, 'gleichgewichtig' der Verpflichtung, seine Parkgebühren (unter gewissen anderen Umständen) zu bezahlen. Das gilt deshalb, weil für den Fall, daß dem Entscheidenden nur die Wahrscheinlichkeiten für das Eintreten der (die Verpflichtungen bedingenden) Ereignisse bekannt sind<sup>2</sup>, diejenige Verpflichtung zu erfüllen ist, deren bedingendes Ereignis wahrscheinlicher ist (A1). Diese Vorrang-

---

<sup>1</sup> "Unter bestimmten Umständen" wird deshalb angefügt, weil A7 keine prima facie Pflichten zuläßt, so daß für das Eintreten beider Ereignisse eine einzige Verpflichtung - sei es nun eine von diesen beiden oder eine dritte - das Handeln leitet.

<sup>2</sup> Etwa die Wahrscheinlichkeit dafür, daß das Kind unter ein Auto kommt und die Wahrscheinlichkeit dafür, daß auch feiertags Parkgebühren zu zahlen sind.

regel bei konkurrierenden Verpflichtungen ist nicht überzeugend, da in ihr die 'Schwere' der Verpflichtung keine Rolle spielt, wobei "Schwere der Verpflichtung" erst noch expliziert werden müßte.

Angenommen, eine solche Explikation sei gelungen (und das ist unabhängig von einer konkreten Theorie normativer Ethik möglich), so würde sich mit einer solchen Modifikation auch die Frage stellen, ob eine adäquatere Vorrangregel nicht das, für die Entwicklung einer konkreten normativen Theorie äußerst restriktive Axiom A7 überflüssig machen würde.

Die Dichotomie in "Verpflichtung" auf der einen und "rationales Verhalten im Sinne der Entscheidungstheorie" auf der anderen Seite, überzeugt in dieser Form nicht. Der gordische Knoten, der Rationalität und Moralität verknüpft, scheint mir mit diesem Ansatz nicht lösbar zu sein. Im Rahmen dieses Verpflichtungsbegriffes wäre es auch in Entscheidungssituationen ohne Interaktion nicht sinnvoll, pflichtgemäßes Handeln mithilfe der Maximierung des Erwartungswertes der Wünschbarkeit (oder des Nutzens) festzustellen. Das steht im Gegensatz zu dem Anspruch, die rationale Entscheidungstheorie sei als generelle Theorie rationalen Verhaltens, u.a. rational moralischen Verhaltens, konzipiert. Gemäß A1 ergibt sich die Präferenzordnung auf der Menge der Verpflichtungen allein aufgrund der Wahrscheinlichkeiten der Ereignisse, die jeweils die Verpflichtungen bedingen.

Es ist nicht einsichtig, warum es nicht möglich sein sollte, auch die Verpflichtungen zu gewichten (in ähnlicher Weise wie auch die Konsequenzen nach ihrer subjektiven Wünschbarkeit gewichtet werden).

Eine rationale moralische Entscheidung in einer Entscheidungssituation ohne Interaktion maximiert den Erwartungswert dieser Gewichtung bei der Anwendung der

Bayesschen Regel rationalen Verhaltens in Risikosituationen bzw. den Wert, den eine Form der Regel  $\Lambda^1$  einer Lotterie von Verpflichtungen zuweist (oder was immer sich als Kriterium rationalen Verhaltens in Risikosituationen ohne Interaktion als adäquat erweist).

Damit läge eine einheitliche Theorie rationalen Verhaltens in Risikosituationen für moralische und außer-moralische Entscheidungen vor.

Die v. Neumann-Morgenstern-Axiomatik hat tatsächlich auch dann eine intuitiv überzeugende Interpretation, wenn wir sie auf die Präferenzordnung  $\succ$  der Verpflichtungen anwenden.<sup>2</sup>

(1)  $\succ$  ist konnexe Ordnungsrelation auf  $V$

(2a)  $v \prec v' \Rightarrow v \prec a \cdot v + (1-a) \cdot v'$ ,  $a \in (0,1)$ ;

wobei:  $v \prec w \Leftrightarrow v \prec w \wedge \sim w \prec v$ ;

$v = w \Leftrightarrow v \prec w \wedge w \prec v$ ;

(2b)  $v \succ v' \Rightarrow v \succ a \cdot v + (1-a) \cdot v'$ ,  $a \in (0,1)$ ;

(2c)  $v_1 \prec v_2 \prec v_3 \Rightarrow \exists a \in (0,1): a \cdot v_1 + (1-a) \cdot v_3 = v_2$ ;

Die algebraischen Axiome gelten natürlich auch hier.

Für " $w \prec a \cdot v + (1-a) \cdot v'$ " gibt es eine naheliegende Interpretation.  $v$ ,  $v'$  und  $w$  sind Verpflichtungen. Es wird dem Entscheidungsträger eine Lotterie angeboten, von deren Ergebnis es abhängt, welche Verpflichtung zu erfüllen dem Entscheidungsträger unmöglich gemacht wird. Für  $a=0,2$  muß er also mit 80 % Wahrscheinlichkeit damit rechnen, daß er der Verpflichtung  $v$  nicht nachkommen kann, sondern nur der Verpflichtung  $v'$ , und mit 20 % Wahrscheinlichkeit damit, daß er der Verpflichtung  $v'$  nicht nachkommen kann, sondern nur der Verpflichtung  $v$ . Der Entscheidungsträger hat dann zwischen der Lotterie  $a \cdot v + (1-a) \cdot v'$  und der (sicheren) Möglichkeit, der Pflicht  $w$  nachzukommen, zu wählen. Je nachdem, wie diese

---

<sup>1</sup> Vgl. §2

<sup>2</sup> Vgl. v. Neumann/Morgenstern (1973), Abs.3.6., S. 26-31 und Anhang S.642-657. Unterdessen gibt es auch einige elegantere und in Nuancen abweichende Axiomensysteme zur Metrisierung einer Präferenzrelation. S. dazu Schneeweiß (1963) und Gottinger (1974), S.30-37.

Wahl ausfällt, gilt:

- $w < a \cdot v + (1-a) \cdot v'$
- $w = a \cdot v + (1-a) \cdot v'$
- $w > a \cdot v + (1-a) \cdot v'$ .

Damit außermoralische Gesichtspunkte (welche Pflichterfüllung ist am angenehmsten, mit den geringsten Mühen verbunden etc.) diese moralische Präferenzrelation auf der Menge der Verpflichtungen nicht verzerren, sollte diese Wahl nicht behavioristisch als die jeweils vollzogene, auch nicht als das Ergebnis einer - mehr oder weniger ehrlichen - Selbsteinschätzung interpretiert werden, sondern ausdrücklich als die Verpflichtungspräferenz unter moralischen Gesichtspunkten, also etwa als Antwort auf die Frage "Welche Wahl, meinen Sie, sollte man in einer solchen (fiktiven) Situation treffen?".

Daß es sich hier um Verpflichtungen handelt, hindert uns also nicht, eine numerische Bewertung vorzunehmen. Hier erschiene es allerdings noch weniger sinnvoll als im üblichen Kontext der 'tatsächlichen individuellen Präferenzen', von einer "Nutzenfunktion" (über  $V$ ) zu sprechen.

Nach  $A_1$  hätte sich die moralische Präferenz zwischen Verpflichtungen auf folgende Weise ergeben:

$$v|_{\epsilon_1} > v'|_{\epsilon_2} \Leftrightarrow \epsilon_1 > \frac{\rho}{\rho} \epsilon_2$$

Das ist sicher inadäquat. Im Fall einer Kollision sich gegenseitig pragmatisch ausschließender Verpflichtungen kann i.a. eine Rangordnung unter den Verpflichtungen angenommen werden. So habe ich die Verpflichtung, höflich zu sein, aber auch die gewichtigere Verpflichtung, jemanden davor zu bewahren, unbedacht die Straße zu betreten und sich dabei zu gefährden. Es ist keine Frage ausschließlich der Wahrscheinlichkeiten der bedingenden Ereignisse, sondern auch eine Frage des moralischen Gewichts einer Verpflichtung, die darüber entscheidet, ob ich ihr nachkommen muß.

Es ist daher zweckmäßig, auch auf der Menge der Verpflichtungen eine numerische Bewertungsfunktion zu definieren.

Konsistentes Wahlverhalten bei Lotterien von Verpflichtungen erlaubt es, je individuell die subjektive Gewichtung als kardinale Bewertungsfunktion zu bestimmen. Rationales moralisches Verhalten kann damit relativ zu dieser subjektiven Bewertungsfunktion bestimmt werden. Welche der subjektiven Gewichtungen moralisch geboten ist, kann erst eine konkrete Theorie normativer Ethik bestimmen.

Der 'abstrakte Nutzen' der Entscheidungs- und Spieltheorie ist nichts anderes (oder ist zumindest als nichts anderes intendiert) als die Metrisierung der resultierenden Präferenzen. Moralisten hoffen, daß die jeweiligen individuellen Nutzenfunktionen in einem hohen Grad moralische Präferenzen repräsentieren. Jedenfalls ist diese Art Nutzenfunktion kein Darstellungsmittel persönlicher Interessen und taugt damit allerdings auch nicht zu einer Präzisierung "individuell rationalen Verhaltens" im Sinne von "den persönlichen Interessen optimal entsprechendem Verhalten".

Suppes' Dichotomie wird dem nicht gerecht:

"Utility and desirability would enter only in choosing between two obligatory acts whose conditional events had equal probability of occurring"<sup>1</sup>

setzt implizit einen anderen Nutzenbegriff voraus, der in irgendeiner Weise die persönlichen Interessen des Entscheidungsträgers wiedergibt, die dann konsequenterweise nur in einem Sonderfall eine Rolle bei der Wahl zwischen zwei Verpflichtungen spielen dürfen, nämlich falls unter normativen Aspekten Indifferenz besteht, und das heißt für Suppes' Axiomatik, falls die (die Verpflichtungen bedingenden) Ereignisse gleich wahrscheinlich sind.

Auch moralisch motiviertes Verhalten ist wünschbar, auch unter Verpflichtungen kann es moralisch motivierte Präferenzen geben, auch diese Präferenzen sollten minimalen Rationalitätsbedingungen genügen und daher metrisierbar sein. Suppes' Dichotomie in Verpflichtung und

---

<sup>1</sup> Suppes (1973), S.5

Rationalität, allgemeiner: in moralisches und rationales Verhalten, schränkt an einer Stelle den allgemeinen Anspruch der rationalen Entscheidungstheorie ein, an der das nicht nötig und nicht zweckmäßig ist.

C KOLLEKTIVE RATIONALITÄT UND MORALITÄT

- § 11 Zur Interpretation normativer Theorien  
der Interessenaggregation: S. 149
- § 12 Rationalität und Moralität als  
konkurrierende Prinzipien: S. 166
- § 13 Begriffsbestimmung: Moralisches System: S. 199
- § 14 Interaktionsituationen vom PD-Typ
- § 15 Begriffsbestimmung: Kollektive  
Kollektive Rationalität: S. 218
- § 16 Moralität als Metapräferenztheorie: S. 223
- § 17 Kollektive Rationalität als Minimalmoral: S. 238

## 11. ZUR INTERPRETATION NORMATIVER THEORIEN DER INTERESSENAGGREGATION:

- AXIOLOGISCHE UND NICHT-AXIOLOGISCHE ETHIK
- ETHIK ERSTER UND ZWEITER ORDNUNG

### VORBEMERKUNG:

Eine ganze Reihe von Fragestellungen der Moralphilosophie, u.a. das Problem der sozialen Gerechtigkeit, des fairen Interessenausgleichs, der Berücksichtigung persönlicher Leistung, der Beziehung von Gemeinwohl und Partikularinteressen etc. kann unter Verwendung der analytischen Werkzeuge, die uns die Theorien der sozialen Entscheidung und die Sozialwahltheorie zur Verfügung stellen, neu - und in vielen Fällen präziser - diskutiert und wohl auch das eine oder andere Mal einer definitiven Lösung zugeführt werden, die ohne diese Mittel nicht erreichbar gewesen wäre. Eine gegenseitige Befruchtung der traditionellen Methoden der Moralphilosophie und der Methoden im Umfeld der rationalen Entscheidungstheorie und politischen Ökonomie setzt allerdings eine besondere Sorgfalt metatheoretischer Klärung voraus. Das betrifft den Status der Theorien rationaler Entscheidung, das Verständnis ihrer Grundbegriffe (wie z.B. "individuelle Präferenz", "kollektive Präferenz" etc.) und die Interpretation normativer Theorien der Aggregation individueller Präferenzen.

Während wir im ersten Teil der Arbeit den Grundbegriff der individuellen Präferenz eingehend untersucht haben, sollten wir einige Vorüberlegungen zur Interpretation normativer Theorien der Aggregation individueller Präferenzen an den Anfang des zweiten Teils stellen, der sich schwerpunktmäßig mit dem Zusammenhang individueller und kollektiver Rationalität einerseits und kollektiver Rationalität und Moralität andererseits befaßt.



Der folgende Abschnitt erfüllt zwei Aufgaben. Erstens soll er die Grenzen normativer Theorien der Interessenaggregation abstecken und zum zweiten soll er einen Beitrag zur Interpretation dieser Theorien leisten, wobei die vorgeschlagene Interpretation einerseits einige Modifikationen für den Ansatz der *collective-choice*-Theorie nach sich zieht, (welche die Begriffsdifferenzierungen des ersten Teils voraussetzen) und andererseits eine wichtige metaethische Implikation hat, die das Begründbarkeitsproblem sozialethischer Theorien dieses Typs abschwächt, wenn nicht behebt.

Eine normativ-ethische Theorie muß nicht notwendigerweise eine axiologische Theorie sein. Eine Theorie, die sagt, was man tun sollte, muß keine Annahmen über Werte machen, muß nicht bestimmte Zielsetzungen menschlichen Verhaltens als gute und andere als schlechte auszeichnen. Unter den normativen Theorien gibt es ästhetische, ethische und andere Theorien<sup>1</sup>, und unter den normativen Theorien, die ethische Theorien sind (und damit Gegenstand der Moralphilosophie), gibt es axiologische und

---

<sup>1</sup> Auch manche Beiträge der Wissenschaftstheorie sind *normativ* zu interpretieren. Die Konzeption des Kritischen Rationalismus z.B. beansprucht weniger einen rekonstruktiven oder deskriptiven, sondern in erster Linie einen normativen Status. Vgl. Poppers *Logik der Forschung*, §10 und §11; vgl.a. Albert (1971), (1972) und Aldrup (1971). Ebenso kann nicht nur die Methodologie empirischer Wissenschaften und empirischen Alltagswissens, sondern auch die Methodologie normativer Wissenschaft und moralischer Alltagsurteile selbst einen normativen Status beanspruchen, was manchen Streit um die 'Einheit der Ethik' überflüssig erscheinen läßt (vgl. Pieper (1971). Da die Metaethik als eigene philosophische Disziplin in engem philosophiehistorischem Zusammenhang mit der sprachanalytischen Philosophie der "ordinary language" steht (vgl. z.B. die Aufsatzsammlung von Meggle und Grewendorf (1974), vgl.a. Albert (1961), Inciarte (1972), Riedel (1979) S.5-16), wird dieser Standpunkt nur selten vertreten, jedoch z.B. explizit von Frankena, *On saying the ethical thing*, Proceedings and progresses of the american philosophical association 39 (1966) 21-42.

nicht-axiologische. Axiologische wie nicht-axiologische ethische Theorien machen, im Gegensatz zu anderen normativen Theorien, zumindest implizit Aussagen darüber, was zu tun ist. Der präskriptive Charakter soll uns daher als 'differentia specifica' normativer ethischer Theorien gelten.

Die Äußerung "dort ist ein Stier" ist wahr oder falsch. Derjenige, der diese Äußerung macht, kann mit dieser Äußerung zugleich den illokutionären Akt des Warnens vollziehen; er könnte das explizit machen durch "ich warne dich, dort ist ein Stier" oder "ich warne dich vor dem Stier dort" - diese Äußerung ist nicht mehr wahr oder falsch.<sup>1</sup> In dem Sinne, in dem "dort ist ein Stier" 'implizit' eine Warnung sein kann, geben axiologische ethische Theorien eine Handlungsempfehlung. Insofern unterscheiden sich moralische Wertaussagen von nicht-moralischen durch ihren präskriptiven Charakter.

Baier hat den Unterschied von performativen Äußerungen der Empfehlungen ("commendation") und konstativen (wahren oder falschen) Äußerungen der Bewertung ("praising") herausgearbeitet.<sup>2</sup> Aber diese und ähnliche Untersuchungen können Kriterien für den präskriptiven Charakter einer Äußerung allein aufgrund sprachlicher Merkmale nicht bereitstellen. Der Unterschied einer Beurteilung 'innerhalb' und 'außerhalb' einer Institution<sup>3</sup> schlägt sich in der sprachlichen Form allein nicht notwendigerweise nieder.

---

<sup>1</sup> Vgl. J.L. Austin, *How to do things with words*, (1962), dt.: *Zur Theorie der Sprechakte*, Stuttgart 1972

<sup>2</sup> Vgl. K. Baier, *The moral point of view: a rational basis of ethics*, Ithaka, N.Y. 1958, S.77-84.

<sup>3</sup> Vgl. J.L. Mackie, *Ethics*, (1977), Kap.3

Der Zusammenhang von präskriptiven Äußerungen und moralischen Wertaussagen ist allerdings komplizierter als man zunächst meinen könnte. Eine ethische Theorie über gerechte Verteilungen gibt nicht notwendigerweise individuelle Handlungsempfehlungen. Man kann sich die Beziehung von präskriptivem und axiologischem Gehalt einer normativ-ethischen Theorie ähnlich der Beziehung von individueller Nutzenfunktion (über die Menge der Konsequenzen) und Präferenzrelation über die Menge der Entscheidungen in der Entscheidungs- und Spieltheorie vorstellen. Dabei ist interessant, daß die Beziehung präskriptiv - axiologisch von den meisten moralphilosophischen Theorien kaum thematisiert wird; vermutlich, weil sie fälschlicherweise als unproblematisch empfunden wird, als wüßte man, welches Verhalten moralisch geboten ist, sobald die richtigen Werturteile bekannt sind.<sup>1</sup> Daß das nicht richtig ist, wird sich im Fortgang der Arbeit zeigen.

Diese Unterscheidung in axiologische und nicht-axiologische normativ-ethische Urteile ist für die analytische Metaethik nichts Neues. Frankena<sup>2</sup> unterscheidet moralische Verpflichtungsurteile von moralischen Werturteilen. Gegenstand von Verpflichtungsurteilen sind Handlungen, Gegenstand von Werturteilen sind Personen, Motive, Ziele etc..<sup>3</sup> Das typische Prädikat eines Werturteils (oder einer Wertaussage - um von der 'juristischen' Terminologie

---

<sup>1</sup> Anders Frankena, der von sich selbst sogar sagt: "my position during this first period, held very tentatively, was a cognitivist one combining naturalism about 'good' and intuitionism about 'ought'", Frankena (1976), S.209.

<sup>2</sup> Vgl. A.K. Frankena, *Ethics*, Englewood Cliffs, N.J. 1963, Kap.1.

<sup>3</sup> Was alles Gegenstand eines moralischen Werturteils sein kann, ist natürlich zwischen den ethischen 'Schulen' umstritten. Für Kant z.B. "ist überall nichts in der Welt, ja überhaupt auch außer derselben zu denken möglich, was ohne Einschränkung für gut könnte gehalten werden, als allein ein guter Wille", so im ersten Satz des ersten Abschnitts der *Grundlegung zur Metaphysik der Sitten* (1785) unter der Überschrift "Übergang von der gemeinen sittlichen Vernunftkenntnis zur philosophischen".

in der Ethik wegzukommen) ist "gut" - "ein guter Charakter", "eine gute Absicht" etc., das typische Prädikat eines Verpflichtungsurteils ist "richtig" - "eine richtige Entscheidung", "eine richtige Tat" etc..

Wertaussagen und Verpflichtungsaussagen sind moralische Urteile. Neben den Verpflichtungsurteilen gibt es weitere Arten nicht-axiologischer normativ-ethischer Sätze. Die Frage nach der besten sozialen Ordnung, die Frage nach der richtigen Verfassungswahl hat vom Standpunkt eines methodischen Individualismus als Antwort kein axiologisches Urteil. Dennoch ist diese Frage keine außermoralische Problemstellung. Bisweilen gibt sich jedoch die Unterscheidung von axiologischen und nicht-axiologischen normativ-ethischen Sätzen als Unterscheidung von moralischen und bloß 'instrumentellen' Urteilen aus.<sup>1</sup>

So meint Harsanyi, die normative Theorie sozialer Entscheidungen habe als Gegenstand weder eine moralische noch eine spieltheoretische Frage. Das Problem der besten Verfassung sei kein wesentlich moralisches, sondern ein instrumentelles Problem:

"We frame constitutions for governments, for clubs, and for associations, not in order to have these constitutions admired for their ethical or aesthetic qualities, but rather in order to maximize the chances that good social decisions will be generated by these constitutions."<sup>2</sup>

Aus der Tatsache, daß - zumindest vom Standpunkt des methodischen Individualismus<sup>3</sup>- Verfassungen nicht um ihrer selbst willen für gut gehalten werden, folgt nicht, daß es sich bei der Beurteilungen von Verfassungen um ein außermoralisches Problem handelt. Nur ein Teil moralischer Wert- und Verpflichtungsaussagen hat zum Inhalt, daß das jeweils beurteilte Objekt, die jeweilige Handlung von intrinsischem Wert sei. Der (wohl größere)

---

<sup>1</sup> Insbesondere in der ökonomischen Methodendiskussion.

<sup>2</sup> J.C. Harsanyi, *Bayesian decision theory, rule utilitarianism, and Arrow's impossibility theorem*, *Theory and decision* 11 (1979) 289-317, S.309.

<sup>3</sup> s.u. S. 158ff.

Teil moralischer Aussagen enthält keine Behauptungen über den intrinsischen Wert von Handlungen, Motiven etc.. Ihre Begründung kann deontologisch (und ohne Rekurs auf intrinsische Werte auskommen) oder teleologisch sein. Ist sie teleologisch und nimmt daher - zumindest implizit - auf intrinsische Werte Bezug, so könnte man die so begründeten moralischen Aussagen mit Harsanyi 'instrumentell' nennen - aber davon unbeschadet bleiben es moralische, oder genauer normativ-ethische Aussagen.

Die analytische Moralphilosophie unterscheidet moralische von außermoralischen Werten. Es gibt eine Interpretation dieser Unterscheidung, die zunächst unproblematisch erscheint. Geht es um die Bewertung von Handlungen, Handlungsmotiven und Personen, handelt es sich um moralische Werturteile,<sup>3</sup> geht es dagegen um die Bewertung von Bildern, Autos und Spülmaschinen, handelt es sich um außermoralische Werturteile. Ein Auto kann nicht 'moralisch gut' sein. Eine Freundschaft, eine wissenschaftliche Arbeit, eine Regierung, kann ebenfalls nicht moralisch gut sein. Nach dieser Interpretation ist das Kriterium für die Unterscheidung von moralischen und außermoralischen Urteilen die Art des Gegenstandes.<sup>1</sup>

Dieses Kriterium ist jedoch in vielen Fällen inadäquat. Die Frage nach der gerechten Staatsverfassung etwa ist eine der ältesten der Moralphilosophie. Wenn Platon für einen Drei-Stände-Staat als Ausdruck der σοφροσύνη plädiert, oder wenn Kant meint, die republikanische sei die einzige mit dem Recht vereinbare Verfassung (als forma regiminis)<sup>2</sup>, so ist das jeweils eine moralische Qualifikation.

Diese Verfassungen werden dabei moralisch beurteilt: Für Kant ist die republikanische forma regiminis

---

<sup>1</sup> Vgl. Frankena (1963).

<sup>2</sup> Vgl. I. Kant, *Zum ewigen Frieden*, (1795), Erster Definitivartikel.

<sup>3</sup> Bei Handlungen sollten wir zur Unterscheidung "Verpflichtungsurteile" sagen, obwohl natürlich auch Verpflichtungsurteile moralisch werten.

die moralisch gebotene, es geht ihm nicht um den außermoralischen Wert dieser Verfassung. Der Satz "X ist eine gute Verfassung" kann eine außermoralische Qualifikation sein, z.B. in einer Diskussion unter Verfassungsjuristen, denen es um Kriterien der inneren Konsistenz, der Einteilung, der Präzision der Formulierungen, etc. geht; dieser Satz kann aber auch eine moralische Qualifikation sein. Die Unterscheidung nach Gegenständen trägt also nicht.

Eine Alternative besteht darin, die Einteilung moralische/außermoralische Werte zu einem der Gegenstände normativer Ethik zu machen (während diese Unterscheidung in der analytischen Moralphilosophie als metaethisches Problem behandelt wird).

Eine utilitaristische Theorie kann z.B. das moralisch Gute als dasjenige definieren, das die Summe menschlicher Glückseligkeit maximiert. Sie kann darüber hinaus den Anspruch erheben, damit das moralisch Gute als dasjenige definiert zu haben, das das außermoralisch Gute maximiert. Dieser Anspruch wäre dann gerechtfertigt, wenn das menschliche Glück das einzige intrinsisch (außermoralisch) Gute ist.

Dieser eudämonistische Utilitarismus würde eine objektivistische Theorie des außermoralisch Wertvollen vertreten. Es ist jedoch auch eine subjektivistische Theorie des außermoralisch Wertvollen denkbar. Sozialwahltheorien z.B. könnten folgendermaßen interpretiert werden: die Relation *moralisch besser* bestimmt die kollektive Präferenzrelation  $R$  in Abhängigkeit von den individuellen Präferenzrelationen. Das außermoralisch Gute ergibt sich je individuell aus den Präferenzen der einzelnen Personen, das moralisch Gute besteht in einer bestimmten interpersonellen Verteilung des außermoralisch Guten, wie sie diese oder jene Aggregationsregel beschreibt.

Da eine normativ-ethische Theorie wenig Sinn hat, wenn sie für keinen einzigen der üblichen Gegenstände

moralischer Qualifikation, nämlich Handlungen, Handlungsmotive, Verhaltensweisen, Charaktereigenschaften oder Personen Kriterien moralischer Beurteilung entwickelt, andererseits jedoch die moralischen Urteile der Person Einfluß auf die Gestalt der Präferenzrelationen haben, und das moralisch Gute ja als Funktion des außermoralisch Guten (und seiner Verteilung) bestimmt werden soll, ergibt es sich auch unter diesem Aspekt, daß das von der normativen Aggregationsregel Aggregierte nicht die overall-Präferenzen der Individuen sein können.

Die collective-choice-Theorie ist ein Abkömmling der ökonomischen Wohlfahrtstheorie, die zur Grundfrage hat, wie kann die (kollektive) Wohlfahrt eines Gemeinwesens definiert werden. Daher liegt es nahe, auch eine weitere Interpretation in Erwägung zu ziehen, die die Aggregation individueller Präferenzen als Bestimmung des außermoralisch Guten versteht. Bei dieser Interpretation bliebe offen, in welcher Beziehung die moralische Rangordnung zur kollektiven Präferenzrelation steht. Die Sozialwahltheorie wäre damit eine außermoralische, aber axiologische Theorie.

Theorien, die sich auf moralische Handlungsempfehlungen beschränken, wollen wir nicht als axiologische Theorien bezeichnen. Verpflichtungsurteile bewerten Handlungen, aber diese Art moralischer Bewertung sollte man von anderen moralischen Bewertungen unterscheiden. Es geht in dieser Arbeit besonders um die Möglichkeit einer nicht-axiologischen Ethik. Uns interessieren nicht-axiologische Theorien normativer Ethik zweiter Ordnung.

Braithwaite nennt seine Überlegungen amoralisch "in the sense that they will not be based upon any first-order moral principles".<sup>1</sup> Da seine Fairneßtheorie

---

<sup>1</sup> R.B. Braithwaite, *Theory of games as a tool for the moral philosopher*, Cambridge 1961 (1. Ausgabe 1955), S.5.

jedoch explizit präskriptiv ist (sie ist den normativ-ethischen Theorien des unparteiischen Beobachters zuzurechnen), ist es nicht sinnvoll, sie amoralisch zu nennen. Nach unserer Kategorisierung handelt es sich vielmehr um eine nicht-axiologische Theorie.

Wie dieses Beispiel zeigt, ist es zweckmäßig, über die Unterscheidungen axiologisch/nicht-axiologisch, normativ-ethisch/metaethisch, moralisch/außermoralisch hinaus auch zwei Ebenen ethischer Theorienbildung auseinander zu halten. Wir nennen die beiden Ebenen Ethik erster und Ethik zweiter Ordnung.

Normative Theorien, die eine Kritik von Bedürfnissen, Wünschen, Interessen begründen, gehören zur Ethik erster Ordnung. Normative Theorien, die aufgrund gegebener Bedürfnisse, Wünsche, Interessen, Handlungsempfehlungen geben, moralische Präferenzen bestimmen oder resultierende Präferenzen in eine moralische Rangordnung bringen (normative Metapreferenztheorie), gehören zur Ethik zweiter Ordnung.

Braithwaites Fairneßtheorie ist nicht nur nicht-axiologisch, sondern gehört auch zur Ethik zweiter Ordnung.

"My treatment will be ethically neutral between the collaborating parties; it will not suppose that one code of values is better than any other,"<sup>1</sup>

und er nennt die aufgrund spieltheoretischer Analyse eines Interessenkonflikts erarbeitete speziellen Empfehlungen "second order moral principles".

Ebenso sind normative Theorien kollektiver Wahl in der ersten Interpretation (s.o.) zur Ethik zweiter Ordnung zu rechnen. Der sog. "individualistische Ansatz" der Sozialwahltheorie ist nur ein anderer Name für diesen Sachverhalt.

In der ökonomischen Diskussion findet sich dafür häufig eine metatheoretische selbst wiederum normative, Begründung, die irreführend ist. Sie läßt sich in einem

---

<sup>1</sup> Braithwaite (1969), S.5



Satz zusammenfassen: Es sollten die Wünsche und Bedürfnisse wie sie sind, akzeptiert werden.

"Respects for the personal freedom and personal interests of other individuals must make us respect their actual explicit preferences unless we have strong reason to think that their preferences are mistaken or are patently antisocial."<sup>1</sup>

Im Kontext normativer Theorien hat solches 'Akzeptieren' der vorgegebenen Interessen oder 'Präferenzen' jedoch gerade nicht die Konsequenz, den Entscheidungsträgern zu empfehlen, sich auch diesen eigenen Präferenzen entsprechend zu verhalten. Ein Utilitarist empfiehlt z.B. ein den Gesamtnutzen maximierendes Verhalten, was eben i.a. nicht im Interesse des Einzelnen ist. Die Mehrheitswahlregel empfiehlt, sich den Präferenzen der Mehrheit gemäß zu verhalten, was, falls diese Regel kollektiver Wahl nicht institutionell sanktioniert ist, nicht in jedem Fall den Interessen des Einzelnen entspricht. So daß diese Begründung für einen individualistischen Ansatz nicht überzeugt.

Es ist in diesem Zusammenhang nötig, zwei Arten des Individualismus zu unterscheiden.

Der methodische Individualismus<sup>2</sup> behandelt (normative) Fragen nach der richtigen kollektiven Entscheidung, nach der richtigen staatlichen, gesellschaftlichen und wirtschaftlichen Organisation allein aufgrund der Haltung (der Präferenzen, der subjektiven Interessen) der betroffenen Individuen angesichts vorliegender Alternativen. Sozialethische Theorien, die dem methodischen

---

<sup>1</sup> J.C. Harsanyi, *Preferences and utilitarian theory: some comments*, Erkenntnis 11 (1978), S.397.

<sup>2</sup> "Methodischer Individualismus" bezieht sich hier ausschließlich auf normativ-ethische Theorien. Er darf nicht verwechselt werden mit dem *methodologischen Individualismus* in den Sozialwissenschaften, wie ihn besonders emphatisch Hayek und Watkins vertreten: "social events like inflation ... must be explained in terms of people: in terms of the situations people confront and the ambitions, fears and ideas which activate them. In short, large scale social phenomena must be accounted for by the situations, dispositions and beliefs of individuals. This I call methodological individualism", J.W.N. Watkins, *Methodological individualism: A reply*, Philosophy of science 22 (1955), S.58

Vgl.a. F.A. Hayek, *The constitution of liberty*, London 1960 und die Aufsatzsammlung *New studies in philosophy, politics, economics and the history of ideas*, London 1978. Vgl. W. Raub/ T. Voss, *Individuelles Handeln und gesellschaftliche Folgen. Das individualistische Programm der Sozialwissenschaften*, Darmstadt 1981.

Individualismus gerecht werden, gehören daher zur Ethik zweiter Ordnung. Der methodische Individualismus kann einen dogmatischen oder heuristischen Charakter haben. Dogmatisch, soweit individuelle Präferenzen (genauer: Interessen, Wünsche, Bedürfnisse) grundsätzlich normativer Kritik entzogen werden, heuristisch, soweit Handlungs- und Entscheidungsempfehlungen zwar als Funktion gegebener individueller Interessen bestimmt werden, jedoch ohne Festlegung zu der Frage, ob diese zugrunde gelegten individuellen Interessen einer (normativen) Kritik unterzogen werden können oder nicht.<sup>1</sup>

Ein epistemologisches Problem kann zur Rechtfertigung des heuristischen Individualismus herangezogen werden: Es bietet sich bisher keine hinlänglich entwickelte allgemeine normative Theorie der Kritik individueller Bedürfnisse und Interessen an, so daß es gerechtfertigt erscheint, aus der Not eine Tugend zu machen - umsomehr,

---

<sup>1</sup> Teitelmann, *The limits of individualism*, Journal of philosophy 69 (1972) 545-556, markiert anhand von Rawls' Gerechtigkeitstheorie die Grenzen 'individualistischer' Ethik, wobei sich seine Argumentation letztlich gegen jede Form normativer Vertragstheorie wendet. Wenn wir hier von (methodisch) individualistischen Theorien sprechen, so verwenden wir den Terminus "individualistisch" in einer anderen Bedeutung als Teitelmann. Rawls' Theorie macht die Gerechtigkeit gesellschaftlicher Institutionen von den Präferenzen fiktiver Individuen in einer fiktiven Entscheidungssituation abhängig. Daher gehört diese Theorie nicht zu den individualistischen Theorien in unserem Sinne, da sie das moralisch Richtige gerade nicht von der Haltung der betroffenen Individuen in der tatsächlichen Entscheidungssituation abhängig macht. Die Wahl der Gerechtigkeitsprinzipien ist nicht individualistisch, das Rawlssche Differenzprinzip selbst kann jedoch individualistisch interpretiert werden - zumindest hat es in dieser Interpretation Eingang in die Sozialwahltheorie gefunden. Vgl. dazu A.K. Sen, *Collective choice and social welfare*, San Francisco 1970, S.135-141; ders., *Rawls versus Bentham: An axiomatic examination of the pure distribution problem*, in: Daniels (1976) 283-292; ders., *Welfare inequalities and Rawlsian axiomatic*, Theory and decision 7 (1976) 243-262; P.J. Hammond, *Equity, Arrow's conditions and Rawls' difference principle*, Econometrica 44 (1976) 793-804; K.W.S. Roberts, *Possibility theorems with interpersonally comparable welfare levels*, Review of economic studies 47 (1980) 409-420; W. Gaertner, *Rawlsianism, utilitarianism, and profiles of extended orderings*, Bielefeld 1980.

als die Ergebnisse eines solchen heuristischen individualistischen Ansatzes durch eine normativ-ethische Theorie erster Ordnung in keiner Weise obsolet würden.

Während die utilitaristische Moral zur (methodisch) individualistischen Ethik gehört, obwohl sie die individuellen Verpflichtungsurteile ausschließlich davon abhängig macht, welche Handlung das Gemeinwohl maximiert und insofern auf persönliche Interessen und Wertvorstellungen wenig Rücksicht nimmt, gibt es eine starke nicht-individualistische Tradition moralphilosophischen Denkens, die am reinsten bei Hegel verkörpert ist und die von einer wechselseitigen (dialektischen) Verkoppelung der persönlichen Interessen mit dem Gemeinwohl ausgeht. Theorien, die dieser nicht-individualistischen Tradition angehören, trennen i.a. ihre normativen Aussagen nicht von den empirisch-deskriptiven, was die Beurteilung, aber auch schon die Rekonstruktion ihres normativen Gehalts erschwert.<sup>1</sup> Ebenfalls in dieser Tradition steht die normativ-ethische Theorie von Nicholas Rescher<sup>2</sup>, die sich jedoch in dem Bemühen um begriffliche und argumentative Klarheit von den meisten Beiträgen dialektischer Philosophie abhebt.

Mit dem methodischen Individualismus in der Sozialethik ist der sozialetische Individualismus nicht zu wechseln.

Der sozialetische Individualismus umfaßt eine Klasse normativ-ethischer Theorien der richtigen Organisation von Staat, Wirtschaft und Gesellschaft, denen die Betonung eines möglichst großen individuellen Entscheidungsspielraums gemeinsam ist.

---

<sup>1</sup> Vgl. jedoch den Versuch von W. Leinfellner, *Marx and the utility approach to the ethical foundations of microeconomics*, in: Gottinger/Leinfellner (1978) 33-58.

<sup>2</sup> Vgl. N. Rescher, *Unselfishness*, Pittsburg 1975. Der Autor, obwohl Vertreter der analytischen Philosophie betont sogar selbst diese Verwandtschaft, vgl.S.106ff.

Eine Theorie normativer Ethik kann individualistisch in methodischem und in sozialem Sinne zugleich sein, muß es aber keineswegs, wie die utilitaristische Moral zeigt.

Es stellt sich nun die Frage, ob nicht-axiologische Theorien normativer Sozialethik, speziell solche zweiter Ordnung, in anderer Weise begründet werden können als axiologische Theorien.

Von zahlreichen normativ-ethischen Theorien wird immer wieder behauptet, sie seien außermoralisch begründbar. Das gilt u.a. für einige wohlfahrtstheoretische Kriterien, besonders das Pareto-Kriterium und für vertragstheoretische Sozialethiken, besonders die von Thomas Hobbes.

Eine 'Minimalmoral' im Hobbesschen Sinne, ein durch Sanktionsdrohungen abzustützendes Normensystem, ist eine nicht-axiologische Theorie normativer Ethik zweiter Ordnung. Denn das Hobbessche Argument lautet: Welche Präferenzen, welche Vorlieben und Wünsche die unterschiedlichen Individuen auch haben, in jedem Fall hat jedes einzelnen dieser Individuen ein Interesse daran, daß jeder gezwungen ist, bestimmte Verhaltensregeln einzuhalten.<sup>1</sup>

Weder Hobbes noch Pareto preisen bestimmte moralische Werte. Der Ausgangspunkt der normativen Theorie liegt jeweils in den gegebenen Präferenzen der Individuen. Diese Präferenzen selbst sind nicht Gegenstand moralischer Bewertung.

Nun wird behauptet, damit könnten Theorien dieser Art, die wir als nicht-axiologische Theorien zweiter Ordnung bezeichnet haben, außermoralisch gerechtfertigt werden. Bezüglich des Pareto-Kriteriums ist diese Ansicht bis

---

<sup>1</sup> Vgl. T. Hobbes, *Leviathan or the matter, form and authority of government* (1651).

heute unter Ökonomen weit verbreitet. Die außermoralische Begründbarkeit normativer Vertragstheorien der Hobbesschen Art ist erneut Gegenstand der philosophischen Diskussion geworden.<sup>1</sup>

In einem strengen Sinn ist dieser Anspruch sicher falsch; auch eine Hobbesianische Ethik hat normativ-ethische Prämissen, falls sie präskriptiv ist - erst eine Umdeutung in eine deskriptive Theorie derjenigen institutionellen Konfliktregelungen, die im allgemeinen Interesse sind, könnte sich von solchen normativ-ethischen Prämissen befreien. Insoweit ist Hoerster zu widersprechen, wenn er zwar einerseits daran festhält, daß es "keine moralische Erkenntnis gibt"<sup>2</sup>, andererseits aber "von einem intersubjektiven Standpunkt aus" im Rahmen einer Hobbesianischen 'Minimalmoral' aufzeigt, daß in jeder menschlichen Gesellschaft bestimmte Normen "allgemein akzeptiert, das heißt in Geltung sein sollten".<sup>3</sup> In unserer Terminologie löst sich dieser Widerspruch: Es mag normative Erkenntnis geben, ohne daß es axiologische Erkenntnis gibt. Die Hobbessche Theorie benötigt zu ihrer Rechtfertigung keine axiologischen Prämissen, ihre Rechtfertigung ist damit nicht außermoralisch, sondern nicht-axiologisch, und diese Eigenschaft teilt sie mit den modernen Fairneßtheorien genauso wie mit der im folgenden diskutierten normativ-ethischen Theorie kollektiver Rationalität. Alle drei Sparten sind unter die nicht-axiologischen Theorien normativer Ethik zweiter Ordnung zu subsumieren.

---

<sup>1</sup> Vgl. z.B. G. Braungart, *Die Ethik und ihre außermoralische Rechtfertigung bei Hobbes*, Ratio 31 (1981) 50-63

<sup>2</sup> N. Hoerster, *Zur Begründung einer Minimalmoral*, in: Morscher/Stranzinger (1981), S.133

<sup>3</sup> Hoerster (1981), S.131

Die Beschränkung der Untersuchung auf nicht-axiologische Theorien hat also einen wesentlichen metaethischen Aspekt: Die Gültigkeit unserer Ergebnisse ist von der Beantwortung der metaethischen Fragestellung, ob es objektive Werterkenntnisse gibt, unabhängig. Die metaethische Skepsis normativ-ethischen Theorien gegenüber ist für normative, aber nicht-axiologische Theorien daher nur zum Teil relevant. M.a.W., wir können einen Großteil der vorgebrachten metaethischen Kritik am 'Moralobjektivismus' in seinen verschiedenen Ausprägungen akzeptieren, ohne unserer normativen Theorie die methodologischen Fundamente streitig zu machen.

Der Wertobjektivist nimmt an, daß es objektive Werte gibt und setzt sich daher der Frage aus, wie er diese erkennt. Darauf geben metaethische Intuitionisten und Deskriptivisten unterschiedliche Antworten. Wer beide Antworttypen aus grundsätzlichen Erwägungen als unbefriedigend empfindet und daher den Wertobjektivismus für eine falsche metaethische Theorie hält, ist jedoch - und das ist hier das allein Wesentliche - dennoch nicht dazu verurteilt, nun auf jegliche normativ-ethische Aussage zu verzichten. Er kann auch weiterhin nicht-axiologische normativ-ethische Theorien zweiter Ordnung vertreten. Auch diese enthalten Aussagen darüber, was getan werden soll, aber das beruht nicht auf einer besonderen Art von Werterkenntnis, sondern auf Interessenanalyse, und dazu können eine methodisch geklärte Entscheidungs- und Spieltheorie und die Ergebnisse der collective-choice-Forschung einen Beitrag leisten.

Es ist wichtig, sich über die Tragweite dieser Konsequenz im klaren zu sein: Ein Wertsubjektivist (in der Bedeutung des normalen philosophischen Sprachgebrauchs) kann danach sehr wohl eine normativ-ethische Theorie vertreten, von der er glaubt, daß sie objektiv gültig und intersubjektiv begründbar ist. (Das ist wohl auch der Kern von Hoersters Überlegungen)

Am Ende dieses Abschnittes sollten noch einige Sätze zu der Beziehung gesagt werden, in der die Ethik zweiter zur Ethik erster Ordnung steht.

Die meisten Theorien normativer Ethik zweiter Ordnung wollen kategorisch verstanden werden. So etwa in der Sozialwahltheorie: Haben  $n$  Individuen folgende Präferenzrelationen  $R_i$ , so ergibt sich die kollektive (und damit i.a. auch moralische) Ordnung der Alternativen  $R$ .

Wir fassen jedoch, um einen denkbaren Konflikt zwischen Ethik erster und zweiter Ordnung auszuschließen, die Sollensbehauptungen einer Ethik zweiter Ordnung grundsätzlich als hypothetische auf.

Das heißt z.B., Eine Fairneß-Empfehlung für einen Interessenkonflikt besagt nicht: Ihr solltet eine kollektive Handlungsweise  $\Sigma$  wählen, da ihr die und die persönlichen Interessen  $\pi_i$  habt, sondern: Relativ zu vorliegenden  $\pi_i$  wäre  $\Sigma$  eine faire Lösung - aber ob  $\Sigma$  zu befolgen ist, kann u.U. noch von der Qualität der  $\pi_i$  abhängen.

Ein Beispiel soll das erläutern: In Braithwaites Untersuchung geht es um zwei Musiker in angrenzenden Zimmern. Aufgrund einer Analyse ihrer persönlichen Interessen wird eine Fairneß-Lösung entwickelt, die empfiehlt, wie die beiden ihre Spielzeiten einteilen sollten. Hier handelt es sich um harmlose, sozusagen 'ethisch neutrale' Interessen, um die faire Verwirklichung außermoralischer Werte. Bei diesem Interessenkonflikt liegt es daher nahe, die Fairneß-Empfehlung kategorisch zu interpretieren: Da ihr diese persönlichen Interessen habt, ist  $\Sigma$  fair (oder gerecht), und damit solltet ihr  $\Sigma$  tun. Da eine solche Fairneß-Empfehlung in ihrer Rechtfertigungsargumentation jedoch an keiner Stelle die Art der persönlichen Interessen wesentlich enthält, also auf Interessenkonflikte beliebiger Art anwendbar ist, wäre diese Interpretation unangemessen. Man denke etwa - um ein

krasses, aber damit um so deutlicheres Beispiel zu geben - an den Interessenkonflikt zwischen der jüdischen Bevölkerung Osteuropas und den Nationalsozialisten (und ihren Anhängern). Es ist anzunehmen, daß eine völlig analoge Analyse als 'Fairneß-Lösung' eine x-prozentige Vernichtung des osteuropäischen Judentums durch die Nazis ergäbe, und x größer Null ist. Dieses Ergebnis kann kein gültiges normativ-ethisches Urteil darstellen.

Daraus lassen sich unterschiedliche Konsequenzen ziehen:

- (1) Diese Fairneßtheorie zurückzuweisen.
- (2) Jegliche Form ethischer Analyse zweiter Ordnung für unzulässig zu halten.
- (3) Die ethische Analyse zweiter Ordnung auf ganz bestimmte Interessenkonflikte zu beschränken.

Diese drei Reaktionen haben eines gemeinsam: Sie beruhen auf einer kategorischen Interpretation normativer Theorien zweiter Ordnung.

Eine grundsätzliche hypothetische Interpretation von Handlungsanleitungen normativer Ethik zweiter Ordnung bietet jedoch eine plausiblere Lösung des Dilemmas. Für Fairneßtheorien heißt das, die Empfehlungen vollständig als " $\Sigma$  ist fair bezüglich  $\langle R_{w_1} \rangle$ " zu formulieren. Um  $\Sigma$  als normativ geboten behaupten zu können, ist dann noch eine zusätzliche Qualifikation der  $R_{w_1}$  nötig.

Auf unser obiges Beispiel angewendet, ergibt eine hypothetische Interpretation: Die x-prozentige Vernichtung ist zwar die 'Fairneß-Lösung' aufgrund vorliegender Interessen, aber da das Interesse der Nazis an einer Ausrottung des Judentums unmoralisch ist (das gehört zur normativen Ethik erster Ordnung), kann diese Fairneß-Lösung nicht als normativ geboten gelten.

Es besteht, wie man hier sieht, eine enge Verbindung zwischen einem methodischen Individualismus der sozial-ethischen Theoriebildung und einer hypothetischen Interpretation normativer Ethik zweiter Ordnung.



## 12. RATIONALITÄT UND MORALITÄT ALS KONKURRIERENDE PRINZIPIEN

Ἡ μὲν ἄρα δίκαια ψυχὴ καὶ ὁ δίκαιος ἀνὴρ εὖ βιώσεται, κακῶς δὲ ὁ ἄδικος.  
φαίνεται, ἔφη, κατὰ τὸν σὸν λόγον.  
Ἀλλὰ μὴν ὅ γε εὖ ζῶν μακάριός τε καὶ εὐδαίμων, ὁ δὲ μὴ ταναυτῖα.  
Πῶς γὰρ οὐ;  
Ὁ μὲν δίκαιος ἄρα εὐδαίμων, ὁ δ' ἄδικος ἄθλιος.  
Ἔστω, ἔφη.  
Ἀλλὰ μὴν ἄθλιόν γε εἶναι οὐ λυσιτελεῖ, εὐδαίμονα δέ.  
Πῶς γὰρ οὐ;  
Οὐδέποτε! ἄρα, ὦ μακάριε Θρασύμαχε, λυσιτελέστερον ἀδικία δικαιοσύνη. (Πολιτεία, A 353e - 354a)<sup>1</sup>

Zahlreiche Ethiker haben sich viel Mühe gegeben, aufzuzeigen, daß moralisches Verhalten zugleich im eigenen Interesse ist.

In der *Politeia* und in anderen Dialogen (insbesondere Protagoras und Gorgias) versucht Platon aufzuzeigen, daß nur ein δίκαιος ἀνὴρ ein gerechter Mensch εὐδαίμων 'glücklich' sei. Neuerdings hat Mackie in ähnlicher Weise für eine weitgehende Übereinstimmung der Forderungen einer utilitaristischen und einer egoistischen Ethik argumentiert. Mackies Überlegungen stützen sich auf die

---

<sup>1</sup> *Platonis opera*. Ed. J. Burnet, Bd.4, Oxford 1902.

Die gerechte Seele und der gerechte Mensch wird also gut leben, und der ungerechte schlecht.

Es folgt dies offenbar aus deinen Worten, versetzte er.

Nun ist aber doch derjenige, welcher gut lebt, selig und glücklich, und wer nicht - das Gegenteil?

Wie wäre es anders möglich?

Der Gerechte ist also glücklich, der Ungerechte unglücklich.

Meinethalben, sagte er.

Unglücklichsein ist nun aber doch nicht vorteilhaft, wohl aber das Glücklichein.

Wie wäre es anders möglich?

Nimmermehr also, mein bester Thrasymachos, ist die Ungerechtigkeit nützlicher als die Gerechtigkeit.

(Übers. W.S. Teuffel, *Platon Sämtliche Werke*, Verlag Lambert Schneider, Berlin o.J., Bd.2, S.45)

dispositionelle Strukturierung menschlichen Verhaltens. Sokrates' Argumente im Gorgias-Dialog kann man ähnlich interpretieren. Die Übereinstimmung der Forderungen von εὐδαιμονία und δικαιοσύνη war notwendig, um die ethischen Forderungen überhaupt rechtfertigen zu können. Gegen common sense der griechischen Kultur hätte eine Ethik nicht bestehen können, die den 'Klugheitsregeln' der εὐδαιμονία zuwiderläuft. Es galt aufzuzeigen, daß die Sophisten mit ihren amoralischen Klugheits-Empfehlungen an ihren eigenen Ansprüchen gemessen, unrecht hatten. Auch für Platon liegt die Rechtfertigung moralischen Verhaltens in dem Nachweis, daß es dem eigenen Interesse der Handelnden entspricht. Allerdings läßt sich dies nach Platon erst bei einer tieferen Analyse aufzeigen, die über die oberflächlichen Betrachtungen der Sophisten hinausgeht. Die Moral ist lehrbar, und da das moralisch Richtige auch als intuitiv empfunden wird, ist diese Lehrtätigkeit eigentlich eine Μαιευτική, eine Hebammenkunst, die nur noch helfend eingreift, um das Verborgenen ans Licht zu bringen. Rationalität und Moralität sind keine Gegensätze, weder in den moralischen intuitiven Überzeugungen des Volkes noch bei wissenschaftlicher Prüfung. Nur der Sophist glaubt, diese Übereinstimmung, intellektuell leichtfertig, anzweifeln zu können.

Während für Platon das moralische Handeln sich auch darin rechtfertigt, daß es dem langfristigen persönlichen Interesse entspricht, stellt sich für Butler<sup>1</sup> der persönliche Vorteil des moralischen Handelns als Folge ein. Pointiert kann man sagen, das Glück ist für Platon Kriterium des Moralischen, für Butler ein erfreulicher Nebeneffekt.<sup>2</sup>

---

<sup>1</sup> Vgl. J. Butler, *Sermons*, 1729.

<sup>2</sup> Vgl. dazu auch die Untersuchung *Ethical egoism in hellenic thought* von J. Wheeler, Ann Arbor 1976.

Auch Hume meint, daß eine Theorie der Moral nur unter der Bedingung sinnvoll ist, daß sie detailliert aufzeigen kann, daß alle von ihr empfohlenen Pflichten auch im Interesse jedes Individuums sind.<sup>1</sup>

Bei Kant hingegen ist zwar moralisches Handeln vernunftgemäß, aber vernunftgemäßes Handeln ist nicht mehr im persönlichen Interesse, im Gegenteil: Ist eine Entscheidung im persönlichen Interesse, ist sie nicht mehr moralisch, sondern, falls in Übereinstimmung mit dem moralischen Gesetz, bloß legal.<sup>2</sup>

"Das Wesentliche alles sittlichen Werts der Handlungen kommt darauf an, daß das moralische Gesetz unmittelbar den Willen bestimme. Geschieht die Willensbestimmung zwar gemäß dem moralischen Gesetze, aber nur vermittelt eines Gefühls, welcher Art es auch sei, das vorausgesetzt werden muß, damit jenes ein hinreichender Bestimmungsgrund des Willens werde, mithin nicht um des Gesetzes willen; so wird die Handlung zwar Legalität, aber nicht Moralität enthalten."<sup>3</sup>

Der klassische Liberalismus behauptet, daß in einem eingeschränkten Bereich, nämlich dem einer freien, kapitalistischen Konkurrenzökonomie, der marktwirtschaftliche Rahmen dafür sorgt, daß das Handeln im persönlichen Interesse, das des Produzenten wie des Konsumenten, zum allgemeinen Wohl optimal beiträgt. Aber im übrigen ist im modernen Denken die Übereinstimmung von moralischem Handeln und Handeln im persönlichen Interesse keineswegs selbstverständlich. Besonders scharf hat diesen Gegensatz Thomas Hobbes im Leviathan herausgearbeitet. Kurt Baier schreibt, an die Hobbessche Tradition des moralphilosophischen Denkens anknüpfend:

"Moralties are systems of principles whose acceptance by everyone as overruling the dictates of self-interest is in the interest

---

<sup>1</sup> Vgl. D. Hume, *An enquiry concerning the principles of morals*, Oxford 1894, sec. 9.

<sup>2</sup> So wie für Platon δίκαιον die moralisch gebotene Handlung schlechthin auszeichnet, so ist für Kant das Rechtliche das moralisch Gebotene - wir nehmen uns in beiden Fällen die Freiheit einer anderen Wortwahl.

<sup>3</sup> I. Kant, *Kritik der praktischen Vernunft*, 1788, 1. Teil, 1. Buch, 3. Hauptstück, Reclam-Ausgabe (1976) S.117.

of everyone alike, though following the rules of morality is not of course identical with following self-interest. If it were, there could be no conflict between a morality and self-interest and no point in having moral rules overriding self-interest."<sup>1</sup>

Und damit ist für Baier auch die Frage beantwortet, warum man moralisch sein sollte:

"We should be moral because being moral is following rules designed to overrule self-interest whenever it is in the interest of everyone alike that everyone should set aside his interests."<sup>2</sup>

und dann

"It is not possible, however, that everyone should do better for himself by following enlightened self-interest rather than morality. The best possible life for everyone is possible only by everyone's following the rules of morality, that is, rules which quite frequently may require individuals to make genuine sacrifices."<sup>3</sup>

Letzteres steht dann allerdings in deutlichem Gegensatz zur Hobbesschen Staatstheorie. Denn Hobbes rechtfertigt ja die staatlichen Institutionen (genaugenommen jegliche Form von staatlicher Autorität) damit, daß nur auf diese Weise das Eigeninteresse der Einzelnen zu einem für alle erträglichen Nebeneinander wird,<sup>4</sup> wogegen Baiers moralische Subjekte ja gerade durch die Erkenntnis des Gemeininteresses ihr Eigeninteresse zurückstellen und damit in einer wohlgeordneten Gesellschaft (im Sinne Rawls')<sup>5</sup> staatlicher Zwänge nicht bedürfen.<sup>6</sup>

---

<sup>1</sup> K. Baier, *The moral point of view: A rational basis of ethics*, Ithaka, N.Y. 1958, S.314

<sup>2</sup> ebd., S.314

<sup>3</sup> ebd., S.314f. Im nächsten Abschnitt versuchen wir die Frage, was ist ein moralisches System, genauer zu klären.

<sup>4</sup> Vgl. auch die Interpretation von Carl Schmitt, *Der Leviathan in der Staatslehre Thomas Hobbes'*, Hamburg 1938, R. Schnur, *Individualismus und Absolutismus. Die politische Theorie von Thomas Hobbes*, Berlin 1963 und D. Gauthier, *The logic of leviathan*, Oxford 1969, bes. S.57-61, S.89-98 und Kapitel IV.

<sup>5</sup> J. Rawls, *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975, S.493-503

<sup>6</sup> In Brocks spieltheoretischen Modell wäre Hobbes' Vertragsvorstellung durch ein 'disagreement payoff' des Verfassungsspiels

Die rationale Entscheidungstheorie könnte eines Tages, aus einer größeren historischen Distanz, als der groß-angelegte Versuch empfunden werden, das platonische Programm wiederaufzunehmen. Rationales Handeln ist nutzenmaximierendes (im Einzelfall nicht den Nutzen, sondern den Nutzererwartungswert maximierendes) Handeln. Der Nutzen ist Ausdruck der persönlichen Ziele, dazu können auch die persönlichen moralischen Ziele gehören. Eine normative Interpretation der rationalen Entscheidungstheorie führt sozusagen zu dem hyperplatonischen Ergebnis: Das individuelle nutzenmaximierende Verhalten ist geboten.

Moralische Kontroversen werden in diesem Modell zurückverlagert auf die Frage nach der richtigen oder moralisch besseren Gestalt individueller Nutzenfunktionen.

---

wiederzugeben, daß jede Person schlechter stellt, als es bei einer beliebigen Verfassungslösung der Fall wäre. Der Souverain hat nach Hobbes das Recht, über Leben und Tod seiner Untertanen zu bestimmen. Verurteilt der Herrscher einen Untertanen, der unschuldig ist, zum Tod, hat der nicht dem Untertan Unrecht getan, sondern gegen das Naturrecht verstoßen, Gott Unrecht getan (vgl. Hobbes 1651, XI). Trotz dieser Radikalität hobbesianischer Staatsphilosophie gibt es auch hier eine Grenze legitimer Staatsgewalt, nämlich konsequenterweise da, wo sie hinter den Urzustand zurückfällt. Im Gesellschaftsvertrag gibt jeder gewisse Rechte ab, mit dem Ziel des innerstaatlichen Friedens und der Verteidigung nach außen. Daher erhält der Untertan seine ursprünglichen Rechte zurück, sobald ihre Abtretung menschliches Leben gefährdet. (Vgl. XIV und XI) (Vgl. H. Brock, *A new theory of social justice based on the mathematical theory of games*, in: Ordeshook (1978), S.565ff. Schwieriger wäre es, Rousseaus Vertragsvorstellung in diesem Modell zu rekonstruieren, da bei Rousseau die Zivilisationsgeschichte als ein dynamisches Element zu berücksichtigen wäre. Vgl. J.J. Rousseau, *Du contrat social*, I,4; I. Fetschers Einleitung zu Hobbes' *Leviathan*, Neuwied/Berlin 1966, und detaillierter: P.C. Mayer-Tasch, *Autonomie und Autorität: Rousseau in den Spuren von Hobbes?*, Neuwied/Berlin 1968.

DER KONFLIKT INDIVIDUELLER RATIONALITÄT UND KOLLEKTIVER OPTIMALITÄT: DAS GEFANGENENDILEMMA

Nun zeigt es sich, daß individuelle Rationalität in Interaktionssituationen bestimmter Art zu einem Ergebnis führt, das den persönlichen Interessen jeder einzelnen beteiligten Person weniger entspricht, als es bei einem anderen Verhalten der Fall gewesen wäre.

Wenn wir annehmen, daß moralische Präferenzen pareto-inklusiv sind, so stimmt individuelle Rationalität und Moralität in manchen Entscheidungssituationen nicht überein. Ein Typ von Interaktionssituationen dieser Art wird für eine Situation mit zwei Personen und je zwei Entscheidungsmöglichkeiten nach einer Episode von Tucker als Gefangenendilemma bezeichnet. Wir brauchen diese Episode hier nicht von neuem zu erzählen<sup>1</sup>, es genügt, die Präferenzstruktur dieses Typs von Interaktionssituationen darzulegen.

Sind die beiden Entscheidungen für die Person A als  $\sigma_{11}$  und  $\sigma_{12}$  und für die Person B als  $\sigma_{21}$  und  $\sigma_{22}$  bezeichnet, und sind die Ergebnisse eindeutig durch die jeweiligen kollektiven Entscheidungen  $\langle \sigma_{11}, \sigma_{21} \rangle$ ,  $\langle \sigma_{12}, \sigma_{21} \rangle$ ,  $\langle \sigma_{11}, \sigma_{22} \rangle$ ,  $\langle \sigma_{12}, \sigma_{22} \rangle$  bestimmt, so können wir die Interaktionssituation vom Typ des Gefangenendilemmas folgendermaßen beschreiben:

1. A hat die Präferenzen<sup>2</sup>:

---

nach außen. Daher erhält der Untertan seine ursprünglichen Rechte zurück, sobald ihre Abtretung menschliches Leben gefährdet. (Vgl. XIV und XI)

<sup>1</sup> Vgl. Luce/Raiffa (1957), S.4

<sup>2</sup> Wir lassen hier zunächst offen, wie die individuellen Präferenzen, die das Gefangenendilemma konstituieren, zu interpretieren sind, und welche Unterschiede sich ergeben, wenn wir die Struktur der Interaktionssituation auf die Präferenzstruktur  $\langle R_{w1}, R_{w2} \rangle$  statt auf  $\langle R_{res1}, R_{res2} \rangle$  beziehen.

$$\begin{aligned} \langle \sigma_{11}, \sigma_{21} \rangle & \underset{1}{>} \langle \sigma_{12}, \sigma_{21} \rangle \\ \langle \sigma_{11}, \sigma_{22} \rangle & \underset{1}{>} \langle \sigma_{12}, \sigma_{22} \rangle \end{aligned}$$

Und B:

$$\begin{aligned} \langle \sigma_{11}, \sigma_{21} \rangle & \underset{2}{>} \langle \sigma_{11}, \sigma_{22} \rangle \\ \langle \sigma_{12}, \sigma_{21} \rangle & \underset{2}{>} \langle \sigma_{12}, \sigma_{22} \rangle \end{aligned}$$

Das führt aufgrund individueller Rationalität (bzw. hier dem spieltheoretischen Dominanzkriterium) zur kollektiven Entscheidung  $\langle \sigma_{11}, \sigma_{21} \rangle \in \Sigma^1 \times \Sigma^2$ .

Diese Präferenzen von A und B schließen jedoch nicht aus, daß dennoch

$$\langle \sigma_{11}, \sigma_{21} \rangle \underset{1}{>} \langle \sigma_{12}, \sigma_{22} \rangle \wedge \langle \sigma_{11}, \sigma_{21} \rangle \underset{2}{>} \langle \sigma_{12}, \sigma_{22} \rangle$$

bzw., daß die individuell rationalen Entscheidungen zu einem Ergebnis führen, daß den Interessen beider Personen weniger entspricht als ein anderes, daß sich infolge anderer individueller Entscheidungen der an der Interaktionssituation Beteiligten ergeben hätte. Eine Entscheidungssituation mit dieser Eigenschaft nennen wir ein Gefangenendilemma oder eine PD-Situation.<sup>1</sup>

Die Entscheidungssituation der nebenstehenden Matrix hat diese Struktur, wobei die Zahlenwerte als ordinale Repräsentation der individuellen Präferenzordnung zu interpretieren sind, um die Allgemeinheit einzuschränken.

		B	
		$\sigma_{21}$	$\sigma_{22}$
A	$\sigma_{11}$	$\langle 2, 2 \rangle$	$\langle 4, 1 \rangle$
	$\sigma_{12}$	$\langle 1, 4 \rangle$	$\langle 3, 3 \rangle$

---

<sup>1</sup> "PD" steht für "Prisoners Dilemma"

Das Gefangenendilemma als nicht-kooperatives Nicht-Konstantsummenspiel hat keine 'Lösung' im strengen Sinne (in der Terminologie der Spieltheorie), da  $\langle \sigma_{12}, \sigma_{22} \rangle \succ_1 \langle \sigma_{11}, \sigma_{21} \rangle$  und damit die kollektive Gleichgewichtsstrategie  $\langle \sigma_{11}, \sigma_{21} \rangle$  nicht pareto-optimal ist. Nicht-kooperative Zwei-Personen-Spiele haben eine Lösung im strengen Sinne, wenn es eine kollektive Gleichgewichtsstrategie gibt, die pareto-optimal ist, und alle pareto-optimalen kollektiven Gleichgewichtsstrategien  $\langle \sigma, \sigma' \rangle$  austauschbar und äquivalent sind.

Wobei:

- (1)  $\langle \sigma, \sigma' \rangle \in \Sigma^1 \times \Sigma^2$  ist pareto-optimal  $\Leftrightarrow$   
 $\sim \exists \langle \hat{\sigma}, \hat{\sigma}' \rangle \in \Sigma^1 \times \Sigma^2$ ,  $[\langle \hat{\sigma}, \hat{\sigma}' \rangle \succ_1 \langle \sigma, \sigma' \rangle \wedge$   
 $\wedge \langle \hat{\sigma}, \hat{\sigma}' \rangle \succ_2 \langle \sigma, \sigma' \rangle]$ .
- (2)  $\langle \sigma, \sigma' \rangle \in \Sigma^1 \times \Sigma^2$  ist Gleichgewichtspunkt (= kollektive Gleichgewichtsentscheidung)  $\Leftrightarrow$   
 $\forall \hat{\sigma} \in \Sigma^1$ ,  $\langle \sigma, \sigma' \rangle \succ_1 \langle \hat{\sigma}, \sigma' \rangle \wedge \forall \hat{\sigma}' \in \Sigma^2$ :  $\langle \sigma, \sigma' \rangle \succ_2 \langle \sigma, \hat{\sigma}' \rangle$ .
- (3) Zwei Gleichgewichtspunkte  $\langle \sigma, \sigma' \rangle \in \Sigma^1 \times \Sigma^2$  und  $\langle \hat{\sigma}, \hat{\sigma}' \rangle \in \Sigma^1 \times \Sigma^2$  sind äquivalent  $\Leftrightarrow$   
 $\langle \sigma, \sigma' \rangle \approx_1 \langle \hat{\sigma}, \hat{\sigma}' \rangle \wedge \langle \sigma, \sigma' \rangle \approx_2 \langle \hat{\sigma}, \hat{\sigma}' \rangle$ .
- (4) Zwei Gleichgewichtspunkte  $\langle \sigma, \sigma' \rangle \in \Sigma^1 \times \Sigma^2$  und  $\langle \hat{\sigma}, \hat{\sigma}' \rangle \in \Sigma^1 \times \Sigma^2$  sind austauschbar  $\Leftrightarrow$   
 $\langle \sigma, \hat{\sigma}' \rangle$  ist Gleichgewichtspunkt  $\wedge \langle \hat{\sigma}, \sigma' \rangle$  ist Gleichgewichtspunkt.

Jedoch hat das Gefangenendilemma eine sog. Nash-Lösung, denn die Bedingung, daß jedes Paar von Gleichgewichtspunkten austauschbar ist, ist bei einem Gleichgewichtspunkt (der nicht-kooperativen kollektiven Entscheidung) trivialerweise erfüllt.

Wir können die Situation des Gefangenendilemmas verallgemeinern zu einem nicht-kooperativen n-Personenspiel, in dem die jeweils individuellen Entscheidungen für die dominante Strategie  $\hat{\sigma}_i$  zu einem Ergebnis  $\langle u_1(\langle \hat{\sigma}_1 \rangle), u_2(\langle \hat{\sigma}_1 \rangle), u_3(\langle \hat{\sigma}_1 \rangle), \dots, u_n(\langle \hat{\sigma}_1 \rangle) \rangle$  führen, das jeden einzelnen schlechter stellt, als es



bei einer anderen (kollektiven) Entscheidung  $\Sigma = \langle \sigma_1', \sigma_2', \sigma_3', \dots, \sigma_v' \rangle$  der Fall wäre - m.a.W. in dem individuelle und kollektive Rationalität<sup>1</sup> auseinanderfällt:

$$\begin{aligned} & \exists \langle \sigma_1', \sigma_2', \sigma_3', \dots, \sigma_v' \rangle \in \Sigma^1 \times \Sigma^2 \times \Sigma^3 \times \dots \times \Sigma^v : \\ & \forall i \in K : \exists \hat{\sigma}_i \in \Sigma_i : \\ & \forall \langle \sigma_1, \sigma_2, \sigma_3, \dots, \sigma_v \rangle \in \Sigma^1 \times \Sigma^2 \times \Sigma^3 \times \dots \times \Sigma^v : \\ & [ [\hat{\sigma}_i \neq \sigma_i \Rightarrow u_i(\langle \sigma_1, \sigma_2, \sigma_3, \dots, \sigma_{i-1}, \hat{\sigma}_i, \sigma_{i+1}, \dots, \sigma_v \rangle) > \\ & \quad u_i(\langle \sigma_1, \sigma_2, \sigma_3, \dots, \sigma_{i-1}, \sigma_i, \sigma_{i+1}, \dots, \sigma_v \rangle) \wedge \\ & \quad \wedge u_i(\langle \sigma_1', \sigma_2', \sigma_3', \dots, \sigma_v' \rangle) > u_i(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \dots, \hat{\sigma}_v) ] ] ; \\ & \text{wobei } \#K = v. \end{aligned}$$

Selbstverständlich ist damit auch das Ergebnis jeweils individuell rationalen Verhaltens  $\langle \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \dots, \hat{\sigma}_v \rangle$  ein Gleichgewichtspunkt des PD-Spieles.

Aus der Definition folgt:

$$\begin{aligned} & u_i(\langle \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \dots, \hat{\sigma}_{i-1}, \hat{\sigma}_i, \hat{\sigma}_{i+1}, \dots, \hat{\sigma}_v \rangle) > \\ & u_i(\langle \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \dots, \hat{\sigma}_{i-1}, \sigma_i, \hat{\sigma}_{i+1}, \dots, \hat{\sigma}_v \rangle) \text{ für } \forall i \in K, \sigma_i \neq \hat{\sigma}_i \end{aligned}$$

Solche Interaktionssituationen müssen natürlich nicht symmetrisch sein. Es sind vielfältige Variationen denkbar, und es ist zweckmäßig, unterschiedliche Typen verallgemeinerter Gefangenendilemmas zu unterscheiden.

Die spieltheoretische Analyse solcher Interaktionssituationen hat in den letzten Jahrzehnten wesentlich zu einer Neuorientierung der Spieltheorie beigetragen. Während die Hoffnung, die die wissenschaftliche Öffentlichkeit auf die Entwicklung der Spieltheorie setzte, nämlich ganz im Sinne des Bayesianisierungsprogrammes der Spieltheorie, Kriterien individuell-rationalen Verhaltens auch für Interaktionssituationen beliebiger Art zu erhalten, auf dem Gebiet der Nicht-Konstantsummen-

---

<sup>1</sup> Der Begriff "kollektive Rationalität" wird später (§13 und §17) formal definiert.

spiele weitgehend enttäuscht wurde<sup>1</sup>, bahnte sich eine 'Ethifizierung' der Spieltheorie eigentlich schon mit Nash, Shapley und Braithwaite<sup>2</sup> an. Diese Entwicklung, die uns hier nicht im Detail interessiert, war jedoch von Anfang an mit der Hypothek einer dafür ungeeigneten Präferenztheorie in Form des modernen behavioristischen Nutzenkonzepts belastet.

Harsanyi ist völlig konsequent, wenn er sagt, in einem Gefangenendilemma kann es für rationale Individuen keine andere 'Lösung' als die nicht-kooperative geben.<sup>3</sup> Kommt es (bei rationalen Individuen) zur kooperativen (kollektiven) Entscheidung, handelt es sich eben nur scheinbar um ein Gefangenendilemma, die Nutzenfunktionen sind falsch ermittelt oder vermutet worden. Aber diese Argumentation zeigt auch, auf welchem schwankendem methodischen Boden die Spieltheorie operiert, denn das kommt in große Nähe zu einer generellen Immunisierung spieltheoretischer Kriterien.<sup>4</sup>

---

<sup>1</sup> Vgl. T.C. Schelling, *The strategy of conflict: Prospectus for a re-orientation of game theory*, Conflict resolution 2 (1958)203-264; ders., *The strategy of conflict*, Cambridge 1963; A. Rapoport, *Fights, games, and debates*, Ann Arbor 1960; N. Howard, *Paradoxes of rationality*, Cambridge/London 1971

<sup>2</sup> Vgl. J.F. Nash, *The bargaining problem*, Econometrica 18 (1950) 155-162; L.S. Shapley, *A value for n-person games*, in: Kuhn/Tucker (1953) 307-317; R.B. Braithwaite, *Theory of games as a tool for the moral philosopher*, Cambridge 1955

<sup>3</sup> Vgl. J.C. Harsanyi, *Rationality postulates for bargaining solutions in cooperative and non-cooperative games*, Management science 9 (1962) 141-153 und ders., *A general theory of rational behavior in game situations*, Econometrica 34 (1966) 613-634

<sup>4</sup> Und zwar bei einer deskriptiven wie bei einer normativen Interpretation. Eine deskriptive Theorie ist immun gegenüber empirischen Widerlegungen, wenn sie wahr ist, unabhängig davon, wie die empirische Welt beschaffen ist. Eine normative Theorie ist immun, wenn sie unabhängig von der Richtigkeit normativer Satzklassen gültig ist. In §14 kommen wir auf diesen Punkt noch einmal zurück.

Harsanyi macht einige Andeutungen, wie er sich eine methodische Klärung vorstellt:

Danach ist die Spieltheorie, die er als Generalisierung der individuellen Entscheidungstheorie ansieht, als rein formale Theorie ohne empirischen oder normativen Gehalt anzusehen. Spezifische faktische Annahmen über die Natur der Präferenzen machen die Spieltheorie erst zu einer Theorie mit empirischem bzw. normativem Gehalt.<sup>1</sup> Da diese faktischen Annahmen von Harsanyi nicht diskutiert werden, handelt es sich bisher nur um die Skizze einer methodischen Konzeption. Um sich mit ihr detaillierter auseinanderzusetzen, müßte zuerst der Versuch einer Konkretisierung gemacht werden. Das würde hier zu weit vom Thema abführen.

Anatol Rapoport kommt zu einem entgegengesetzten Schluß aus der - auch experimentellen - Analyse von Interaktionssituationen des PD-Typs: Er hält die Ergebnisse der Spieltheorie nur in Konstantsummen-Spielen für normativ, hier kann das Optimierungsverhalten bestimmt werden:

"even if real conflicts were no more complex than theoretically tractable formal games, game theory would be powerless to prescribe 'optimal strategies' in any but total bi-polar conflicts" <sup>2</sup>.

Begründungen dafür finden sich schon in *Fights, games, and debates* (1960), in der deutschen Übersetzung auf den Seiten 127-267.

"Es scheint notwendig, normative von beschreibender Behandlung strategischer Konflikte zu unterscheiden. Wie wir gesehen haben, ist es, abgesehen von der Theorie des Nullsummen-Zweipersonenspiels und der recht unbestimmten Vorstellung einer 'Lösung' in von Neumanns Sinn, fast unmöglich, die normative Spieltheorie auszuweiten, weil die Kriterien der Rationalität sich bei dem Zusammenprall von individuellen und Gruppennormen verwirren."<sup>3</sup>

---

<sup>1</sup> Vgl. Harsanyi (1966), S.613f und A.Rapoport, *Prisoners' dilemma*, in: Rapoport (1974) 17-34

<sup>2</sup> A. Rapoport, *Game theory as a theory of conflict resolution*, Dordrecht/Boston 1974, Introduction S.7, wobei er mit 'total bi-polar conflicts' Zwei-Personen-Konstantsummen-Spiele meint.

<sup>3</sup> ders., *Kämpfe, Spiele und Debatten. Drei Konfliktmodelle*, Darmstadt 1976, S.243

## PROBLEME NUTZENMAXIMIERENDEN VERHALTENS IN INTERAKTIONSSITUATIONEN

Dieser Selbstbescheidung der Spieltheorie genau entgegengesetzt ist das Programm einer Bayesianisierung der Spieltheorie, d.h. der Versuch, spieltheoretische Kriterien individueller Handlungsrationalität als spezielle Anwendung des Bayesschen NEW-Maximierungskriteriums auf Interaktionssituationen zu konzipieren bzw. zu rechtfertigen.

Die Problematik dieses Ansatzes liegt darin, daß hier Ereignissen subjektive Wahrscheinlichkeiten zugeordnet werden müssen, die durch personale Entscheidungen bedingt sind. Das kann mit zwei epistemischen Merkmalen von spieltheoretisch idealisierten Interaktionsstrukturen in Konflikt geraten: erstens der Annahme, daß alle an der Interaktion Beteiligten wissen, daß alle 'Spieler' rational im Sinne der jeweiligen Regeln rationalen Verhaltens sind, und zweitens der Annahme, daß die Elemente der Interaktionsstruktur - die Anzahl der Spieler, ihre Entscheidungspotentiale  $\Sigma^1$ , die Konsequenzen der möglichen kollektiven Entscheidungen  $\Xi$  und die individuellen Nutzenfunktionen  $u_i$  über  $\Xi$  - den Beteiligten bekannt sind, und daß jeder auch weiß, das dem so ist.

Diese epistemischen Annahmen der üblichen spieltheoretischen Modelle rechtfertigen (schon bei Konstantsummen-Spielen) Kombinationen gemischter Strategien als Lösung. Aus entscheidungstheoretischer Sicht ist das nicht befriedigend, denn es mögen zwar in einigen Entscheidungssituationen zwei Entscheidungen den gleichen Nutzenerwartungswert haben, so daß das NEW-Maximierungskriterium indifferent bei diesen beiden Entscheidungen ist, aber die Empfehlung, nach einem Zufallsexperiment mit einer bestimmten Wahrscheinlichkeit die eine und mit einer bestimmten Wahrscheinlichkeit die andere Entscheidung zu wählen, kann aufgrund eines ent-

scheidungstheoretischen Rationalitätskriterium nicht gerechtfertigt werden. Die 'Lösung' von Interaktionssituationen durch Kombinationen gemischter Strategien kann nicht mehr als Konsequenz einer Theorie unabhängiger individueller Entscheidungsrationaliät interpretiert werden. In diesem Sinne erscheint die spieltheoretische 'Empfehlung' von Gleichgewichtspunkten in gemischten Strategien als eine Empfehlung, die sich nicht an den Einzelnen, sondern an die Gruppe richtet.

Die Problematik NEW-maximierenden Verhaltens in Interaktionssituationen läßt sich an folgendem Beispiel veranschaulichen:

		B	
		$\sigma_{21}$	$\sigma_{22}$
A	$\sigma_{11}$	$\langle 2, 2 \rangle$	$\langle 3, 3 \rangle$
	$\sigma_{12}$	$\langle 3, 3 \rangle$	$\langle 1, 1 \rangle$

Diese Interaktionsstruktur ist keine Konfliktsituation im üblichen Sinne, denn es besteht eine völlige Übereinstimmung der Interessen der Beteiligten:

Beide sind indifferent zwischen den kollektiven Entscheidungen  $\langle \sigma_{12}, \sigma_{21} \rangle$  und  $\langle \sigma_{11}, \sigma_{22} \rangle$  und stimmen auch sonst in ihren Präferenzen über die Menge der kollektiven Entscheidungen überein:

$$\langle \sigma_{12}, \sigma_{21} \rangle \underset{1}{\approx} \langle \sigma_{11}, \sigma_{22} \rangle \underset{1}{>} \langle \sigma_{11}, \sigma_{21} \rangle \underset{1}{>} \langle \sigma_{12}, \sigma_{22} \rangle$$

$$\langle \sigma_{12}, \sigma_{21} \rangle \underset{2}{\approx} \langle \sigma_{11}, \sigma_{22} \rangle \underset{2}{>} \langle \sigma_{11}, \sigma_{21} \rangle \underset{2}{>} \langle \sigma_{12}, \sigma_{22} \rangle.$$

Ist diese Interaktionsstruktur jedoch völlig symmetrisch, und sind die beiden o.g. epistemischen Annahmen erfüllt, so dürfen wir annehmen, daß die beiden Betroffenen ziemlich ratlos sind. Beide wissen, daß ihre Optimierungsüberlegungen aufgrund der Interaktionsstruktur zum gleichen Ergebnis führen müßten, d.h. zu den kollektiven Entscheidungen  $\langle \sigma_{11}, \sigma_{21} \rangle$  oder  $\langle \sigma_{12}, \sigma_{22} \rangle$ , oder zu einer Kombination gemischter Strategien.

Dieses Wissen kommt jedoch merkwürdigerweise in Konflikt mit der Gleichgewichtstheorie. Das kann man sich folgendermaßen klarmachen:

Für beliebiges  $\alpha \in [0,1]$  kann die gemischte Strategie  $(\alpha \cdot \sigma_{11} \text{ \& } (1-\alpha) \cdot \sigma_{12})$  keine Gleichgewichtsstrategie sein, denn angenommen, es wäre die Gleichgewichtsstrategie der ersten Person, so wäre sie aufgrund der Symmetrie der Interaktionsstruktur auch Gleichgewichtsstrategie der zweiten Person. Unsere epistemischen Voraussetzungen lassen die zweite Person jedoch Kenntnis von der Gleichgewichtsstrategie der ersten Person haben, damit ist aber  $(\alpha \cdot \sigma_{21} \text{ \& } (1-\alpha) \cdot \sigma_{22})$  nicht mehr Gleichgewichtsstrategie der zweiten Person.

Obwohl diese Interaktionsstruktur ohne Interessenkonflikt zwei Gleichgewichtspunkte in reinen Strategien hat, können individuell rationale Individuen unter naheliegenden epistemischen Annahmen die Gleichgewichtspunkte nicht erreichen.

Dieses Beispiel mag als nicht weiter 'tragisch' für die Theorie NEW-maximierenden Verhaltens in Interaktionssituationen empfunden werden, da beide Individuen indifferent den beiden kollektiven Gleichgewichtsentscheidungen gegenüber sind - das Gefangenendilemma gibt Anlaß zu vergleichsweise gravierenderen Bedenken. Sobel argumentiert jedoch, daß selbst nicht-symmetrische Interaktionssituationen ohne interpersonellen Konflikt mit einem pareto-optimalen Gleichgewichtspunkt keine 'Lösung' für individuelles (unabhängiges) NEW-maximierendes Verhalten bieten.<sup>1</sup> Er führt das folgende Beispiel an<sup>2</sup>:

		B	
		$\sigma_{21}$	$\sigma_{22}$
A	$\sigma_{11}$	$\langle 2, 2 \rangle$	$\langle 4, 4 \rangle$
	$\sigma_{12}$	$\langle 3, 3 \rangle$	$\langle 1, 1 \rangle$

<sup>1</sup> Vgl. J.H. Sobel, *Interaction problems of utility maximizers*, Canadian journal of philosophy 4 (1975), S.68ff

<sup>2</sup> Wir passen die Notation Sobels an die unsrige an.

Sobel meint, es sei für gutinformierte Nutzenmaximierer in dieser Interaktionssituation unmöglich, eine Strategie - sei sie nun rein oder gemischt - als rational auszuzeichnen. Die rationale Entscheidung ist abhängig von der Erwartung, was der andere tun wird:

Erwartet A, daß sich B für  $\sigma_{21}$  entscheidet, ist es für A rational, sich für  $\sigma_{12}$  zu entscheiden; erwartet A, daß sich B für  $\sigma_{22}$  entscheidet, ist es für A rational, sich für  $\sigma_{11}$  zu entscheiden. Erwartet B, daß sich A für  $\sigma_{11}$  entscheidet, ist es rational für B, sich für  $\sigma_{22}$  zu entscheiden; erwartet B, daß sich A für  $\sigma_{12}$  entscheidet, ist es rational für B sich für  $\sigma_{21}$  zu entscheiden. Erwartet A, daß sich B für die gemischte Strategie  $\langle 0,75\sigma_{21} \text{ \& } 0,25\sigma_{22} \rangle$  entscheidet, so ist A zwischen  $\sigma_{11}$  und  $\sigma_{12}$  indifferent, da diese Entscheidungen dann den gleichen Nutzenerwartungswert für ihn haben. Erwartet A, daß sich B für eine gemischte Strategie  $\langle \rho \sigma_{21}, (1-\rho)\sigma_{22} \rangle$  mit  $\rho \in [0, 0,75)$  entscheidet, ist  $\sigma_{11}$  rational für A, erwartet er, daß B die gemischte Strategie  $\langle \rho \sigma_{21}, (1-\rho)\sigma_{22} \rangle$  mit  $\rho \in (0,75, 1]$  wählt, ist für A  $\sigma_{12}$  rational.

Die These lautet nun: Ideale Nutzenmaximierer mit den genannten epistemischen Voraussetzungen könnten solche Erwartungen nicht haben.

Rationales Verhalten im Sinne des Bayesschen Entscheidungskriteriums maximiert für jede Person  $i \in K$   $\sum \cup_1 (\xi_{K\lambda}) \cdot p_1(\varepsilon_\lambda)$ . Im Zwei-Personen-Spiel ist  $\varepsilon_\lambda \in \Sigma^2$  für  $i=1$  und  $\varepsilon_\lambda \in \Sigma^1$  für  $i=2$ . Die Erwartungen über das Entscheidungsverhalten des anderen  $\rho_1, \rho_2$  richten sich danach, von welcher Entscheidung (rein oder gemischt) angenommen werden kann, daß sie für den anderen rational ist. Sobels These ist nun, daß eine Annahme darüber, welches Verhalten für den jeweils anderen rational ist und damit eine Erwartung über dessen Entscheidungsverhalten in einer Gesellschaft idealer rationaler Nutzenmaximierer bestimmte gemeinsame Überzeugungen darüber

voraussetzt, welche Entscheidung im eigenen Interesse ist. Das aber kann nur dann bestimmt werden, wenn klar ist, welche Entscheidung im Interesse des anderen ist - es ergibt sich so ein Begründungszirkel. Da die individuellen Handlungen nach Voraussetzung kausal unabhängig sind, gibt es in vielen Fällen keine Möglichkeit, die individuell rationalen Entscheidungen in einer Gesellschaft idealer rationaler Nutzenmaximierer anzugeben. Merkmal der Interaktionssituationen, die für eine Gesellschaft idealer rationaler Nutzenmaximierer nicht 'lösbar' sind, ist demnach, daß eine wechselseitige Abhängigkeit der je individuell rationalen (da NEW-maximierenden) Entscheidung von der Entscheidung anderer an der Interaktionssituation beteiligter Personen besteht.<sup>1</sup> Das Gefangenendilemma gehört daher nicht zu den in diesem Sinne 'unlösbaren' Interaktionssituationen, denn hier ist die nicht-kooperative Entscheidung jeweils rational, unabhängig davon, welche Entscheidung der andere Gefangene trifft (und damit auch, welche Entscheidung der andere für rational hält).

Eine andere Kritik der Theorie NEW-maximierenden Verhaltens in Interaktionssituationen stellt Gauthiers Behandlung des rationalen Egoismus dar. Im Gegensatz zu Sobel gibt es für Gauthiers in beliebigen Zwei-Personen-Spielen eine 'Lösung' auch für ideale rationale

---

<sup>1</sup> Hier gibt es eine interessante Parallele zum Ansatz der Metaspieltheorie. Howard hat 1966 gegenüber Kritikern ein 'existentialist axiom' eingeführt, daß er in Howard (1971) unter dem etwas irreführenden Titel 'The free will paradox' folgendermaßen zusammenfaßt: "If certain factors affect a subject's behavior, and he is conscious of this, he may always choose to make his behavior obey any function of the set of possible values of those factors to the set of his alternative choices. It is true that this implies that a conscious decision maker can always choose to disobey any theory predicting his behavior. We may say that he can always 'transcend' such a theory"

N. Howard, *Paradoxes of rationality*, Cambridge/London 1971, S.64  
Vgl.a. ders., *The theory of metagames*, General systems - Yearbook of the society for gen. systems res. 11 (1966) 167-186



(NEW-maximierende) Personen (bei den gleichen epistemischen Voraussetzungen)<sup>1</sup>, da es für beliebige Zwei-Personen-Spiele mit endlichen Mengen reiner Strategien mindestens einen Gleichgewichtspunkt gemischter Strategien gibt<sup>2</sup>, der nicht von einem anderen Gleichgewichtspunkt dominiert wird.

Gauthier hat an anderer Stelle<sup>3</sup> die Empfehlung NEW-maximierenden Verhaltens u.a. am Beispiel des Gefangenendilemmas kritisiert, da ihre Befolgung auch in Zwei-Personen-Interaktionssituationen zu inakzeptablen Konsequenzen führen kann. Aber jene Kritik ist eine normative, während es in *The impossibility of rational egoism* um eine logische Frage geht: Liefern die Axiome NEW-maximierenden Verhaltens zusammen mit bestimmten epistemischen Annahmen konsistente Verhaltensempfehlungen? Und die Antwort lautet: In allen Zwei-Personen-Spielen ist das so, aber in einigen Drei-Personen-Spielen führt NEW-maximierendes individuelles Verhalten zu inkonsistenten Verhaltensempfehlungen.

Die Argumentation läßt sich in folgender Weise rekonstruieren:

- (1) Ein Egoist verhält sich NEW-maximierend. NEW-maximierendes Verhalten ist eine notwendige (aber vielleicht keine hinreichende) Bedingung für Egoismus.
- (2) Eine Verhaltensregel schreibt jedem Handelnden in jeder Interaktionssituation genau eine Entscheidung vor.

---

<sup>1</sup> Vgl. D.P. Gauthier, *The impossibility of rational egoism*, *Journal of philosophy* 14 (1974) 439-456 und ders., *Reason and maximization*, *Canadian journal of philosophy* 4 (1975) 411-433

<sup>2</sup> Wie Nash sehr elegant bewiesen hat. Vgl. J.F. Nash, *Non-cooperative games*, *Annals of mathematics* 54 (1951) 286-295, vgl.a. Luce/Raiffa (1957) 5.8. und Appendix 2

<sup>3</sup> Vgl. Gauthier, *Morality and advantage*, *Philosophical review* 76 (1976) 460-475

- (3) Eine vollständige Verhaltensregel ordnet jeder Interaktionssituation eine individuell empfohlene Handlung zu.
- (4) Eine Kombination  $\Sigma$  individueller Strategien  $\sigma_1 \in \Sigma^1$ ,  $\sigma_2 \in \Sigma^2$ , die von einer Verhaltensregel  $V$  empfohlen werden, muß ein Gleichgewichtspunkt sein.
- (5) Eine Kombination individueller Strategien  $\Sigma$ , die von einer Regel individuell rationalen Verhaltens empfohlen werden, darf nicht von einer erreichbaren Strategiekombination  $\hat{\Sigma}$  dominiert werden.
- (6) Wenn eine Verhaltensregel  $V$  dem Handelnden in einer Interaktionssituation  $\Gamma$  bestimmte Strategien empfiehlt, so darf  $V$  in einer gegenüber  $\Gamma$  bzgl.  $V$  reduzierten Interaktionssituation  $\hat{\Gamma}$  nur solche Strategien empfehlen, die sie auch in  $\Gamma$  empfohlen hat.

Def.:  $\Gamma$  ist bzgl.  $V$  stabil bei Reduktion  $\Leftrightarrow$   
 $\Gamma$  erfüllt Axiom (6) (d.h. (6) gilt für alle gegenüber  $\Gamma$  reduzierten Interaktionssituationen)

Def.:  $\hat{\Gamma}$  ist eine gegenüber  $\Gamma$  bzgl.  $V$  reduzierte Interaktionssituation  $\Leftrightarrow$   
 $\hat{\Gamma}$  geht aus  $\Gamma$  dadurch hervor, daß eine oder mehrere der empfohlenen Strategien als befolgt angenommen werden.

$\hat{\Gamma}$  ist gegenüber  $\Gamma$  bzgl.  $V$  einfach reduziert  $\Leftrightarrow$

$\hat{\Gamma}$  geht aus  $\Gamma$  dadurch hervor, daß eine der empfohlenen Strategien als gegeben angenommen wird.

$\hat{\Gamma}$  ist gegenüber  $\Gamma$  n-fach reduziert  $\Leftrightarrow$

es eine Folge  $\Gamma, \Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_{n-1}, \hat{\Gamma}$  gibt, so daß für  $\forall v \in \{1, 2, 3, \dots, n-1\}$  gilt:  $\Gamma_v$  ist gegenüber  $\Gamma_{v-1}$  bzgl.  $V$  einfach reduziert und  $\hat{\Gamma}$  ist gegenüber  $\Gamma_{v-1}$  bzgl.  $V$  einfach reduziert.

Korollar: Wenn  $\hat{\Gamma}$  gegenüber  $\Gamma$  n-fach reduziert ist und  $\Gamma$  ein  $m$ -Personenspiel ist, so ist  $\hat{\Gamma}$  ein  $m$ - $n$ -Personen-Spiel.

Def.: Ein Gleichgewichtspunkt  $\Sigma$  ist erreichbar,  $\Leftrightarrow$   
seine Wahl ist aufgrund von Symmetrieüberlegungen nicht ausgeschlossen.

Die Unmöglichkeit des rationalen Egoismus läßt sich dann so formulieren:

Es gibt keine vollständige (3), egoistische (1) Verhaltensregel (2), die nicht (4) $\vee$ (5) $\vee$ (6) verletzt.

Beweis durch Beispiel:

Da es sich in diesem Beispiel um eine 3x2-Interaktion handelt, brauchen wir zu seiner Darstellung zwei 2x2-Matrizen: (Das Entscheidungsverhalten der dritten Person nehmen wir dabei als Parameter):

Die 3x2-Interaktionssituation  $\Gamma$

$\sigma_{31}$	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 8, 8, 8 \rangle$	$\langle 12, 4, 6 \rangle$
$\sigma_{12}$	$\langle 4, 12, 6 \rangle$	$\langle 12, 12, 4 \rangle$

$\sigma_{32}$	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 16, 16, 0 \rangle$	$\langle 12, 3, 7 \rangle$
$\sigma_{12}$	$\langle 3, 12, 7 \rangle$	$\langle 3, 3, 13 \rangle$

Diese beiden Matrizen zeigen, daß  $\langle \sigma_{11}, \sigma_{21}, \sigma_{31} \rangle$  die einzige kollektiven Gleichgewichtsentscheidung in  $\Gamma$  ist.

Denn

(1) Wie der Vergleich der Elemente der ersten Spalte der oberen Matrix zeigt,

$$\text{gilt: } \langle \sigma_{11}, \sigma_{21}, \sigma_{31} \rangle \underset{1}{>} \langle \sigma_{12}, \sigma_{21}, \sigma_{22} \rangle$$

(2) Wie der Vergleich der Elemente der ersten Zeile der oberen Matrix zeigt,

$$\text{gilt: } \langle \sigma_{11}, \sigma_{21}, \sigma_{31} \rangle \underset{2}{>} \langle \sigma_{11}, \sigma_{22}, \sigma_{31} \rangle$$

(3) Wie der Vergleich der oberen und der unteren Matrix zeigt,

$$\text{gilt: } \langle \sigma_{11}, \sigma_{21}, \sigma_{31} \rangle \underset{3}{>} \langle \sigma_{11}, \sigma_{21}, \sigma_{32} \rangle$$

(4) Es gibt keine andere kollektive Entscheidung

$$\langle \sigma_{1\kappa}, \sigma_{2\lambda}, \sigma_{3\mu} \rangle, \text{ für die gilt:}$$

$$\langle \sigma_{1\kappa}, \sigma_{2\lambda}, \sigma_{3\mu} \rangle \underset{1}{>} \langle \sigma_{1\chi}, \sigma_{2\lambda}, \sigma_{3\mu} \rangle \wedge$$

$$\langle \sigma_{1\kappa}, \sigma_{2\lambda}, \sigma_{3\mu} \rangle \underset{2}{>} \langle \sigma_{1\kappa}, \sigma_{2\psi}, \sigma_{3\mu} \rangle \wedge$$

$$\langle \sigma_{1\kappa}, \sigma_{2\lambda}, \sigma_{3\mu} \rangle \underset{3}{>} \langle \sigma_{1\kappa}, \sigma_{2\lambda}, \sigma_{3\omega} \rangle$$

wobei  $\kappa, \lambda, \mu \in \{1, 2\}$ ;  $\chi, \psi, \omega \in \{1, 2\}$  und  $\kappa \neq \chi, \lambda \neq \psi, \mu \neq \omega$ .

Betrachten wir jedoch die einfach reduzierte Interaktionssituation, die dadurch aus  $\Gamma$  hervorgeht, daß wir  $\sigma_{31}$  als gegeben annehmen, so ergibt sich  $\hat{\Gamma}$ :

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 8, 8 \rangle$	$\langle 12, 4 \rangle$
$\sigma_{12}$	$\langle 4, 12 \rangle$	$\langle 12, 12 \rangle$

In  $\hat{\Gamma}$  gibt es zwei kollektive Gleichgewichtsentscheidungen  $\langle \sigma_{11}, \sigma_{21} \rangle$  und  $\langle \sigma_{12}, \sigma_{22} \rangle$ , wobei  $\langle \sigma_{12}, \sigma_{22} \rangle \underset{1}{>} \langle \sigma_{11}, \sigma_{21} \rangle \wedge \langle \sigma_{12}, \sigma_{22} \rangle \underset{2}{>} \langle \sigma_{11}, \sigma_{21} \rangle$ .

Damit ist  $\langle \sigma_{12}, \sigma_{22} \rangle$  kollektiv besser als  $\langle \sigma_{11}, \sigma_{21} \rangle$ ;  $\langle \sigma_{11}, \sigma_{21} \rangle$  wird von  $\langle \sigma_{12}, \sigma_{22} \rangle$  dominiert und das widerspricht Postulat (5).  $\langle \sigma_{12}, \sigma_{22} \rangle$  ist trivialerweise auch stabil bei Reduktion (Postulat 6).

Das Beispiel zeigt, daß es keine vollständige 'egoistische' Verhaltensregel geben kann, die (4), (5) und (6) erfüllt. Es kann natürlich eine nicht-vollständige egoistische Verhaltensregel geben, die (4), (5) und (6) erfüllt.

Da es auch für beliebige n-Personen-Interaktionssituationen eine kollektive Gleichgewichtsentscheidung in den gemischten individuellen Strategien gibt, bleiben die jeweils empfohlenen individuellen Gleichgewichtsstrategien auch in allen gegenüber der ursprünglichen

reduzierten Interaktionssituation Gleichgewichtsstrategien. Aber es kann vorkommen, daß eine, von keiner anderen kollektiven Gleichgewichtsentscheidung in der ursprünglichen Interaktionssituation dominierte, kollektive Entscheidung in einer reduzierten Interaktionssituation von einer kollektiven Gleichgewichtsentscheidung dominiert wird. Das ist der spieltheoretische Kern dieser Egoismus-Kritik. Sie ist insofern immanent, als nur die innere Konsistenz einer egoistischen Handlungsempfehlung überprüft wird, oder wie es Gauthier formuliert:

"I am about to demonstrate, what I had long believed false. Not only intuitively, but also in reflection, it had seemed to me that egoism, whatever its defects, provided a systematic comprehensive basis of action. I am about to demonstrate that it does not."<sup>1</sup>

Gauthiers Argumentation gegen den rationalen Egoismus hat zwei Schwachpunkte.

Eine Schwäche liegt darin, daß die Eindeutigkeit für Verhaltensregeln postuliert wird. Diese Forderung ist zu stark, aber, wie eine Prüfung der Argumentation zeigt, auch entbehrlich, denn (2) ist für den Beweis des Theorems nicht erforderlich.

Ein gravierenderes Problem aber besteht in folgendem: Der Zusammenhang von NEW-maximierendem Verhalten und der Wahl der dominanten Gleichgewichtsstrategie ist nicht so eindeutig, wie es Gauthier erscheint. Auch dann nicht, wenn aus NEW-maximierendem Verhalten nur unter der Bedingung die individuelle Entscheidung für den dominanten Gleichgewichtspunkt folgen soll, daß dieser Gleichgewichtspunkt erreichbar ist.

---

<sup>1</sup> Gauthier (1974), S.440

	$\sigma_{21}$	$\sigma_{22}$	$\sigma_{23}$
Nach Gauthier wäre in einer Interaktionssituation, wie sie die nebenstehende Matrix darstellt, die kollektive Gleichgewichtsentscheidung $\langle \sigma_{12}, \sigma_{22} \rangle$ zwar dominiert durch die Gleichgewichts-	$\sigma_{11}$ <5,5>	<1,1>	<1,1>
	$\sigma_{12}$ <1,1>	<4,4>	<1,1>
	$\sigma_{13}$ <1,1>	<1,1>	<5,5>

entscheidungen  $\langle \sigma_{11}, \sigma_{21} \rangle$  und  $\langle \sigma_{13}, \sigma_{23} \rangle$ , aber nicht erreichbar dominiert, denn es gibt ohne Kommunikation keine Möglichkeit, eine von beiden dominanten Gleichgewichtsentscheidungen zu garantieren. Würden sich beide Personen ausschließlich an die Empfehlung halten, eine dominante Gleichgewichtsstrategie zu wählen, so ergäbe sich (aufgrund der Symmetrie der Situation) ein individueller Nutzenerwartungswert von 3 für jeden der beiden Beteiligten, wenn keine weiteren Informationen verfügbar sind. Denn man kann dann annehmen, daß jede Person mit der Wahrscheinlichkeit 0,5 eine der beiden dominanten Gleichgewichtspunkte anstrebt, und damit die vier kollektiven Entscheidungen  $\langle \sigma_{11}, \sigma_{21} \rangle$ ,  $\langle \sigma_{13}, \sigma_{21} \rangle$ ,  $\langle \sigma_{13}, \sigma_{23} \rangle$ ,  $\langle \sigma_{11}, \sigma_{23} \rangle$  bzw. die 'Auszahlungen'  $\langle 5,5 \rangle$ ,  $\langle 1,1 \rangle$ ,  $\langle 5,5 \rangle$ ,  $\langle 1,1 \rangle$  gleichwahrscheinlich werden. Nach Gauthier wäre deshalb  $\sigma_{12}$  bzw.  $\sigma_{22}$  die jeweils beste individuelle Entscheidung. Schon bei einer unbedeutenden Asymmetrie der Interaktionssituation jedoch, z.B. derjenigen, daß für  $\langle \sigma_{13}, \sigma_{23} \rangle$  die Auszahlung  $\langle 5,001,5 \rangle$  erfolgt, könnte man argumentieren, daß damit - ceteris paribus - nicht nur für die eine Person  $\sigma_{13}$ , sondern auch für die andere Person  $\sigma_{23}$  zur besten Entscheidung würde, obwohl sich ihre Auszahlungsfunktion nicht geändert hat. Andere Asymmetrien, z.B. bestehende Konventionen, die sich in den Bewertungsfunktionen nicht äußern müssen, können die gleiche Rolle spielen.

Unbeschadet dieser Einschränkung hat die Annahme, daß NEW-maximierendes Verhalten zur Wahl der

<sup>1</sup> Zum Problem der Koordination individueller Entscheidungen in Interaktionssituationen, die nicht von Interessenkonflikt, sondern von weitgehender Übereinstimmung der individuellen Interessen geprägt sind, vgl. die mit spieltheoretischen Hilfsmitteln durchgeführte sprachphilosophische Analyse von David Lewis, *Konventionen*, Berlin/New York 1975; T.C. Schelling, *The strategy of conflict*, Cambridge (Mass.) 1963, Teil 2 und ders., *Game theory and the study of ethical systems*, *Journal of conflict resolution* 12 (1968) 34-44



und argumentiert folgendermaßen:

"Since we are both egoists, I know we would both do best if we were both to show heads. Since you know what I know, you know this. Therefore I know that we both know that we would both do best if we were both to show heads. And you know that we both know... This mutuality of knowledge and of advantage ensures that your decision will parallel mine; I may then treat the situation as if my decision were a joint decision. Therefore, as an egoist, I should show heads."<sup>1</sup>

Wäre die Interaktionssituation von folgender Art gewesen

	'You show heads'	'You show tails'
'I show heads'	<3,3>	<1,4>
'I show tails'	<4,1>	<2,2>

hätte Gauthier mit den gleichen Sätzen begründet, daß rationale Egoisten beide 'show heads', bzw. daß im Gefangenendilemma rationale Egoisten kooperativ entscheiden würden. Andererseits ist aber die Annahme, daß rationale Egoisten in einer Situation von der Art des Gefangenendilemmas nicht kooperativ entscheiden würden, ja gerade Grundlage für die normative Kritik individuell nutzenmaximierenden Verhaltens.<sup>2</sup>

NEW-maximierendes Verhalten muß nicht egoistisch sein,<sup>3</sup> aber egoistisches Verhalten ist NEW-maximierend, und individuell NEW-maximierend ist in einer Interaktionssituation vom Typ des Gefangenendilemmas die nicht-kooperative Entscheidung.

<sup>1</sup> Gauthier (1974), S.448

<sup>2</sup> So auch Gauthier selbst an anderer Stelle, vgl. Gauthier (1976)

<sup>3</sup> Vgl. B. Medlin, *Ultimate principles and ethical egoism*, Australasian journal of philosophy 35 (1957) 111-118, dagegen J. Kalin, *In defense of egoism*, in: Gauthier (Hg.), *Morality and rational self-interest*, Eaglewood Cliffs, N.J. 1970, 64-87; auch H. Sidgwick, *The method of ethics*, London 1907, dagegen G.E. Moore, *Principia Ethica*, Cambridge 1903



Es sind mehrere Versuche einer Bayesianisierung der Spieltheorie gemacht worden, deren z.T. großer mathematische Aufwand<sup>1</sup> jedoch nicht darüber hinwegtäuschen kann, daß ein befriedigendes Konzept dieser Art bisher nicht vorliegt. Eine erfolgreiche Bayesianisierung der Spieltheorie würde diese zu einer Subdisziplin der Entscheidungstheorie machen.<sup>2</sup>

#### DIE 'ETHIFIZIERUNG' DER THEORIE RATIONALEN VERHALTENS IN INTERAKTIONSSITUATIONEN ALS ALTERNATIVE

Als eine Alternative zur Bayesianisierung der Spieltheorie bietet sich eine 'Ethifizierung'<sup>3</sup> der Theorie rationalen Verhaltens in Interaktionssituationen an. Diese Tradition nimmt den Anspruch eines normativen Status der rationalen Entscheidungstheorie ernst.

---

<sup>1</sup> Vgl. z.B. W. Armbruster, *Spiele mit unbekanntem Nutzenfunktionen und ihre Gleichgewichtspunkte*, Heidelberg 1978; W. Armbruster, W. Böge, *Bayesian game theory*, in: O. Moeschlin/D. Pallaschke (Hg.), *Game theory and related topics*, Amsterdam/New York/Oxford 1979, 17-28

<sup>2</sup> Besonders Harsanyi hat sich um eine Bayesianisierung der Spieltheorie bemüht. Vgl. J.c. Harsanyi, *A general theory of rational behavior in game situations*, *Econometrica* 34 (1966) 613-634, ders., *Games with incomplete informations played by Bayesian players*, *Management science* 14 (1967/68) 159-182, ders., *The tracing procedure: A bayesian approach to defining a solution for n-person non-cooperative games*, *International journal of game theory* 5 (1975) 61-94, und ders., *A new solution concept for both cooperative and non-cooperative games*, Opladen 1979. Eine überzeugende und detaillierte Kritik von Harsanyis 'tracing procedure' gibt A. Bjerring, *The 'tracing procedure' and a theory of rational interaction*, in: Hooker/Leach/McClennen (1978) Bd.I 1-54. Zur Bayesianisierung der Spieltheorie vgl. a. die Überlegungen von W. Spohn (1982).

<sup>3</sup> s. nächste S.

In den letzten beiden Jahren waren beide Entwicklungen zu beachten, die einer Ethifizierung und die einer Bayesianisierung der Spieltheorie.

Am Beispiel des Voters' Paradox lassen sich beide Ansätze verdeutlichen. Unter naheliegenden Vermutungen über die Gestalt individueller Nutzenfunktionen ist die individuelle Teilnahme an Wahlen nicht NEW-maximierend, also i.a. irrational. Überlegen wir uns, welche Voraussetzungen gegeben sein müßten, um die Teilnahme für eine Person  $\uparrow$  NEW-maximierend zu machen.

Angenommen, es gibt 12 Millionen Wahlberechtigte, von denen, wie die Statistiken der vergangenen Wahlen zeigen, jeweils etwa 10 Millionen wählen. Wie groß ist dann die Wahrscheinlichkeit dafür, daß meine Stimme wahlentscheidend wird? Das hängt ganz davon ab, welche subjektive Wahrscheinlichkeitsverteilung man über die möglichen Wahlergebnisse hat, und welche Konstellationen bei welchem Wahlergebnis möglich sind. Machen wir es uns einfach: Es stehen sich nur zwei Parteien gegenüber, und wir sind uns ganz sicher, daß keine über 60% und unter 40% der Stimmen erhalten wird; zwischen 40% und 60% halten wir aber jedes Ergebnis für gleichwahrscheinlich. Damit beträgt die Wahrscheinlichkeit dafür, daß meine Stimme die Wahl entscheidet  $0,5 \cdot 10^{-6}$ . Nehmen wir darüber hinaus idealisierend an, daß es ausschließlich darauf ankommt, welche Partei mehr Stimmen hat, und nicht wie groß der Unterschied ist, so müßte im intrapersonellen kardinalen Nutzenvergleich die Unannehmlichkeit zum Wahl-

---

<sup>3</sup> Viele Spieltheoretiker sind heute der Ansicht, daß dies die Entwicklung der Spieltheorie in den letzten Jahren prägte und in Zukunft weiter prägen wird. Vgl. die Aufsatzsammlungen Hooker/Leach/McClennen (1978) und Gottinger/Leinfellner (Hg.), *Decision theory and social ethics*, Dordrecht 1978.

lokal zu gehen, eventuell Schlange zu stehen, den Kaffee-Plausch bei Tante Frieda vielleicht frühzeitig abubrechen etc., für den individuellen Nutzen einen geringeren Unterschied machen als die Hälfte eines Millionstels der Nutzendifferenz, die sich für die betreffende Person aus den beiden möglichen Konsequenzen der Wahl ergibt. Es ist nicht anzunehmen, daß das regelmäßig für etwa 80% der Wahlberechtigten zutrifft. Die meisten Wähler sind jedoch durchaus davon überzeugt, sich rational zu verhalten.

Ja auch die ökonomische Theorie der Politik, die neue politische Ökonomie<sup>1</sup>, die einige interessante Ergebnisse durch die Anwendung entscheidungstheoretischer Modelle zur Analyse politischer und organisatorischer Entscheidungen erzielt hat, möchte dieses für jede Demokratie notwendige Verhalten nicht gerne als Ausdruck einer weitverbreiteten Entscheidungsirrationalität interpretiert wissen. Daher der Titel Voters' Paradox.

Nun bieten sich zwei Lösungen des Voters' Paradox an. Erstens könnte man der Wahlhandlung als solcher einen intrinsischen Wert zuweisen. Wenn jedoch allein die Tatsache, daß eine bestimmte Person  $v$  zur Wahl geht, im Rahmen einer bestimmten Interpretation der rationalen Entscheidungstheorie ausreichen würde, um bei einigen idealisierenden Annahmen (wohlüberlegte Entscheidung etc.) die Existenz einer Nutzenfunktion  $u_v$  zu begründen, bezüglich der die Entscheidung zur Wahl zu gehen NEW-maximierend ist, so ist die rationale

---

<sup>1</sup> Bahnbrechend für die ökonomische Theorie der Politik bzw. der Neuen Politischen Ökonomie war die Arbeit von M. Olson, *The logic of collective action*, Cambridge 1965; vgl. dazu auch die kritische Darstellung von B. Barry, *Neue Politische Ökonomie*, Frankfurt 1975 und die Monographien von Downs (1968), Riker/Ordeshook (1973), Kirsch (1974), Frey (1977) und Meyer-Krahmer (1979)

Entscheidungstheorie bei dieser Interpretation keine gehaltvolle Theorie: sie könnte dann kein Kriterium rationalen Verhaltens, sondern nur eine neue Beschreibung menschlichen Verhaltens, in der solche Termini wie Nutzen und subjektive Wahrscheinlichkeit vorkommen, liefern. Gibt es jedoch ein von der konkreten Entscheidung unabhängiges Verfahren, um die subjektive Bewertung des Wählens zu bestimmen, und läßt sich mit diesem Verfahren zeigen, daß die meisten Personen eine entsprechende Nutzenbewertung dieser Handlung haben, so wäre damit das Paradoxon behoben.

Wenn man aber die Wähler fragt, weshalb sie zur Wahl gehen, werden ihre Antworten diese Interpretation kaum stützen. Die meisten werden ihr Verhalten mit Generalisierbarkeitskriterien unterschiedlicher Art begründen:

- "Wenn alle Anhänger meiner Partei aus Bequemlichkeit nicht wählen würden, verlöre die Partei diese Wahl."
- "Wenn alle Bürger einer Demokratie aus Bequemlichkeit nicht wählen würden, ginge die Demokratie rasch zugrunde."

etc.

Läßt sich diese Argumentation, die ja gerade die Vernünftigkeit, die Rationalität dieses Verhaltens begründen soll, mit der Bayesschen Theorie rationalen Verhaltens vereinbaren? Mir scheint, das ist nur dann möglich, wenn das Konzept Bayesscher Entscheidungs rationalität seines theoretischen (normativen resp. empirischen resp. begriffsexplikativen) Gehaltes im oben erwähnten Sinne fast völlig entkleidet wird, das aber ist - auch unter methodologischen Gesichtspunkten - eine unbefriedigende Lösung.

Jedoch lassen sich die üblichen Begründungen, die sich auf Generalisierbarkeitskriterien beziehen, mit der Theorie kooperativer Spiele sehr gut in Einklang bringen. In einem kooperativen Spiel, d.h. in einer Interaktions-

Situation, in der erstens jeder Informationsaustausch zwischen den Beteiligten vor der Entscheidung unverfälscht möglich ist, zweitens alle Abmachungen bindend sind und durch Merkmale der Interaktion selbst (Regeln des Spieles) sanktioniert sind, und drittens die subjektiven Bewertungsfunktionen der Beteiligten von den Abmachungen und Verhandlungen vor der Entscheidung nicht beeinflußt werden<sup>1</sup>, können sich die Personen auch auf Entscheidungsregeln einigen, die Generalisierbarkeitsargumenten der o.g. Art gerecht werden, denn es ist zu vermuten, daß nur die Entscheidungsregeln konsensfähig sind, die unterschiedliche Behandlung von Personen von ethisch relevanten oder relevant erscheinenden Unterschieden abhängig machen.

Interpretieren wir das Wahlverhalten als  $10.000.000 \times 2$  - Spiel kooperativer Art, so können einzelne Entscheidungsregeln (z.B.: Alle Personen, die nicht schwer körperbehindert sind, oder eine wichtige persönliche Verpflichtung am Wahltag haben, oder politisch uninformiert sind ... , gehen zur Wahl) als Lösungen dieses kooperativen Spiels diskutiert werden.

Auch Harsanyi, der wohl profilierteste Vertreter des Bayesianisierungsprogramms der Theorie rationaler Entscheidung in Interaktionssituationen, hat vorgeschlagen, moralische Entscheidungsregeln als Lösungen kooperativer Spiele zu interpretieren und meint, damit eine Form des Regelutilitarismus begründen zu können.<sup>2</sup> Regelutilitaristisch läßt sich die Entscheidung zur Wahl zu gehen, problemlos rechtfertigen, unabhängig von der damit verbundenen Einbuße persönlichen Wohlergehens.<sup>3</sup>

---

<sup>1</sup> Vgl. Luce/Raiffa (1957) Kap.6

<sup>2</sup> Vgl. J.C. Harsanyi, *Rule utilitarianism, rights, obligations and the theory of rational behavior*, *Theory and decision* 12 (1980) 115-133

<sup>3</sup> Vgl. Harsanyi (1980), S. 129-130

Ist  $\hat{\Gamma}$  das der nicht-kooperativen realen Interaktionssituation  $\Gamma$  zugeordnete fiktive kooperative Spiel, so gilt damit nach dieser Theorie: die moralisch gebotene Entscheidung in  $\Gamma$  ist die Lösung des kooperativen Spieles  $\hat{\Gamma}$ . Im Rahmen einer allgemeinen bayesianisierten Spieltheorie<sup>1</sup> (für kooperative und nicht-kooperative Spiele) wäre damit zwar die moralische Entscheidung  $\sigma_1$  in  $\Gamma$  die individuell rationale, NEW-maximierende Entscheidung in  $\hat{\Gamma}$ , aber damit ist  $\sigma_1$  i.a. nicht auch zusätzlich die individuell rationale, bzw. NEW-maximierende Entscheidung in  $\Gamma$ . Die in  $\Gamma$  moralisch gebotene Entscheidung erfordert nach dieser Interpretation also eine irrationale Entscheidung in dieser realen Entscheidungssituation  $\Gamma$ . (Irrational deshalb, weil sie den Nutzenerwartungswert nicht maximiert.) Für unser Beispiel heißt das: *Wählen gehen* mag Bestandteil einer Lösung einer fiktiven kooperativen Interaktionssituation sein, in der die Beteiligten sanktionierte und jeden bindende Abmachungen treffen können; für einen bestimmten spieltheoretischen Ansatz mag damit *wählen gehen* sogar die individuell NEW-maximierende Entscheidung in dieser kooperativen Interaktionssituation sein, aber damit ist *wählen gehen* noch nicht individuell NEW-Maximierend in der realen Entscheidungssituation des Wahltages.

Die Arbeiten, die zu einer 'Ethifizierung' der Spieltheorie beigetragen haben, suchen nicht mehr vordringlich nach Kriterien unabhängig je individuell rationalen Verhaltens in Interaktionssituationen, sondern Kriterien 'vernünftiger' Entscheidung, 'vernünftigen' Interessenausgleichs etc. Dabei bleibt jedoch die Frage offen, was die individuelle Befolgung dieser Kriterien begründen könnte. Warum sollten die Beteiligten Kriterien dieser Art befolgen, wenn das im Widerspruch

---

<sup>1</sup> wie sie z.B. von Harsanyi vorgeschlagen wurde, vgl. ders., *A new solution concept for both cooperative and non-cooperative games*, Opladen 1979

zur Bayesschen Entscheidungsregel steht, also in diesem Sinne ihren subjektiven Wünschen und Zielen nicht entspricht? Eine Antwort auf diese Frage müßte offensichtlich Handlungspräferenzen und Zustandspräferenzen in einer gegenüber der Bayesschen Entscheidungstheorie modifizierten Form aufeinander beziehen. Vor diesem Schritt scheuen die neueren Beiträge zur Ethifizierung der Spieltheorie jedoch fast durchgängig zurück.

Zwar wird in einigen neueren Untersuchungen wie selbstverständlich (wenn auch zumeist nur implizit) ein differenzierter Präferenzbegriff vorausgesetzt,<sup>1</sup> so wenn von "meta-processes" gesprochen wird, die von höherer Ordnung sind, "selecting or determining among other things, the preference structures, relevant action possibilities, and decision procedures of the action"<sup>2</sup> unter ihnen auch "the model of the process (meta-process) of evaluating and choosing among alternative preference relations"<sup>3</sup>, es bleibt dabei aber ungeklärt, in welcher Beziehung diese 'Präferenzen über Präferenzen' mit dem behavioristischen Präferenzmodell der klassischen Spieltheorie stehen. Die Schwierigkeit ergibt sich insbesondere aus der handlungsanleitenden Rolle dieser Metapräferenzen. Dies gilt letztlich auch für die Konzeption der Metarationalität, auf die später noch detaillierter eingegangen wird.

Bei einigen neueren Schriften, die sich mit der Anwendung ökonomischer Methoden, speziell der Entscheidungs- und Spieltheorie, im Bereich normativer Theorien befassen, kann man den Eindruck gewinnen, die bahnbrechende

---

<sup>1</sup> Vgl. R. Hardin, *Collective action as an agreeable n-persons prisoner dilemma*, Behavioral science 16 (1971) 472-481; R.C. Jeffrey, *Preferences among preferences*, Journal of philosophy 71 (1974) 377-391; T.Burns/L.D. Meeker, *Structural properties and resolutions of the prisoners' dilemma game*, in: Rapoport (1974) 35-62; J. Waldner, *Good desires*, in: M. Black, *Problems of choice and decision*, Ithaka 1975, 453-481; D. Braybrooke, *Variety among hierarchies of preference*, in: Hooker/Leach/McClennen (1978), Bd. I 55-65; K. Boulding, *Ethics of the critique of preferences*, in: W. Finnin/G.A. Smith, *The morality of scarcity*, Baton Rouge/London 1979, 9-23

<sup>2</sup> Burns/Meeker (1974), S.35

<sup>3</sup> ebd., S.37

neue Nutzentheorie als Metrisierungstheorie individueller Präferenzen sei ebenso wie die vorausgegangene Methodendiskussion um den ökonomischen Nutzenbegriff längst obsolet geworden. So beginnt z.B. der erste Abschnitt einer "Diagrammatic exposition of justice" von Donald Wittman<sup>1</sup> mit dem Satz: "It is useful to start with a utility (happiness) tradeoff frontier." Ungeklärt bleibt, in welcher Weise das zu verstehen ist: Als Konsequenz der (empirischen) Hypothese, die individuellen Präferenzen  $\xi \succ \xi'$  ordnen die Zustände nach dem Glücksstatus des betreffenden Individuums in diesen Zuständen - also:  $\xi \succ \xi' \Leftrightarrow \iota$  ist in  $\xi$  glücklicher als in  $\xi'$ , wobei diese Äquivalenz nicht logisch sondern empirisch gültig ist - oder ob dieser Zusammenhang analytisch ist, das hätte zur Konsequenz, daß hier die Nutzenfunktion als Metrisierung der individuellen (overall-) Präferenzen aufgegeben wurde, und wir dann entweder "Nutzen" schlicht als einen anderen Namen für Glückseligkeit auffassen müssen, oder - was mir sinnvoller erschiene - diese Art Nutzenfunktion als Metrisierung nicht der individuellen overall-Präferenzen, aber als Metrisierung der zweistelligen Index-Relation -  $\iota$  ist in  $\xi$  glücklicher als in  $\xi'$  - zu interpretieren haben.<sup>2</sup>

Nicholas Rescher<sup>3</sup> operiert dagegen explizit mit zwei Arten von Nutzenfunktionen, eine mit und eine ohne "vicarious participation in the welfare of others". Rescher knüpft jedoch in keiner Weise an die moderne

---

<sup>1</sup> D. Wittman, *A diagrammatic exposition of justice*, in: Brock, *Game theory, social choice and ethics*, Theory and decision 11 (1979) 207-237, S.208

<sup>2</sup> Ähnliche Fragen stellen sich bei F. Schick, *Toward a theory of sociality*, in: Hooker/Leach/McClennen (1978), Bd.II 121-130. Hier ist die Beziehung von Interesse und Nutzen in ähnlicher Weise unklar wie bei Wittman die Beziehung 'happiness' und Nutzen. Auch Brew legt seinen interessanten Überlegungen keinen Nutzenbegriff zugrunde, der sich in die v.Neumann-Morgensternsche Präferenztheorie einpaßt. Vgl. J.S. Brew, *An altruism parameter for prisoner's dilemma*, Journ. of conflict resolution 17(1973) 351-367

<sup>3</sup> N. Rescher, *Unselfishness*, Pittsburg 1975

<sup>4</sup> ebd., S.23



Nutzenkonzeption an. Seine Polemik gegenüber ökonomischen Rationalitätstheorien, deren Versäumnis "to maintain a sufficiently sharp distinction between prudence and rationality"<sup>1</sup> er kritisiert und hervorhebt, daß fast alle ökonomischen Beiträge fälschlich Rationalität mit etwas identifizieren, "what is, in effect, simply self-interested prudence"<sup>1</sup>, läßt vermuten, daß er eine klassische Nutzeninterpretation zugrundelegt.

Reschers Präferenztheorie leidet trotz dieses Unterschieds ebenfalls an dem ungeklärten Verhältnis der beiden utility-status<sup>2</sup>, denn die zwischen beiden vermittelnde 'transfer-rate' "is designed to reflect the whole iterative process of reactions to reactions, insofar as such a feedback effect comes into play. The p.t.r. (=transfer rate; d.A.) is to represent the all-comprising upshot that results 'when everything is said and done'."<sup>3</sup>

Wir haben aber bei der Behandlung zirkulär bestimmter Präferenzen festgestellt, daß es ein Ende eines solchen iterativen Prozesses in vielen Fällen, d.h. für zahlreiche Interdependenzstrukturen, nicht gibt, und in anderen Fällen das Ergebnis bzw. die Grenzwerte eines solchen Prozesses inadäquate Zustandsbewertungen darstellen. Darüberhinaus ist es nicht zweckmäßig, von Bewertungsfunktionen als Metrisierungen von Präferenzen auszugehen (vgl. §10).

---

<sup>1</sup> Rescher (1975), S.38

<sup>2</sup> "given any allocation or distribution of 'utility' (welfare, well-being, happiness, etc.) to various people, the utility-status of any individual can be regarded from two perspectives.  
(1) His first-order utility-status, determined on what might be termed a prima facie basis, that is, without any reference to his affective involvement with the fate of others.  
(2) His second-order utility-status, incorporating into his utility condition a part of the condition of others, duly internalized through a vicarious participation in the welfare of others"

Rescher (1975), S.23

<sup>3</sup> ebd., S.24

### 13. BEGRIFFSBESTIMMUNG: MORALISCHES SYSTEM

So wie man fragen kann, was unterscheidet das Recht von anderen Arten von Verhaltensregeln, so kann man fragen, was unterscheidet moralische Regeln, oder Systeme von moralischen Regeln, von anderen Verhaltensregeln bzw. Systemen von Verhaltensregeln. Ein Kriterium, das auf diese Frage Antwort gibt, darf nicht verwechselt werden mit einer Bestimmung des moralisch Richtigen.

Eine Klasse von Verhaltensregeln  $S$  für eine Personengruppe  $K$  ist genau dann ein moralisches System, wenn

(1) es für jede Person aus  $K$  vorteilhaft ist, wenn jede Person aus  $K$  in Übereinstimmung mit  $S$  handelt,

und

(2) ein Verhalten in Übereinstimmung mit  $S$  von allen Personen unter bestimmten Umständen persönlich nachteilige Handlungen verlangt.

Die drei Formulierungen

(a) Die Verhaltensweise der Person  $i$  ist in Übereinstimmung mit der Verhaltensregel  $V$

(b)  $i$  hält sich an  $V$

(c)  $i$  akzeptiert  $V$

haben in der Alltagssprache voneinander abweichende Bedeutungen.

Bei (a) handelt es sich um die Feststellung einer objektiven Übereinstimmung, unabhängig von den Intentionen des Handelnden. Bei (c) geht es in erster Linie um die Intentionen des Handelnden. Kann es richtig sein, daß  $i$   $V$  akzeptiert, aber  $V$  zuwiderhandelt? Das ist denkbar, auch wenn es i.a. ein starkes Indiz dafür ist, daß  $i$   $V$  nicht akzeptiert, wenn  $i$  nicht  $V$  gemäß handelt. Die Formulierung (b) ist hingegen nur wahr, wenn (a) wahr ist, aber (a) scheint mir keine hinrei-

chende Bedingung für (b) zu sein, denn (b) hat eine intentionale Konnotation: (b) kann nur sinnvoll auf intentionales Verhalten angewendet werden, während das für (a) nicht der Fall ist.

Im folgenden gehen wir davon aus, daß jemand, der V akzeptiert, i.a. auch V gemäß handelt, m.a.W. wir gehen davon aus, daß empirisch gilt: (c)  $\Rightarrow$  (a) und (c)  $\Rightarrow$  (b). Der Unterschied zwischen (a) und (b) andererseits ist für menschliches bewußtes Verhalten - und nur darauf bezieht sich i.a. die ethische Analyse - irrelevant. Jedoch ist (a) höchstens ein Indiz für (c). Wenn i V akzeptiert, dann würde er in jeder Situation nach bestem Wissen V gemäß handeln, wenn i in allen bisherigen realen Situationen V befolgt hat, so ist anzunehmen, daß er das in vielen Fällen bewußt getan hat und, wenn er bisher in keiner Situation V bewußt zuwider gehandelt hat, dann kann man u.U. vermuten, daß er in jeder Situation V befolgen würde bzw., daß die Person i die Verhaltensregel V akzeptiert.

Die Definition (1)  $\wedge$  (2) läßt sich abschwächen: Eine Klasse von Verhaltensregeln S für eine Personengruppe K ist genau dann ein moralisches System, wenn (1a) es für jede Person aus K vorteilhaft ist, wenn fast alle Personen aus K in Übereinstimmung mit S handeln,

und

(2a) ein Verhalten in Übereinstimmung mit S von den meisten Personen unter bestimmten Umständen persönlich nachteilige Handlungen verlangt.

Eine weitere Abschwächung von (2) zu

(2b) ein Verhalten in Übereinstimmung mit S von mindestens einer Person unter bestimmten Umständen persönlich nachteilige Handlungen verlangt.

(2b) scheint mir schon zu schwach zu sein, aber darüber kann man geteilter Meinung sein.

Sicher inadäquat wäre jechoche eine entsprechende weitergehende Abschwächung von (1) zu

(1b) es für jede Person aus K vorteilhaft ist, wenn sich mindestens eine Person aus K in Übereinstimmung mit S verhält,

denn dies würde auch Verhaltensregeln ohne Koordinationsfunktion als moralische kennzeichnen.

Der Ausdruck "vorteilhaft" in (1) kann mißverstanden werden. Wenn alle Personen aus K sich in Übereinstimmung mit S verhalten, oder S akzeptieren, wie wir kürzer sagen wollen, dann ist das nicht erst dann vorteilhaft für die Person i, wenn es kein anderes moralisches System S' gibt, das - allgemein akzeptiert - i besser stellen würde, sondern genau dann, wenn i in S (d.h. wenn alle sich in Übereinstimmung mit S verhalten) besser dran ist, als wenn von den Personen aus K kein gemeinsames moralisches System akzeptiert ist.

Damit ist also moralisches Verhalten nicht vorteilhaft für jeden in dem Sinne, daß jeder seinen Nutzen maximiert, wenn er moralisch handelt, sondern moralisches Verhalten ist insofern für jeden vorteilhaft, als gilt, daß es für jede Person nachteilig ist, wenn niemand moralisch handelt.

Dieser Sachverhalt läßt sich auch anders formulieren:

Für jede Person ist der persönliche Vorteil, der sich (für sie) aus den persönlich nachteiligen Handlungen der anderen (infolge eines akzeptierten Systems moralischer Verhaltensregeln) ergibt, größer als der persönliche Nachteil, der für diese Person daraus erwächst, daß sie die Regeln des moralischen Systems befolgt. Das ist eine Konsequenz der Definition, die deshalb sinnvoll ist, weil andernfalls jeder ein System von Verhaltensregeln S' vorziehen würde, das sich von dem ursprünglichen S nur darin unterscheidet, daß es keine persönlich nachteiligen Handlungen verlangt. Ein solches System würden wir nicht als ein System morali-

scher Verhaltensregeln bezeichnen.<sup>1</sup>

Für eine moralische Verhaltensregel V sind folgende individuelle Präferenzen charakteristisch:

Für jeden einzelnen ist es am vorteilhaftesten, wenn die anderen sich an die Verhaltensregel halten, es ihm persönlich jedoch freigestellt ist. Schon weniger vorteilhaft ist es für jeden einzelnen, wenn alle (einschließlich ihm selbst) die Verhaltensregel befolgen. Noch nachteiliger ist es jedoch für jeden einzelnen, wenn sich niemand an die Verhaltensregel hält. Darüber hinaus ist es für jeden einzelnen i.a. noch nachteiliger, wenn ausschließlich er selbst sich an die Verhaltensregel hält.

Wenn wir die vier Möglichkeiten auflisten

A: Jeder hält sich an die Verhaltensregel V

B: Niemand hält sich an die Verhaltensregel

C: Nur die Person i hält sich an die Verhaltensregel

D: Alle außer der Person i halten sich an die Verhaltensregel

ergibt sich als typische individuelle Wohlfahrtspräferenz, wenn V eine moralische Verhaltensregel ist:<sup>2</sup>

$$\forall i \in K : \langle A, B \rangle \in R_{w_i} \wedge \langle D, C \rangle \in R_{w_i} \wedge \\ \wedge \langle D, A \rangle \in R_{w_i} \wedge \langle B, C \rangle \in R_{w_i}.$$

Diese typische individuelle Wohlfahrtspräferenz ist in vielen Fällen nicht die tatsächliche (overall-)Präferenz der einzelnen Personen, denn sonst gäbe es (per definitionem) kein moralisches Verhalten. Diese  $R_{w_i}$ -Präferenzen sind jedoch den moralisch handelnden Personen als

---

<sup>1</sup> Vgl. dazu Baier (1958) S.308-320 und S.191-213

"Morality is a system of principles whose acceptance by everyone as overruling the dictates of self-interest is in the interest of everyone alike, though following the rules of a morality is not of course identical with following self-interest." S.314

vgl.a. D.P. Gauthier, *Morality and advantage*, Philosophical Review 76 (1976), S.463

<sup>2</sup> Vgl. dazu N. Hoerster, *Utilitaristische Ethik und Verallgemeinerung*, Freiburg/München 1977, S.123-139

Ausdruck ihrer persönlichen Interessen zumeist bewußt.

In unserer Definition eines Systems moralischer Verhaltensregeln spielten nur zwei Präferenzen eine Rolle:

$\forall i \in K : \langle A, B \rangle \in R_{w_i}$  und

$\forall i \in K : \langle D, A \rangle \in R_{w_i}$

Statt  $D \underset{w_i}{>} A \underset{w_i}{>} B \underset{w_i}{>} C$  beschränkt sich die Definition eines System von moralischen Verhaltensregeln auf  $D \underset{w_i}{>} A \underset{w_i}{>} B$ .

Wenn wir sagten "i hält sich an V", meinten wir "i befolgt in jeder Situation V" und, wenn wir sagten "i hält sich nicht an V", meinten wir damit nicht, daß i in jeder Situation V zuwider handelt, sondern nur, daß i sich nicht grundsätzlich an V hält, sondern (insbesondere wenn es im persönlichen Interesse ist) in manchen Situationen davon abweicht.

Zur Illustration greifen wir auf eine sprachphilosophische Argumentation zurück: David Lewis meint,

"es gibt in P (=eine Sprechergemeinschaft) eine allgemeine Konformität mit der Regularität der Wahrhaftigkeit und des Vertrauens in (der Sprache) L. Unter der Voraussetzung, daß sich die meisten an die Regularität halten, wollen die Mitglieder von P, daß sich alle daran halten. Sie wollen, daß die anderen, wie auch sie selbst wahrhaftig und vertrauensvoll in L sind. Diese allgemeine Präferenz wird durch ein gemeinsames Interesse an Kommunikation aufrecht erhalten"<sup>1</sup>

Es wird nicht ganz klar, was diese Sätze genau meinen. Es liegt nahe, sie so zu verstehen, daß sich jeder einzelne aufgrund seines persönlichen Interesses an Kommunikation wahrhaftig und vertrauensvoll verhält. Wesentlich erscheint mir jedoch, daß es sich bei der Regularität der Wahrhaftigkeit und des Vertrauens um eine normative Vertrauensregel handelt, deren allgemeine

---

<sup>1</sup> D. Lewis, *Die Sprache und die Sprachen*, in: G. Meggle (Hg.), *Handlung, Kommunikation, Bedeutung*, Frankfurt/M. 1979, 197-240, S.204

Befolgung in L durch P eine notwendige Bedingung dafür ist, daß die Population P die Sprache L verwendet.

Lewis spricht vom gemeinsamen Interesse an und von der allgemeinen Präferenz für eine allgemeine Regularität der Wahrhaftigkeit und des Vertrauens. Wir können jedoch auch bei dieser Verhaltensregel davon ausgehen, daß zwar jeder lieber möchte, daß sich alle daran halten, als daß sich keiner daran hält - insofern kann man von einem gemeinsamen Interesse für eine allgemeine Konformität mit dieser Regularität sprechen, jedoch gibt es in einer Hinsicht gerade kein gemeinsames Interesse dieser Art: Wir können davon ausgehen, daß zwar jeder ein persönliches Interesse daran hat, daß die anderen sich an die genannten Konventionen halten, aber es ist nicht einzusehen, warum jemand ein persönliches Interesse daran hat, daß er selbst sich in jeder Situation an diese Konventionen hält - schließlich kommuniziert er nur mit anderen, nicht mit sich selbst. Allerdings hat ein Interesse daran, daß die anderen meinen, er hielte sich in jeder Situation an diese Konventionen, denn andernfalls würde seinen eigenen Äußerungen kein Vertrauen mehr entgegengebracht. Man mag einwenden "Lügen haben kurze Beine", m.a.W. das individuelle Interesse an Kommunikation ließe es angeraten erscheinen, sich auch in jedem Einzelfall an die Verhaltensregel der Wahrhaftigkeit und des Vertrauens zu halten. Träfe das zu, würde es sich bei der Regularität der Wahrhaftigkeit und des Vertrauens tatsächlich um eine nicht-normative Verhaltensregel handeln, da ihre Befolgung in jeder Entscheidungssituation im Interesse des Einzelnen ist. Wenn dieser - nur empirisch prüfbare - Sachverhalt jedoch für eine normale, hinreichend große Sprechergemeinschaft nicht zutrifft, was anzunehmen ist, dann ist es irreführend, zu unterstellen, die je individuelle Konformität mit der Regularität der Wahrhaftigkeit und des Vertrauens sei individuell rational (im üblichen Sinne von im persönlichen Interesse) aufgrund eines persönlichen

Kommunikationsbedürfnisses. In diesem Fall wäre die Konformität mit der Regularität der Wahrhaftigkeit und des Vertrauens eine normative Verhaltensregel, und die für die Sprechergemeinschaft konstitutiven Konventionen bildeten insgesamt ein moralisches System im oben definierten Sinne.

Das zweite Beispiel führt uns zurück zum Gefangenendilemma und zur Frage der kollektiven Rationalität. Ist die Verhaltensregel "Wähle in Entscheidungssituationen vom Typ des Gefangenendilemmas die kooperative Handlungsalternative" eine moralische Verhaltensregel im Sinne der Definition (1)  $\wedge$  (2) oder einer ihrer Abschwächungen? Ja diese Verhaltensregel erfüllt die Bedingungen (1) und (2) sogar in ihrer strengen Form. Damit ist allerdings nicht gesagt, daß die kooperative Entscheidung jeweils moralisch geboten ist.

Interessanter ist die umgekehrte Fragestellung: Überträgt man die vier Alternativen A,B,C,D auf den Zwei-Personen-Fall, ergibt sich mit

- +v<sub>i</sub>: i hält sich an V
- v<sub>i</sub>: i hält sich nicht an V

folgende Matrix:

	+v <sub>2</sub>	-v <sub>2</sub>
+v <sub>1</sub>	A	C <sub>1</sub> bzw. D <sub>2</sub>
-v <sub>1</sub>	C <sub>2</sub> bzw. D <sub>1</sub>	B

- wobei C<sub>i</sub>: Nur i hält sich an V  
und D<sub>i</sub>: Alle außer i halten sich an V.

Ist v eine moralische Verhaltensregel, haben also die Personen noch folgende individuellen Wohlfahrtspräferenzen:

Person 1: D<sub>1</sub> > A > B

Person 2: D<sub>2</sub> > A > B



Ergänzt man die Präferenzstruktur zu

Person 1:  $D_1 > A > B > C_1$

Person 2:  $D_2 > A > B > C_2$

so ist diese Entscheidungssituation von der Art des Gefangenendilemmas. Interaktionssituationen vom PD-Typ spielen daher eine besondere Rolle für die Sozialethik.

## 14. INTERAKTIONSSITUATIONEN VOM PD-TYP

Eine Gefangenendilemma-Situation im engeren Sinne ist eine Zwei-Personen-Interaktionssituation mit einer Präferenzstruktur, wie in §12 dargestellt. Die individuell dominanten Strategien führen zu einer kollektiven Entscheidung, die bezüglich der zugrundegelegten Präferenzen nicht pareto-optimal ist. Verallgemeinert sagen wir von einer Interaktionssituation mit einem nicht-pareto-optimalen Gleichgewichtspunkt in dominanten Strategien, sie sei vom PD-Typ, oder kurz eine PD-Situation. Wenn man die in §9 eingeführten Präferenztypen unterscheidet, kann eine Interaktionssituation in dreierlei Hinsicht vom PD-Typ sein: Bezüglich der individuellen Wohlfahrtspräferenzen, aber auch bezüglich der resultierenden und der moralischen Präferenzen der beteiligten Personen.

Betrachten wir zunächst Situationen, die bezüglich der Wohlfahrtspräferenzen vom PD-Typ sind. Das Gefangenendilemma zeigt: Verfolgen in einer Gruppe von Personen alle (rational) ihre persönlichen Interessen, so kann das zu einem für jeden einzelnen ungünstigeren Ergebnis (relativ zu  $Rw_i$ ) führen, als wenn die Beteiligten von der rationalen Verfolgung ihrer persönlichen Interessen abgehen. In welchem Sinne ist das ein Dilemma?

Ein theoretisches Dilemma ist dadurch gekennzeichnet, daß zwei oder mehrere plausible (oder gut begründete) Annahmen sich als logisch unvereinbar erweisen. Ein praktisches Dilemma dadurch, daß zwei zugleich bestehende Zielvorstellungen aus logischen Gründen (wie beim Liberalen Paradoxon, vgl. §9) oder aufgrund bestimmter empirischer Umstände nicht gemeinsam verwirklicht werden können.

In einer Interaktions-Situation vom PD-Typ befinden sich die Beteiligten in einem praktischen Dilemma: Einerseits würden sie, etwa als unparteiische Beobachter, dazu raten, sich im eigenen Interesse kooperativ zu verhalten, und sie würden sich als Betroffenen wünschen, daß von den möglichen symmetrischen Verhaltensregeln eine kooperative eingehalten wird. Andererseits wissen sie, daß die nicht-kooperative Entscheidung (in  $Rw_i$ ) dominant ist, d.h. daß es unabhängig davon, wie die anderen entscheiden, für sie vorteilhafter ist, die unkooperative Strategie zu wählen. Nur wenn man annimmt, daß sich aus der Befürwortung einer Verhaltensregel zugleich ein entsprechendes Entscheidungsmotiv ergibt, stellt eine Situation vom PD-Typ den einzelnen Beteiligten vor ein praktisches Dilemma.<sup>1</sup>

Inwiefern könnten Situationen vom PD-Typ ein theoretisches Dilemma aufzeigen? Wer davon überzeugt ist, daß die Interessen einer Gruppe von

---

<sup>1</sup> Anatol Rapoport, der sich auch experimentell mit Interaktions-Situationen vom PD-Typ auseinandergesetzt hat, (vgl. *Prisoners Dilemma*, Ann Arbor 1965), schlägt eine 'Lösung' dieses praktischen Dilemmas vor, indem er ausschließlich dieses Entscheidungsmotiv zugrundelegt: "Anstatt als Kalkulationsgrundlage die Frage zu stellen: 'Wobei bin ich besser dran?', nehmen wir an, daß jeder Gefangene von der folgenden Annahme ausgeht: 'Mein Partner ist wie ich, deshalb wird er wahrscheinlich handeln wie ich. Wenn ich folgere, daß ich bekennen sollte, wird er wahrscheinlich zu demselben Schluß kommen. Wenn ich folgere, daß ich nicht bekennen sollte, denkt er wahrscheinlich auch so. Im ersten Fall erhalten wir beide -5, im zweiten Fall erhalten wir beide +5. Das zeigt, daß ich persönlich Nutzen davon habe, nicht zu bekennen.'" (Rapoport (1976, S.194)) Rapoport stellt die Entscheidungssituation des Gefangenendilemmas in der folgenden Gewinnmatrix dar:

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	<-5,-5>	<10,-10>
$\sigma_{12}$	<-10,10>	< 5,5 >

Wobei  $\sigma_{11}, \sigma_{21}$  die Handlungen 1 bzw.2 'bekennt' und  $\sigma_{12}, \sigma_{22}$  die Handlungen 1 bzw. 2 'bekennt nicht' bezeichnen soll.

Personen immer dann am besten befriedigt werden, wenn jeder seine Interessen optimal verfolgt, ist mit dem Gefangenendilemma widerlegt. Zugespitzt könnte man sagen, das Gefangenendilemma (bzgl.  $Rw_i$ ) erscheint in diesem Sinne als ein theoretisches Dilemma nur vor dem Hintergrund einer extrem liberalistischen oder anarchistischen Auffassung.<sup>1</sup> Da das nicht-kooperative Verhalten in Interaktions-Situationen vom PD-Typ, von beiden befolgt, für jeden Einzelnen ein unerfreuliches Ergebnis zeitigt, ist es kollektiv nicht empfehlenswert, genauer: es ist nicht kollektiv rational bzgl.  $Rw_i$ . Ein außenstehender Beobachter wird vielleicht sagen, die

---

<sup>1</sup> So argumentiert Anatol Rapoport (*Kämpfe, Spiele und Debatten*, Darmstadt 1976, S.196f.): "Das Gefangenendilemma ist ein nachdrückliches Beispiel für eine gesellschaftliche Situation, in welcher die "Summe" der beiden individuellen Interessen sich zu einem Nachteil für beide Seiten addiert. Der Grund dafür, daß dies paradox erscheint, liegt vermutlich darin, daß wir allzu lange daran gewöhnt waren, unkritisch den Laissez-faire-Grundsatz hinzunehmen, nämlich, daß eine Gesamtheit von Individuen, die ihre jeweiligen Eigeninteressen durch kurzsichtige Kalkulationen suchen, tatsächlich zur Verwirklichung dieses Eigeninteresses infolge des Wirkens ökonomischer Gesetze gelangen wird, die aus der Annahme einer freien Marktwirtschaft abgeleitet werden."

Allerdings ist der Ausdruck "kurzsichtig" wohl selbst wieder Folge dieser Befangenheit, denn die kooperative kollektive Entscheidung ist ja auch bei einer einmaligen Entscheidung in einer Konfliktsituation der Art des Gefangenendilemmas vorteilhafter für alle Beteiligten als die individuell rationale kollektive Entscheidung.

Ähnlich C.F.v. Weizsäcker (*Die Einheit der Natur*, München 1974, S.31): "Der geniale Gedanke der klassischen Nationalökonomie, daß das Zusammenspiel der konsequent verfolgten Eigeninteressen das Gesamtinteresse automatisch fördere - er wurde für Hegel zum Modell seiner 'List der Idee', zum Beweis, daß das Gemeinwesen nicht des platonischen Philosophenkönigs bedürfe, der als Individuum das Wohl des Ganzen adäquat denkt; er wurde ebenso zum Modell der Darwinschen Selektionslehre -, dieser Gedanke erweist sich in seiner Gültigkeit von gesellschaftlichen Voraussetzungen abhängig, die wir heute wieder kritischer betrachten als in der Blüte des Wirtschaftswunders der kapitalistischen Welt nach dem letzten Weltkrieg."

beiden verhalten sich irrational - distributiv ist das falsch, kollektiv ist das richtig. Es ist für die Beteiligten kollektiv vorteilhaft, wenn sie sich irrational verhalten. Individuell rationales Verhalten bzgl.  $Rw_i$  ist keine gute allgemeine Empfehlung an eine Gruppe, das eigene individuell rationale Verhalten ist dennoch im Interesse jedes Beteiligten (das ist nicht trivial).<sup>1</sup> Die Tatsache, daß es für die Beteiligten kollektiv vorteilhaft ist, sich im Sinne der Entscheidungstheorie irrational zu verhalten, könnte als Begründung irrationalen Verhaltens interpretiert werden, und das beunruhigt nicht nur Spieltheoretiker.<sup>2</sup> Auch Moralphilosophen wünschen nicht, daß ihre normativ-ethischen Theorien als Empfehlungen irrationalen Verhaltens gelten müssen.

Anders stellt es sich dar, wenn wir Situationen betrachten, die bezüglich der resultierenden Präferenzen vom PD-Typ sind. In diesem Fall ist das Verhalten der Person zugleich Kriterium für die Beschreibung der Situation als Gefangenendilemma in  $Rres_i$ . Wir wollen die Frage jedoch an dieser Stelle offen lassen, ob eine kooperative Entscheidung zeigt, daß es sich nicht um eine PD-Situation in  $Rres_i$  handelt - darauf kommen wir zum Schluß noch einmal zurück. Es sei nur festgehalten, daß der Schluß von kooperativem Verhalten auf eine entsprechende Präferenzstruktur aufgrund des entscheidungstheoretischen Rationalitätsmodells bestimmte subjektive overall-Bewertungen der Konsequenzen nach sich ziehen würde, die selbst gewissen Plausibili-

---

<sup>1</sup> Vgl. Parfit, *Prudence, morality and the prisoner's dilemma*, Oxford 1981.

<sup>2</sup> Vgl. N. Howard, *Paradoxes of rationality: Theory of meta-games and political behavior*, Cambridge/London 1971. Bei Howard firmiert das Dilemma als 'second breakdown of rationality' und begründet zusammen mit zwei weiteren 'breakdowns of rationality' einen auf der Analyse von Metaspielen beruhenden neuen Begriff der Entscheidungsrationalität.

tätsbedingungen genügen sollten.<sup>1</sup> Nur wenn die Rubri-  
zierung einer Interaktionssituation mit zwei Per-  
sonen und je zwei Handlungsalternativen unter den  
Typ des Gefangenendilemmas nicht analytisch vom Ent-  
scheidungsverhalten der Personen in der Interaktions-  
Situation abhängt, kann das Gefangenendilemma einer  
empirischen und auch normativen Analyse unterzogen  
werden. So wäre etwa die groß angelegte empirische  
Studie von Rapoport und Chammah (1965) zum *Prisoner's  
dilemma* selbstverständlich gegenstandslos, wenn das Auf-  
treten längerer Abschnitte kollektiv rationalen Ent-  
scheidungsverhaltens in den Testreihen wiederholter PD-  
Entscheidungen so interpretiert werden müßte, daß sich  
in diesen Phasen die Konsequenzenbewertungen der Perso-  
nen entsprechend geändert haben, so daß im Augenblick  
der jeweiligen kooperativen Entscheidung nur scheinbar  
- nämlich in Gestalt der monetären Auszahlung<sup>2</sup> - eine

---

<sup>1</sup> Orthodoxe Spieltheoretiker sehen in PD-Situationen kein Di-  
lemma, denn für die ergeben sich die subjektiven Bewertungsfunktionen im Sinne des revealed-preference-Modells aus den tatsächlichen Handlungswahlen. Kooperatives Entscheiden hebt also PD-Situationen auf. Man muß sich jedoch über folgendes im klaren sein: Wenn allein die Tatsache, daß in einer Interaktions-Situation mit zwei Personen und zwei Handlungsalternativen eine rationale Person Handlungspräferenzen hat, die dem sure-thing-principle bezüglich einer PD-Konsequenzenbewertung widersprechen, ausschließt, daß die Charakterisierung der Situation als vom Typ des Gefangenendilemmas adäquat ist, und dieser Sachverhalt analytischer Natur ist, so würde die Entscheidungstheorie auf eine 'logische' Theorie in dem Sinne reduziert, daß sie weder einer handlungsanleitenden (normativen) noch einer handlungserklärenden (empirischen) Interpretation fähig ist. Es muß eine von der aktuellen Handlungspräferenzrelation des Individuums unabhängige (=autonome) Methode geben, die individuellen Konsequenzenpräferenzen und damit u.U. die entsprechenden ordinalen bzw. im günstigsten Falle auch kardinalen Bewertungsfunktionen zu bestimmen, wenn die Entscheidungstheorie ihrem Anspruch als generelle Theorie rationalen Verhaltens gerecht werden soll.

<sup>2</sup> Die experimentellen Studien zur Spieltheorie modellieren die gewünschten Interaktions-Situationen fast immer durch entsprechende monetäre Auszahlungen. Vgl. Vanacke (1957); Thibaut u. Kelly (1959); Wilson u. Bixenstine (1962); Vogelsang (1963); Rapoport u. Chammah (1965); Rapoport u. Orwant (1965); Howard (1971) Appendix B; Rapoport u. Perner (1974); Horowitz (1977); Rapoport, Kahan, Funk und Horowitz (1979).

Interaktionssituation vom Typ des Gefangenendilemmas vorlag, in Wirklichkeit- d.h. was die individuelle Nutzenbewertung der Konsequenzen betrifft - jedoch eine Interaktionssituation vom Typ der rechten Matrix.<sup>1</sup>

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	<3,3>	<1,4>
$\sigma_{12}$	<4,1>	<2,2>

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	<4,4>	<1,3>
$\sigma_{12}$	<3,1>	<2,2>

<sup>1</sup> Die mangelnde Klärung metatheoretischer Fragen der modernen Entscheidungs- und Spieltheorie macht sich hier bemerkbar. Das erklärt auch die verwirrende Vielfalt unvereinbarer Positionen in der Interpretation z.B. des Gefangenendilemmas, die von der Diagnose "Jenseits des normativen Anwendungsbereichs der Spieltheorie" (Rapoport) bis zur Diagnose "Anderer Rationalitätsbegriff notwendig" (Nigel) reicht. Unser Ziel ist es nicht, das Problem der Spieltheorie zu lösen (falls es sich als ein solches der Spieltheorie tatsächlich aufrecht erhalten läßt), etwa indem wir den Ansatz Nigels weiterverfolgen, sondern einige Grundprobleme der Sozialethik klarer zu fassen und bei dem Versuch ihrer Lösung das entscheidungs- und spieltheoretische Instrumentarium fruchtbar zu machen. Rapoport (1965) spricht von "multiordinality of the meaning of 'rationality'", wobei die Rationalitätstypen nach Konfliktarten unterschieden werden. In Nullsummenspielen liefert die Spieltheorie eine eindeutige und überzeugende Festlegung 'rationaler Entscheidung': "it is within this context that game theory can be most justifiably said to have laid the mathematical foundations of a theory of rational conflict. For in this context there is no question of what it means to 'play rationally'" (S.16), nämlich falls ein Sattelpunkt vorliegt, die Minimax-Strategie, und bei Nullsummenspielen ohne Sattelpunkt die gemischte Gleichgewichtsstrategie.

Jedoch: "The two-person zero-sum game without a saddlepoint constitutes in our opinion the limit of applicability of game theory as a normative (or prescriptive) theory. That is to say, the minimax mixed strategy can be both rationally defended and intuitively accepted as a 'best course of action' in a certain sense. Once the limits of two-person zero-sum games are transcended, game theory, while remaining a powerful tool for analyzing the logic structure of conflict of interest, loses its prescriptive power". Vgl.a. ders., *Strategy and conscience*, New York 1964 und ders., M.J. Guyer, D.G. Gordon, *The 2x2-game*, Ann Arbor 1976, bes. S.41-49 und S.151-158.

Das führt Rapoport und Perner zu einer sehr vorsichtigen Beurteilung der Möglichkeiten der empirischen Widerlegung bzw. Bestäti-

Wir können jedoch schon jetzt festhalten, daß in einer Interaktionssituation vom PD-Typ bzgl.  $Rw_1$  ohne externe Effekte die kollektiv rationale Entscheidung moralisch besser als die individuell rationale ist. Die kürzeste Begründung wäre die Pareto-Inklusivität moralischer Aggregationsregeln bzgl.  $Rw_1$ . Eine detailliertere Begründung würde eine Reihe moralischer Kriterien daraufhin prüfen, ob sie die kooperative oder die individuell rationale Verhaltensweise auszeichnet.

Das Rawlssche Differenzprinzip<sup>1</sup>, alle Spielarten des

---

gung spieltheoretischer Lösungsmodelle:

"One question, to be sure, remains. Do the payoff numbers in the game matrix represent the 'real' utilities of the players? To raise this question means to become concerned with the problem of assessing the players' 'utility functions' for money or whatever the payoffs represent. While methods for assessing these functions experimentally have been proposed, the results are anything but conclusive. That is to say, it is by no means possible to get definite consistent answers to the questions posed: so a diversion into an altogether different experimental area purportedly to 'refine' the formulation of a game by determining the players' 'real' utilities, in order to test Nash's theory experimentally, does not seem to be warranted." (*Testing Nash's solution of a cooperative game*, in: Rapoport (1974) 103-115, S.107f.

Jedoch wird dann eine überraschende Wendung vollzogen:

"A much simpler way out is to put the proposed solution of the cooperative game to experimental test directly by assuming that the payoffs represent utilities" (S.108). Aber schließlich hat das die unbefriedigende Konsequenz: "Whether the theory is then corroborated or refuted will not constitute a final verdict (because of our ignorance of the 'real' utilities)" (S.108), unbefriedigend deshalb, weil diese Konsequenz das Eingeständnis beinhaltet, daß wegen mangelnder metatheoretischer Klärung unklar bleiben muß, was den empirischen (und/oder gegebenenfalls normativen) Gehalt der zu prüfenden Theorie ausmacht.

<sup>1</sup> Vgl. J. Rawls, *A theory of justice* (1971), dt.: *Eine Theorie der Gerechtigkeit*, Frankfurt/M. 1975.

Der zweite Grundsatz der Rawlsschen Gerechtigkeitstheorie lautet, soziale und wirtschaftliche Ungleichheiten sind nur dann gerecht, wenn sich aus ihnen Vorteile für jedermann ergeben, insbesondere für die schwächsten Mitglieder der Gesellschaft (S.32). Der erste Grundsatz garantiert das größtmögliche Maß an gleichen persönlichen Freiheiten für alle, er ist dem zweiten Grundsatz 'lexikographisch' vorgeordnet, d.h. erst wenn der erste Grundsatz voll erfüllt ist, kann die Gesellschaft gemäß dem zweiten geordnet werden. Der zweite Gerechtigkeitsgrundsatz wird auch Unterschiedsprinzip oder Differenzprinzip genannt.



Utilitarismus<sup>1</sup>, Suppes' 'grading principle of justice'<sup>2</sup>, alle mir bekannten Fairnessprinzipien<sup>3</sup>, Kants moralisches Gesetz<sup>4</sup>, Sidgwick's 'equity principle'<sup>5</sup>, Singers Argument der Verallgemeinerung<sup>6</sup> zeichnen die kollektiv rationale Verhaltensweise als die moralische aus.

- 
- <sup>1</sup> Vgl. J. Bentham, *An introduction to the principles of morals, and legislation*, London 1789; J.S. Mill, *Utilitarianism*, London 1863, dt.: *Der Utilitarismus*, Stuttgart 1976; H. Sidgwick, *The methods of ethics*, London 1907, bes. Kap.IV; G.E. Moore, *Principia ethica*, Cambridge 1903, dt.: *Principia Ethica*, Stuttgart 1970, bes. Kap.6 (Idealer Utilitarismus); J.J.C. Smart, *An outline of a system of utilitarian ethics*, Melbourne 1961 (Handlungsutilitarismus); J.J.C. Smart u. B. Williams, *Utilitarianism - For and against*, Cambridge 1973; N. Hoerster, *Utilitaristische Ethik und Verallgemeinerung*, Freiburg/München 1977 (Regelutilitarismus); J.L. Mackie, *Ethik*, Stuttgart 1981, Kap.6
- <sup>2</sup> Vgl. P. Suppes, *Some formal models of grading principles*, Synthese 16 (1966), 284-306
- <sup>3</sup> Vgl. R.B. Braithwaite, *Theory of games as a tool for the moral philosopher*, Cambridge 1969; L.G. Svensson, *Social justice and fair distributions*, Lund 1977; H.R. Varian, *Distributive justice, welfare economics, and the theory of fairness*, in: *Philosophy and public affairs* 4 (1975), 223-247; ders., *Equity, envy, and efficiency*, in: *Journal of economic theory* 9 (1974), 63-91; E.A. Pazner u. D. Schmeidler, *Egalitarian equivalent allocations: A new concept of economic equity*, in: *Quarterly journal of economics* 92 (1978), 671-687; H.W. Brock, *A new theory of social justice based on the mathematical theory of games*, in: *Ordeshook* (1978) 563-627
- <sup>4</sup> Vgl. I. Kant, *Kritik der praktischen Vernunft*, (1788), Stuttgart 1975, S.53
- <sup>5</sup> Vgl. H. Sidgwick, *The methods of ethics*, London 1907, "Whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances." (S.379) Nur wenn "alle" kollektiv zu interpretieren ist, zeichnet dieses Prinzip das kooperative Verhalten in einer Interaktionssituation vom PD-Typ als das moralische aus.
- <sup>6</sup> Vgl. M.G. Singer, *Generalizations in ethics: An essay in the logic of ethics with the rudiments of a system of moral philosophy* (1961) dt.: *Verallgemeinerung in der Ethik: Zur Logik moralischen Argumentierens*, Frankfurt/M. 1975, Kap.4

Kann es Situationen geben, die bezüglich der subjektiven moralischen Präferenzen  $Rm_i$  der Beteiligten vom PD-Typ sind? Dagegen scheint zu sprechen, daß teleologische ethische Theorien i.a. pareto-inklusiv sind und daß meist die eine oder andere Form von Verallgemeinerbarkeitsprinzipien eine Rolle spielt. Andererseits könnten die subjektiven moralischen Präferenzen wesentlich von den Vorgaben ethischer Theorien abweichen. Vereinfachen wir die Fragestellung dadurch, daß wir uns auf solche subjektiven moralischen Präferenzen beschränken, die in einer Beziehung zu den (vermuteten) Funktionen individueller Wohlfahrt  $\pi_i$  stehen, die sich durch eine lineare Funktion

$$\mu_i = \sum_{j=1}^n r_{ij} \cdot \pi_j$$
 darstellen läßt. Bei  $r_j=1$  für alle  $j=1, \dots, n$  gibt es keine Interaktionssituation, die bezüglich der subjektiven moralischen Bewertungsfunktionen  $\mu_j$  vom PD-Typ ist, denn für beliebige individuelle Wohlfahrtsfunktionen  $\langle \pi_j \rangle$  sind die  $\mu_j$  in diesem Fall untereinander gleich.<sup>1</sup> Für alle anderen Transferfaktoren gibt es jedoch Interaktionssituationen, die bezüglich  $\langle \mu_j \rangle$  vom PD-Typ sind (bei Transferfaktoren unter 1 gibt es entsprechend scharf ausgeprägte PD-Situationen, die auch nach der 'Transformation' PD-Situationen bleiben und bei Transferfaktoren über 1 gibt es Situationen, die vorher keine PD-Situationen waren, aber nach Transformation dazu werden).<sup>2</sup>

Für andere funktionale Beziehungen zwischen  $\langle \pi_i \rangle$  und  $\mu_j$  läßt sich i.a. ein ähnlicher Zusammenhang herstellen. Es ist daher völlig unplausibel, anzunehmen, daß PD-Situationen bezüglich der subjektiven moralischen Präferenzen unmöglich seien.

---

<sup>1</sup> Damit wären natürlich auch die strengsten Universalisierbarkeitsbedingungen erfüllt.

<sup>2</sup> Vgl. S.258 ff.

Nachdem man davon ausgehen kann, daß es PD-Situationen bezüglich  $Rw_i$  ( $\pi_i$ ), aber auch bezüglich  $Rm_i$  ( $\mu_i$ ) gibt, ist anzunehmen, daß es auch PD-Situationen bezüglich der resultierenden subjektiven Präferenzen gibt. Allerdings zwingt uns diese Möglichkeit zu einer Festlegung, die wir bei der Einführung des Begriffes 'resultierende Präferenz' aufgeschoben haben (s.S.112ff). Wir sind u.E. davon ausgegangen, daß die resultierenden Präferenzen in etwa den handlungsleitenden Präferenzen im revealed-preference-concept der modernen Nutzentheorie entsprechen. Dies müssen wir nun auf einen Teilbereich von Entscheidungssituationen beschränken: In PD-Situationen ohne Interaktion entsprechen die resultierenden individuellen Präferenzen den handlungsleitenden Präferenzen; es gilt:  
 $U_i(\xi) \geq U_i(\xi') \Leftrightarrow \langle \xi, \xi' \rangle \in Rres_i$ . Das ist aber eine bloß extensionale Äquivalenz. Die resultierenden Präferenzen ergeben sich aus einer Abwägung persönlicher Interessen (repräsentiert durch  $Rw_i$  bzw.  $\pi_i$ ) und moralischer Aspekte (repräsentiert durch  $Rm_i$  bzw.  $\mu_i$ ). Die resultierenden Präferenzen bilden eine Ordnung über die Alternativenmenge  $\Xi$ . Wir gebrauchen den Terminus 'resultierende Präferenzen' also so, daß kein begrifflicher Zusammenhang mit den Handlungspräferenzen der betreffenden Person besteht.

Für PD-Situationen bezüglich  $Rw_i$  wissen wir, welche Konsequenz die moralisch bessere ist, und welches Verhalten kollektiv rational ist. Und dennoch sind damit die normativ-ethischen Probleme nicht gelöst; insbesondere wissen wir nicht, welches individuelle Verhalten normative geboten ist.

Soll sich die eine Person auch dann kooperativ verhalten, wenn sie eine subjektive Wahrscheinlichkeit nahe Eins hat, daß sich die andere Person inkooperativ verhält?

Unter welchen Umständen ist kooperatives Verhalten

normativ geboten?

Welche Annahmen machen wir über die Präferenzstrukturen einer Person, wenn wir ein kooperatives Verhalten erwarten?

Es geht dabei auch um die Klärung dreier zentraler Probleme normativer Sozialethik:

- (1) Das Problem der sozialen Kooperation, das immer dann auftaucht, wenn das Ergebnis individuell rationaler Entscheidungen jeden einzelnen schlechter stellt, als bei einer anderen kollektiven Entscheidung.
- (2) Das Problem der sozialen Koordination bzw. das Problem der wechselseitigen Abstimmung, um eine der kollektiven Entscheidungen zu erreichen, die ein allgemein besseres Ergebnis als die aufgrund bloßer individueller Rationalität getroffene kollektive Entscheidung haben.
- (3) Das Problem des sozialen Vertrauens bzw. das Problem, daß es für ein kooperatives individuelles Verhalten i.a. erst dann gute Gründe gibt, wenn ich davon ausgehen kann, daß sich die anderen, oder wenigstens die meisten anderen, ebenfalls kooperativ verhalten.<sup>1</sup>

---

<sup>1</sup> Vgl. a. Baier (1977)

## 15. KOLLEKTIVE RATIONALITÄT

Wir haben zwar den Begriff "kollektive Rationalität" schon mehrfach verwendet, ihn aber bisher noch nicht präzise definiert.

Wir holen das jetzt in zwei Stufen nach. Zunächst definieren wir kollektive Rationalität im engeren Sinne und diskutieren diesen Begriff anhand einzelner Interaktionssituationen. Anschließend führen wir einen erweiterten Begriff kollektiver Rationalität ein.

Beide Rationalitätsbedingungen haben eine enge Verwandtschaft mit dem Konzept der Pareto-Optimalität. Der wesentliche Unterschied beruht auf der hier angewandten Differenzierung des Präferenzbegriffes.

Ein Zustand  $\xi$  ist kollektiv besser im engeren Sinne als ein Zustand  $\xi'$  genau dann, wenn gilt, daß  $\xi$  für jede Person bezüglich ihrer persönlichen Interessen besser ist:

$$\langle \xi, \xi' \rangle \in Kb_w \Leftrightarrow \forall i \in K: \langle \xi, \xi' \rangle \in Rw_i.$$

Sind die Konsequenzen eindeutig durch die kollektiven Entscheidungen festgelegt, spielen also zufällige Ereignisse der Natur keine Rolle, und sind die individuellen Entscheidungen reine Strategien, können wir auf "Zustände" verzichten und die individuellen Präferenzrelationen als Relationen über die Menge der logisch möglichen kollektiven Entscheidungen definieren.

Kollektive Entscheidungen  $\Sigma$  sind dabei nichts anderes als  $n$ -Tupel individueller Entscheidungen. Es gilt dann analog:

$$\langle \Sigma, \Sigma' \rangle \in Kb_w \Leftrightarrow \forall i \in K: \langle \Sigma, \Sigma' \rangle \in Rw_i.$$

Sind die beiden Bedingungen der sicheren Ereignisse und der reinen Strategien nicht erfüllt, ist es naheliegend, das Bayessche Entscheidungskriterium oder ein vergleich-

bares Kriterium anzuwenden, um aufgrund der (vollständigen)  $Rw_1$ -Relation über  $\Xi$  die jeweilige  $Rw_1$ -Relation über  $S = \Sigma^1 \times \Sigma^2 \times \dots \times \Sigma^n$  abzuleiten. Allerdings beachte man, daß es sich hierbei um eine spezielle Art von Bewertungsfunktion handelt, die nicht mit dem 'abstrakten Nutzen' der normalen Entscheidungs- und Spieltheorie verwechselt werden darf.

Ist auch die zweite Bedingung erfüllt, hat man sich  $\Xi$  als n-Tupel individueller reiner Entscheidungen vorzustellen und nicht als n-Tupel individueller Entscheidungen, zu denen auch gemischte (individuelle) Strategien gehören. Der Übergang von  $\Sigma$  zu n-Tupeln gemischter individueller Strategien ist ein eigenes Thema.

Da die  $Kb_w$ -Relation alles andere als vollständig ist, ist sie für einen Vergleich von Zuständen in den (logisch) meisten Fällen ungeeignet.

Ein Zustand ist daher kollektiv rational nur bezüglich einer Klasse von Zuständen, mit denen er  $Kb_w$ -vergleichbar ist.

" $\xi$  ist kollektiv rational im engeren Sinne", definieren wir daher als 4-stellige Relation:

$$\langle \xi, \Xi^0, \Xi, K \rangle \in Kr_w \Leftrightarrow \forall \xi' \in \Xi^0 \setminus \{\xi\} : \langle \xi, \xi' \rangle \in Kb_w.$$

wobei  $\Xi^0 \subset \Xi \wedge \xi \in \Xi^0$

Und, da ja insbesondere interessiert, ob die Vergleichsklasse  $\Xi^0$  alle  $Kb_w$ -vergleichbaren Zustände umfaßt, definieren wir die maximale kollektive Rationalität im engeren Sinne:

$$\begin{aligned} \langle \xi, \Xi^0, \Xi, K \rangle \in Kr_w^* & \Leftrightarrow \\ \forall \xi' \in \Xi^0 \setminus \{\xi\} : \langle \xi, \xi' \rangle \in Kb_w & \wedge \\ \sim \exists \xi^* \in \Xi \setminus \Xi^0 : [\langle \xi^*, \xi \rangle \in Kb_w & \vee \langle \xi, \xi^* \rangle \in Kb_w] \\ \text{wobei } \Xi^0 \subset \Xi \wedge \xi \in \Xi^0 & \end{aligned}$$

Falls  $\xi$  bezüglich  $\Xi^0$ ,  $\Xi$  und  $K$  maximal kollektiv rational ist, gibt es keine kollektiv bessere Alternative in  $\Xi$  -  $\sim \exists \xi^* \in \Xi : \langle \xi^*, \xi \rangle \in Kb_w$  - und alle kollektiv schlechteren Alternativen sind in  $\Xi^0$  enthalten -  $\sim \exists \xi^* \in \Xi \setminus \Xi^0 : \langle \xi, \xi^* \rangle \in Kb_w$  -.

Die Entsprechung für den einfacheren Fall, in dem wir uns auf die Beurteilung der kollektiven Entscheidung beschränken können, ist dementsprechend:

$$\begin{aligned} &\langle \Sigma, S^0, S, K \rangle \in Kr_w \quad , \Leftrightarrow \\ &\forall \Sigma \in S^0 \setminus \{\Sigma\} : \langle \Sigma, \Sigma' \rangle \in Kb_w \wedge \\ &\sim \exists \Sigma^* \in S \setminus S^0 : [\langle \Sigma^*, \Sigma \rangle \in Kb_w \vee \langle \Sigma, \Sigma^* \rangle \in Kb_w] \\ &\text{wobei } S^0 \subset S \wedge \Sigma \in S^0 \end{aligned}$$

Anstelle von " $S^0 \setminus \{\Sigma\}$ " bzw. " $\Xi^0 \setminus \{\xi\}$ " hätte man auch bloß " $S^0$ " bzw. " $\Xi^0$ " setzen können, da die  $Rw_i$  ja schwache Präferenzrelationen sind. So lassen sich diese Präferenzrelationen jedoch problemlos abschwächen durch Einsetzen der starken individuellen Präferenzrelationen  $Pw_i$ .

Die kollektive Irrationalität definieren wir folgendermaßen

$$\begin{aligned} &\langle \xi, \Xi^0, \Xi, K \rangle \in \bar{K}r_w \quad , \Leftrightarrow \\ &\exists \xi' \in \Xi^0 : \langle \xi', \xi \rangle \in Kb_w \wedge \sim \langle \xi, \xi' \rangle \in Kb_w \\ &\text{wobei } \Xi^0 \subset \Xi \wedge \xi \in \Xi^0 \end{aligned}$$

bzw. im Sicherheitsfall:

$$\begin{aligned} &\langle \Sigma, S^0, S, K \rangle \in \bar{K}r_w \quad , \Leftrightarrow \\ &\exists \Sigma' \in S^0 : \langle \Sigma', \Sigma \rangle \in Kb_w \wedge \sim \langle \Sigma, \Sigma' \rangle \in Kb_w , \\ &\text{wobei } S^0 \subset S \wedge \Sigma \in S^0 \end{aligned}$$

Allerdings mag ein Zustand bezüglich einer Alternativenmenge  $\Xi^0$  in  $\Xi$  und einer Personengruppe  $K$  nicht kollektiv irrational sein, ohne damit kollektiv rational zu sein: Aus  $\sim \langle \xi, \Xi^0, \Xi, K \rangle \in \bar{K}r_w$  folgt nicht  $\langle \xi, \Xi^0, \Xi, K \rangle \in Kr_w$ , da die  $Kb_w$ -Relation nicht vollständig ist.

Über Alternativen, die bezüglich der Relation  $Kb_w$  nicht vergleichbar sind, läßt sich unter dem Aspekt der kollektiven Rationalität (in diesem hier definierten engeren Sinne) nichts aussagen. Es ist daher auch möglich, daß alle Alternativen weder kollektiv irrational noch kollektiv rational im engeren Sinne sind.

In endlichen Mengen  $Kb_w$ -vergleichbarer Alternativen gibt es jeweils mindestens eine kollektiv rationale Alternative, bzw. im Sicherheitsfall eine kollektiv rationale kollektive Entscheidung.

Anhand einiger einfacher Interaktionssituationen läßt sich der hier definierte Begriff kollektiver Rationalität erster Art verdeutlichen. Der Zwei-Personen-Fall läßt sich durch Matrizen darstellen.

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 2, 3 \rangle$	$\langle 4, 0 \rangle$
$\sigma_{12}$	$\langle 0, 4 \rangle$	$\langle 3, 2 \rangle$

Diese Interaktionssituation mit

$S = \{ \langle \sigma_{11}, \sigma_{12} \rangle, \langle \sigma_{11}, \sigma_{22} \rangle, \langle \sigma_{12}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle \}$   
 enthält kein  $S^0 \subset S$  - wobei  $\#S^0 > 1$  - mit einem kollektiv rationalen Element. Keine der vier hier möglichen Entscheidungen ist mit einer anderen  $Kb_w$ -vergleichbar. Für  $S^0$  mit  $\#S^0 = 1$  ist  $\langle \Sigma, S^0, S, K \rangle$  trivialerweise kollektiv rational.

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 1, 1 \rangle$	$\langle 4, 3 \rangle$
$\sigma_{12}$	$\langle 0, 3 \rangle$	$\langle 3, 4 \rangle$
$\sigma_{13}$	$\langle 2, 2 \rangle$	$\langle 5, 0 \rangle$

$S = \{ \langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{11}, \sigma_{22} \rangle, \langle \sigma_{12}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle, \langle \sigma_{13}, \sigma_{21} \rangle, \langle \sigma_{13}, \sigma_{22} \rangle \}$

$K = \{ 1, 2 \}$

In dieser Interaktionssituation haben folgende Mengen kollektiver Entscheidungen ein kollektiv rationales Element:

- Alle Einermengen

- Folgende Zweiermengen:

$\{ \langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{13}, \sigma_{21} \rangle \}$ , mit  $\Sigma_1 = \langle \sigma_{13}, \sigma_{21} \rangle$

$\{ \langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{11}, \sigma_{22} \rangle \}$ , mit  $\Sigma_2 = \langle \sigma_{11}, \sigma_{22} \rangle$

$\{ \langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle \}$ , mit  $\Sigma_3 = \langle \sigma_{12}, \sigma_{22} \rangle$

$\{ \langle \sigma_{12}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle \}$ , mit  $\Sigma_4 = \langle \sigma_{12}, \sigma_{22} \rangle$



$\{\langle \sigma_{13}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle\}$ , mit  $\Sigma_5 = \langle \sigma_{12}, \sigma_{22} \rangle$

$\{\langle \sigma_{13}, \sigma_{21} \rangle, \langle \sigma_{11}, \sigma_{22} \rangle\}$ , mit  $\Sigma_6 = \langle \sigma_{11}, \sigma_{22} \rangle$

- und die beiden Dreiermengen:

$S_1^0 = \{\langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{13}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle\}$ , mit

$\Sigma_1' = \langle \sigma_{12}, \sigma_{22} \rangle$

$S_2^0 = \{\langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{13}, \sigma_{21} \rangle, \langle \sigma_{12}, \sigma_{22} \rangle\}$ , mit

$\Sigma_2' = \langle \sigma_{11}, \sigma_{22} \rangle$

Maximal kollektiv rational sind

$\langle \Sigma_1', S_2^0, S, K \rangle$  und  $\langle \Sigma_2', S_2^0, S, K \rangle$ ;

$\Sigma_1'$  und  $\Sigma_2'$  sind nicht  $Kb_w$ -vergleichbar.

Obwohl wir "kollektiv rational" und "kollektiv irrational" als 4-stelligen Relationsbegriff eingeführt haben, sprechen wir auch in Zukunft davon, daß die kollektive Entscheidung bezüglich der Interessenstruktur einer Interaktionssituation irrational sei, ohne die 'Bezugsklassen' , und K anzugeben, wenn diese sich aus dem Kontext unmißverständlich ergeben.

Ist x kollektiv irrational in der hier definierten Bedeutung, so ist x auch intuitiv irrational. Der später eingeführte Begriff kollektiver Irrationalität zweiter Art umfaßt weitere Fälle, die auch intuitiv als irrational erscheinen. Ebenso gilt umgekehrt, daß es kollektive Entscheidungen gibt, die intuitiv rational sind, aber nicht unter den hier definierten Begriff fallen. Einige von diesen Entscheidungen, aber nicht alle, sind kollektiv rational im o.g. weiteren Sinne.

## 16. MORALITÄT ALS METAPRÄFERENZTHEORIE

Moralität unterscheidet sich von Rationalität u.a. darin, daß sie Kriterien für die Beurteilung von Präferenzen, von subjektiven Bewertungen enthält, während die Verhaltensrationalität im Sinne der Entscheidungstheorie beurteilt, ob Handlungen bzgl. jeweils gegebener Präferenzen bzw. subjektiver Bewertungen rational sind.<sup>1</sup>

Daraus könnte geschlossen werden, ein Konflikt zwischen Forderungen der Rationalität und der Moralität sei nicht möglich, weil sie Unterschiedliches zum Gegenstand haben. Viele Ökonomen und Entscheidungstheoretiker sind dieser Ansicht und sehen daher kein Problem in der Beziehung dieser beiden Arten normativer Kriterien.

Sen<sup>2</sup> hat dem zwei Argumente entgegengehalten.

(1) In einer Interaktionssituation, also in einer Situation, in der das Ergebnis der persönlichen Entscheidung von den Handlungen der anderen Personen abhängt, gibt es keine klare 'Übersetzung' von Präferenzen in Handlungen (die die betreffende Person wählen sollte), da die rationale Entscheidung von den Präferenzen der anderen beteiligten Personen abhängt. Andererseits gibt es auch Modelle der Moralität, die bestimmte Handlungen als unmoralisch oder unfair auszeichnen, und zwar auch bzgl. gegebener Präferenzen. Ein Konflikt zwischen Rationalitäts- und Moralitätsempfehlungen ist also nicht ausgeschlossen. Dies ist für Sen ein Grund, der eine genauere Untersuchung des Verhältnisses von moralischem und rationalem Handeln rechtfertigt.

(2) Der zweite Grund ist, daß in einigen Interaktionssituationen (wie z.B. solche vom Typ des Gefangenendilemmas) nach den üblichen spieltheoretischen Entschei-

---

<sup>1</sup> Wenn wir ohne Spezifizierung von "Präferenzen" sprechen, meinen wir Präferenzen über, bzw. Bewertungen von Zuständen - nicht von Handlungen.

<sup>2</sup> Vgl. A.K. Sen, *Choice, orderings and morality*, in: Körner (1974) 54-82.

dungskriterien ein Verhalten als individuell rational gilt, das ein für alle ungünstiges Ergebnis hat. Woraus Sen (in Übereinstimmung mit Rapoport) folgert, daß damit die Definition individuell rationalen Verhaltens äußerst schwierig wird<sup>1</sup> und "an attempt to escape from this problem through the use of the notion of collective - as opposed to individual - rationality would involve ideas that relate to the concept of morality."<sup>2</sup>

Sen geht nun das Problem an, indem er zwei - wie wir es einmal nennen wollen - 'ALS-OB-SPIELE' untersucht.

(1) DAS VERTRAUENSSPIEL (Assurance Game)<sup>3</sup>

Hier wird angenommen, die beiden Personen, die sich in einer Interaktionssituation vom Typ des Gefangenendilemmas befinden<sup>4</sup>, hätten folgende Präferenzen über die vier möglichen Konsequenzen, bzw. über die vier möglichen kollektiven Entscheidungen

$\langle \sigma_{11}, \sigma_{21} \rangle$ ,  $\langle \sigma_{12}, \sigma_{21} \rangle$ ,  $\langle \sigma_{11}, \sigma_{22} \rangle$ ,  $\langle \sigma_{12}, \sigma_{22} \rangle$ .<sup>5</sup>

A hat die Präferenzen:

$\langle \sigma_{12}, \sigma_{22} \rangle > \langle \sigma_{11}, \sigma_{22} \rangle > \langle \sigma_{11}, \sigma_{21} \rangle > \langle \sigma_{12}, \sigma_{21} \rangle$

B hat die Präferenzen:

$\langle \sigma_{12}, \sigma_{22} \rangle > \langle \sigma_{12}, \sigma_{21} \rangle > \langle \sigma_{11}, \sigma_{21} \rangle > \langle \sigma_{11}, \sigma_{22} \rangle$ .

---

<sup>1</sup> Diese Folgerung ist nicht zwingend; vgl. §12

<sup>2</sup> Dies ist die Gegenposition zum Harsanyi/Selten-Konzept der Entscheidungsrationalität in Interaktionssituationen. Natürlich ist auch in Situationen eines Spieles 'gegen die Natur' das Ergebnis der persönlichen Entscheidung abhängig von den 'Handlungen der Natur', von den Ereignissen; dennoch gibt es in diesem Fall eine 'klare Übersetzung' von Präferenzen zu Handlungsempfehlungen gemäß der Bayesschen Entscheidungsregel oder einem modifizierten Kriterium, wie wir es z.B. im Absatz vorgeschlagen haben. Sen geht also davon aus, daß die Entscheidungen anderer Personen in einer Interaktionssituation nicht wie 'natürliche Ereignisse' (in einer Theorie rationaler Entscheidung) behandelt werden können.

<sup>3</sup> Vgl. Sen (1974a).

<sup>4</sup> Im folgenden kurz als "die Gefangenen" bezeichnet.

<sup>5</sup> Wobei  $\sigma_{11}$ : die erste Person (A) gesteht  
 $\sigma_{12}$ : A gesteht nicht  
 $\sigma_{21}$ : die zweite Person (B) gesteht  
 $\sigma_{22}$ : B gesteht nicht

Soweit der andere sich kooperativ verhält, zieht es jeder vor, sich ebenfalls kooperativ zu verhalten. Verhält sich der andere nicht kooperativ, zieht es jeder vor, sich ebenfalls nicht kooperativ zu verhalten. Voraussetzung für das Zustandekommen der kooperativen kollektiven Entscheidung  $\langle \sigma_{12}, \sigma_{22} \rangle$  ist also das wechselseitige Vertrauen darauf, daß sich der andere kooperativ verhält.<sup>1</sup>

Zum Vergleich: Die Präferenzen im Gefangenendilemma sind:

$$\begin{array}{l} \langle \sigma_{11}, \sigma_{22} \rangle \underset{1}{\succ} \langle \sigma_{12}, \sigma_{22} \rangle \underset{1}{\succ} \langle \sigma_{11}, \sigma_{21} \rangle \underset{1}{\succ} \langle \sigma_{12}, \sigma_{21} \rangle \\ \langle \sigma_{12}, \sigma_{21} \rangle \underset{2}{\succ} \langle \sigma_{12}, \sigma_{22} \rangle \underset{2}{\succ} \langle \sigma_{11}, \sigma_{21} \rangle \underset{2}{\succ} \langle \sigma_{11}, \sigma_{22} \rangle \end{array}$$

Im Vertrauensspiel gibt es nun im Gegensatz zum Gefangenendilemma zwei Gleichgewichtspunkte:  $\langle \sigma_{11}, \sigma_{21} \rangle$  u.  $\langle \sigma_{12}, \sigma_{22} \rangle$ . Da die individuelle Entscheidung für die kooperative, bzw. nicht-kooperative Entscheidung davon abhängt, welche Entscheidung von der anderen Person erwartet wird, ist es keineswegs garantiert, daß es zu einer kollektiven Gleichgewichtsentscheidung kommt. Nur bei gleichen subjektiven Erwartungen der Beteiligten wird eine Gleichgewichtsentscheidung erfolgen. Ein wesentliches Merkmal dieses Spieles ist, daß die kooperative Entscheidung durch eine Abmachung ohne Sanktionen erreicht wird (im Gegensatz zum Gefangenendilemma).

## (2) DAS ALTRUISMUSSPIEL<sup>2</sup>

Es wird angenommen, die Gefangenen<sup>3</sup> hätten folgende Präferenzen:

$$\begin{array}{l} \langle \sigma_{12}, \sigma_{22} \rangle \underset{1}{\succ} \langle \sigma_{12}, \sigma_{21} \rangle \underset{1}{\succ} \langle \sigma_{11}, \sigma_{22} \rangle \underset{1}{\succ} \langle \sigma_{11}, \sigma_{21} \rangle \\ \langle \sigma_{12}, \sigma_{22} \rangle \underset{2}{\succ} \langle \sigma_{11}, \sigma_{22} \rangle \underset{2}{\succ} \langle \sigma_{12}, \sigma_{21} \rangle \underset{2}{\succ} \langle \sigma_{11}, \sigma_{21} \rangle \end{array} \quad ^4$$

<sup>1</sup> Sen nennt dieses Spiel 'assurance game'. Im Deutschen erscheint mir jedoch "Vertrauensspiel" treffender als "Versicherungsspiel".

<sup>2</sup> Bei Sen 'other regarding', vgl. Sen (1974a).

<sup>3</sup> Vgl. S.224, FN 4.

<sup>4</sup> Vgl. S.224, FN 5.

Unabhängig von der Entscheidung des anderen Gefangenen wird die kooperative individuelle Entscheidung von beiden Gefangenen vorgezogen. Die kooperative Entscheidung ist hier die dominante Strategie, während im Gefangenendilemma die nicht-kooperative Entscheidung die dominante Strategie ist.

Vorläufige SCHLUSSFOLGERUNG:

- (1) Im Vertrauensspiel kommt es zur kollektiv besten, zur kollektiv rationalen Entscheidung, wenn ein entsprechendes wechselseitiges Vertrauen in die Kooperationsbereitschaft des anderen besteht. Es gibt keine Veranlassung, einer entsprechenden Vereinbarung oder auch Konvention zuwiderzuhandeln. Falls wechselseitiges Vertrauen in die Kooperationsbereitschaft besteht, ist das Ergebnis, die kollektive Entscheidung  $\langle \sigma_{12}, \sigma_{22} \rangle$  nicht nur kollektiv rational bezüglich der individuellen Präferenzen des Vertrauensspieles (= V-Präferenzen), sondern ebenfalls bezüglich der individuellen Präferenzen des Gefangenendilemmas (= G-Präferenzen). Würden sich die Gefangenen aber verhalten als ob sie V-Präferenzen hätten und würden sie von einer Kooperationsbereitschaft wechselseitig ausgehen, so kämen sie zur kollektiv rationalen Entscheidung der Interaktionssituation Gefangenendilemma, während sie kollektiv irrational entscheiden, falls sie ihre Entscheidung an den G-Präferenzen orientieren.
  
- (2) Im Altruisspiel hingegen ist die Bedingung des wechselseitigen Vertrauens in die Kooperationsbereitschaft entbehrlich. Hier ist die kooperative kollektive Entscheidung  $\langle \sigma_{12}, \sigma_{22} \rangle$  dominant. Diese Entscheidung ist darüberhinaus ebenfalls kollektiv rational sowohl bezüglich der individuellen Präferenzen des Altruisspieles (A-Präferenzen) wie auch bezüglich der G-Präferenzen.

Die Gefangenen würden also besser fahren, wenn sie sich so verhalten würden, als ob sie A-Präferenzen hätten.

Bezüglich der V-Präferenzen und der A-Präferenzen ist die kollektive Entscheidung  $\langle \sigma_{12}, \sigma_{22} \rangle$  das für jeden individuell bestmögliche Ergebnis.  $\langle \sigma_{12}, \sigma_{22} \rangle$  ist bezüglich der G-Präferenzen zwar nicht die individuell beste Konsequenz, aber  $\langle \sigma_{12}, \sigma_{22} \rangle$  ist für die Beteiligten individuell besser als das Ergebnis individuell rationalen Entscheidungsverhaltens  $\langle \sigma_{11}, \sigma_{21} \rangle$ . V-Präferenzen garantieren das kollektiv rationale Ergebnis bezüglich der G-Präferenzen (das auch kollektiv rational bezüglich V- und A-Präferenzen ist, aber das ist hier unwesentlich) unter der epistemischen Bedingung wechselseitiger Kooperationserwartung, und A-Präferenzen garantieren das kollektiv rationale Ergebnis sogar ohne jede epistemische Bedingung.

Welche Konsequenzen ergeben sich daraus vorläufig für die Beziehung von Rationalität und Moralität?

Es gibt Interaktionssituationen, in denen individuell rationales Verhalten bezüglich anderer Präferenzen als sie die Handlungen tatsächlich haben, zu kollektiv rationalen Ergebnissen führt.

Ist es deshalb sinnvoll, eine moralische Ordnung der Präferenzen, anstatt einer moralischen Ordnung der Ergebnisse (z.B. aufgrund von Fairnessprinzipien) vorzunehmen? In unserem Beispiel würden dann die A-Präferenzen moralisch vor den V-Präferenzen, und die V-Präferenzen vor den G-Präferenzen rangieren.

Obwohl Sen so weit geht, festzustellen, daß "social inoptimality might be avoidable only by a moral code of behaviour that drives a wedge between preference and welfare" und die übliche Gleichsetzung individueller Präferenzen "as reflection of individual welfare" kritisiert und es angesichts dieser Ergebnisse für inad-

äquat hält "to treat welfare as a numerical representation of preferences revealed by individual choices", ja von der Notwendigkeit einer "dichotomy between revealed preferences and welfare"<sup>1</sup> überzeugt ist; was ja ganz im Sinne der Überlegungen ist, die wir im ersten Teil der Arbeit angestellt haben,<sup>2</sup> bleibt er dennoch bei einer "als ob"-Interpretation der Ergebnisse. Sen plädiert zwar für eine moralische Präordnung über die Menge der möglichen individuellen Präferenzordnungen. Während die individuellen Präferenzordnungen auf der Menge der Konsequenzen definiert sind, ist die moralische Präordnung<sup>3</sup> auf der Menge der individuellen Präferenzrelationen definiert. Jedoch ist die Interpretation dieser moralischen Präferenzordnung dunkel.

Naheliegend ist es, sie im Anschluß an das diskutierte Beispiel als eine Zusammenfassung impliziter Handlungsempfehlungen zu interpretieren. Denn die tatsächliche Interaktionssituation ist ja vom Typ des Gefangenendilemmas. Moralisch empfohlen wird kollektiv rationales Entscheidungsverhalten. Diese Empfehlung ist jedoch in eine besondere (spieltheoretische) Form gebracht: es wird ein Verhalten empfohlen, das individuell rational ist bezüglich bestimmter Als-ob-Präferenzen.

---

<sup>1</sup> Sen (1974 a), S.62

<sup>2</sup> Vgl. bes. §7 und §8

<sup>3</sup> Sen spricht von *Quasiordnung* und meint damit eine reflexive transitive, also nicht notwendigerweise konnexe zwei-stellige Relation über die Menge der möglichen Ordnungsrelationen (über die Menge der Konsequenzen). Reflexive und transitive binäre Relationen werden i.a. als Präordnungen bezeichnet, reflexive, transitive und asymmetrische binäre Relationen als Ordnungen. Das was Sen "Ordnung" nennt, wird i.a. als "totale", "vollständige" oder "konnexe" Ordnung bezeichnet. (Vgl. z.B. Kasch/Pareigis 1974); zur Terminologie zwei-stelliger Relationen vgl. Mathematischen Appendix. Diese moralische Präferenzrelation ist deshalb im Gegensatz zu den individuellen (overall-)Präferenzrelationen i.a. nicht konnex (oder vollständig), weil die *Vergleichbarkeit unterschiedlicher Präferenzordnungen aufgrund moralischer Kriterien nicht immer möglich ist.*

Es ist hier letztlich nicht wesentlich, ob das die einzig vertretbare Deutung des *Senschen* Ansatzes ist.<sup>1</sup> Wesentlich ist, daß es zwei Möglichkeiten gibt, eine moralische Metapräferenz zu deuten, die sorgfältig auseinanderzuhalten sind.

Die erste läßt sich folgendermaßen skizzieren: Die spieltheoretischen Rationalitätskriterien werden als generelle Verhaltensempfehlungen aufgegeben. Aufgrund gegebener Präferenzen gibt es moralische Entscheidungskriterien, die gegen die Gebote individueller Rationalität verstoßen können und die sich u.U. als Ergebnis individuell rationalen Entscheidens bezüglich einer Als-ob-Präferenzrelation (über die Menge der Konsequenzen) zusammenfassen lassen. Da es unter diesen Als-ob-Präferenzrelationen bessere und schlechtere gibt - A-Präferenzen sind z.B. besser als V-Präferenzen - konstituiert dies eine moralische Metapräferenzrelation, eine Präordnung von individuellen konnexen Präferenzordnungen.

Die Frage bleibt, ob damit die Als-ob-Präferenzen selbst sinnvoll als moralische Präferenzen bezeichnet werden können. Moralisch beurteilt werden ja zunächst die Entscheidungen bei gegebener Präferenz- und Interaktionsstruktur (hier dem Gefangenendilemma), des weiteren gibt es eine moralische Beurteilung der verschiedenen Als-ob-Präferenzen, nämlich je nachdem, mit welcher Sicherheit und in welchem Ausmaß diese Als-ob-Präferenzen die moralisch gebotene kollektive Entscheidung garantieren. Damit steht jedoch in keiner Weise fest, daß die Als-ob-Präferenzen selbst moralische Präferenzen sind. Sie beurteilen Konsequenzen, aber beurteilen sie diese Konsequenzen aufgrund moralischer Kriterien? Lautet die Antwort "Ja", so hätten wir drei Gegenstände (Ebenen) moralischer Beurteilung:

---

<sup>1</sup> Um diese Frage hatte sich eine interessante Diskussion entsponnen. Vgl. J.W.N. Watkins, *Comment: 'Self-interest and morality'*, in: Körner (1974) 67-77 und *Sens* Erwiderung, ebd. 78-82, in der deutlicher wird, daß die 'als-ob'-Interpretation bzw. die erste der beiden Interaktionsmöglichkeiten die von Sen intendierte ist. Vgl. dazu a. die interessante Kritik Kurt Baiers, *Rationality and morality*, *Erkenntnis* 11 (1977) 197-223.



- (1) Konsequenzen (im Sicherheitsfall also kollektive Entscheidungen), die von moralischen (Als-ob-) Präferenzen beurteilt werden,
- (2) Individuelle Entscheidungen aufgrund gegebener (sozusagen 'primärer') Präferenzen (die in dem o.g. Beispiel Bestandteil des Gefangenendilemmas sind),
- (3) Individuelle Präferenzrelationen, die von moralischen Metapräferenzen (der *Senschen* Quasiordnung) in eine moralische Ordnung gebracht werden.

Es ist nicht leicht, die Begründungsbeziehung dieser drei Ebenen zu bestimmen.

Im Fall des Gefangenendilemmas ist die moralische Bewertung der Konsequenzen einfach: die kooperative kollektive Entscheidung ist aufgrund des Pareto-Kriteriums besser als die nicht-kooperative kollektive Entscheidung (Ebene 1).<sup>1</sup> Schon der (normativ-ethische, nicht logisch-deduktive) 'Schluß' von (1) darauf, daß auch die individuelle kooperative Entscheidung besser ist als die individuelle nicht-kooperative (Ebene 2), ist nicht selbstverständlich, da dessen Plausibilität von der moralischen Bewertung der asymmetrischen kollektiven Entscheidung abhängt.

Mangels Information (Ordinalität und interpersonelle Unvergleichbarkeit) müssen die beiden asymmetrischen kollektiven Entscheidungen untereinander moralisch gleich bewertet werden:  $\langle \sigma_{11}, \sigma_{22} \rangle \approx \langle \sigma_{12}, \sigma_{21} \rangle$ , aber sie könnten zwischen die beiden symmetrische Entscheidungen oder erst an die dritte Stelle gesetzt werden:

$$\text{Rm: } \langle \sigma_{12}, \sigma_{22} \rangle \underset{\perp}{\succ} \langle \sigma_{11}, \sigma_{22} \rangle \underset{\perp}{\succ} \langle \sigma_{11}, \sigma_{21} \rangle \\ \underset{\perp}{\succ} \langle \sigma_{12}, \sigma_{21} \rangle$$

$$\text{Rm: } \langle \sigma_{12}, \sigma_{22} \rangle \underset{\perp}{\succ} \langle \sigma_{11}, \sigma_{21} \rangle \underset{\perp}{\succ} \langle \sigma_{11}, \sigma_{22} \rangle \\ \underset{\perp}{\succ} \langle \sigma_{12}, \sigma_{21} \rangle$$

---

<sup>1</sup> Das Pareto-Kriterium selbst ist allerdings strenggenommen erst in Bezug auf die Wohlfahrtspräferenzen und nicht bei resultierenden Präferenzen intuitiv überzeugend, wie das liberale Paradoxon gezeigt hat, s. §9.

Werden die asymmetrischen kollektiven Entscheidungen gemäß  $R_M$  schlechter als die nicht-kooperative kollektive Entscheidung bewertet, so ist die darauf gründende moralische Bewertung der individuellen Entscheidung nicht eindeutig. Je nach der Wahrscheinlichkeit  $\rho$ , mit der der andere Beteiligte kooperativ entscheidet, könnte sowohl  $\sigma_{12} \succ \sigma_{11}$  (falls  $\rho$  nahe 1) wie  $\sigma_{11} \succ \sigma_{12}$  (falls  $\rho$  nahe 0) gerechtfertigt werden.

Nehmen wir an, die moralischen Beurteilungen der Ebene 2 seien aufgrund der moralischen Beurteilungen der Ebene 1 möglich gewesen, dann fehlt immer noch der letzte Schritt zu einer moralischen Metapräferenzrelation (Sens 'moral quasiordering of orderings'), der die individuellen Präferenzordnungen aufgrund der moralischen Beurteilung der individuellen Entscheidungen in eine moralische Rangordnung bringt. Wobei der Moralphilosoph schon sehr zufrieden wäre, wenn ihm auf diese Weise eine eher rudimentäre moralische Metapräferenzrelation zur Verfügung stünde, die je eine individuelle Präferenzordnung als die moralisch gebotene und alle übrigen als moralisch verboten bestimmen könnte.<sup>1</sup>

Abgesehen von dem Interpretationsproblem der Präferenzrelationen, die in eine moralische Ordnung gebracht werden sollen, - offensichtlich dreht es sich ja hier nicht um tatsächliche, sondern um fiktive Präferenzen - stellt sich die Frage, was die Beurteilungsgrundlage ist: die moralische Konsequenzenbewertung oder die moralische Bewertung der individuellen Entscheidungen. Die von Sen intendierte Begründungsbeziehung führt wohl von (1) zu (2) und (vermittels des spieltheoretischen Instrumentariums) von (2) zu (3).

---

<sup>1</sup> Die große Unvollständigkeit dieser Metapräferenzrelation - auf einer Menge von 10 Präferenzordnungen einer Person sind von 45 geordneten Paaren nur 9 Element der Metapräferenzrelation - sollte jedoch nicht dadurch behoben werden, daß man zwischen zwei moralisch verbotenen Präferenzordnungen moralische Indifferenz annimmt, denn das würde weiteren moralischen Erkenntnisfortschritt, der auch eine Differenzierung in mehr oder weniger unmoralisch zuläßt, von vorneherein ausschließen.

Diese normative Metapräferenztheorie scheitert m.E. an folgendem: Eine Interaktionssituation von  $n$  Beteiligten ist definiert durch die individuellen (reinen) Strategiemengen  $\Sigma = \Sigma^1 \times \Sigma^2 \times \dots \times \Sigma^n$  und den Nutzen, den die Beteiligten aus den jeweiligen Strategiekombinationen (den Endpunkten des Spielbaums in extensiver Form)  $\{\langle \sigma_1 \times \sigma_2 \times \sigma_3 \times \dots \times \sigma_n \rangle\}$  ziehen<sup>1</sup>. Der Einfachheit halber wollen wir hier annehmen, daß die, für die Nutzenbewertung der Beteiligten relevanten Merkmale der erreichten Zustände völlig durch die Strategiekombinationen bzw. die kollektiven Entscheidungen bestimmt sind (deterministischer Fall).

Auch das Gefangenendilemma ist durch eine bestimmte individuelle Nutzenbewertung der kollektiven Entscheidungen definiert:

$$\begin{aligned} &U_1(\langle \sigma_{11}, \sigma_{22} \rangle) > U_1(\langle \sigma_{12}, \sigma_{22} \rangle) > U_1(\langle \sigma_{11}, \sigma_{21} \rangle) > U_1(\langle \sigma_{12}, \sigma_{21} \rangle) \wedge \\ \wedge &U_2(\langle \sigma_{12}, \sigma_{21} \rangle) > U_2(\langle \sigma_{12}, \sigma_{22} \rangle) > U_2(\langle \sigma_{11}, \sigma_{21} \rangle) > U_2(\langle \sigma_{11}, \sigma_{22} \rangle) \end{aligned}$$

Als-ob-Präferenzen ziehen eine entsprechende andere Bewertung der kollektiven Entscheidungen nach sich.

A-Präferenzen:

$$\begin{aligned} &U_1(\langle \sigma_{12}, \sigma_{22} \rangle) > U_1(\langle \sigma_{12}, \sigma_{21} \rangle) > U_1(\langle \sigma_{11}, \sigma_{22} \rangle) > U_1(\langle \sigma_{11}, \sigma_{21} \rangle) \wedge \\ \wedge &U_2(\langle \sigma_{12}, \sigma_{22} \rangle) > U_2(\langle \sigma_{11}, \sigma_{22} \rangle) > U_2(\langle \sigma_{12}, \sigma_{21} \rangle) > U_2(\langle \sigma_{11}, \sigma_{21} \rangle) \end{aligned}$$

V-Präferenzen:

$$\begin{aligned} &U_1(\langle \sigma_{12}, \sigma_{22} \rangle) > U_1(\langle \sigma_{11}, \sigma_{22} \rangle) > U_1(\langle \sigma_{11}, \sigma_{21} \rangle) > U_1(\langle \sigma_{12}, \sigma_{21} \rangle) \wedge \\ \wedge &U_2(\langle \sigma_{12}, \sigma_{22} \rangle) > U_2(\langle \sigma_{12}, \sigma_{21} \rangle) > U_2(\langle \sigma_{11}, \sigma_{21} \rangle) > U_2(\langle \sigma_{11}, \sigma_{22} \rangle) \end{aligned}$$

Sind diese Präferenzen moralische Präferenzen (unterschiedlicher 'Dignität'), so ergibt sich eine Ordnung der kollektiven Entscheidungen unter moralischen Gesichtspunkten:

<sup>1</sup>  $\sigma_i \in \Sigma^i$ , im abzählbaren Fall  $\sigma_{1k} \in \Sigma^1$ ,  $k \in \{1, 2, 3, \dots, \#\Sigma^1\}$

Falls die 'Auszahlungen' allein durch die Strategiekombinationen bestimmt sind, erhalten wir eine Vektorfunktion  $U$  von

$\Sigma^1 \times \Sigma^2 \times \Sigma^3 \times \dots \times \Sigma^v \ni \Sigma \rightarrow U\{\Sigma\} \in \mathbb{R}^v$ ,  $v = \#K$

für die erste Person:

$\langle \sigma_{12}, \sigma_{22} \rangle > \langle \sigma_{12}, \sigma_{21} \rangle > \langle \sigma_{11}, \sigma_{22} \rangle > \langle \sigma_{11}, \sigma_{21} \rangle$

für die zweite Person:

$\langle \sigma_{12}, \sigma_{22} \rangle > \langle \sigma_{11}, \sigma_{22} \rangle > \langle \sigma_{12}, \sigma_{21} \rangle > \langle \sigma_{11}, \sigma_{21} \rangle$

bei A-Präferenzen und bei V-Präferenzen:

für die erste Person:

$\langle \sigma_{12}, \sigma_{22} \rangle > \langle \sigma_{11}, \sigma_{22} \rangle > \langle \sigma_{11}, \sigma_{21} \rangle > \langle \sigma_{12}, \sigma_{21} \rangle$

für die zweite Person:

$\langle \sigma_{12}, \sigma_{22} \rangle > \langle \sigma_{12}, \sigma_{21} \rangle > \langle \sigma_{11}, \sigma_{21} \rangle > \langle \sigma_{11}, \sigma_{22} \rangle$

Die moralische Metapräferenzrelation sagt uns dann, daß die individuellen Ordnungen der ersten Art gegenüber den individuellen Ordnungen der zweiten Art moralisch besser sind. Nur: Was ist der Gegenstand der moralischen Beurteilung? Ist der Gegenstand in den tatsächlichen Präferenzen zu suchen, die das Gefangenendilemma konstituieren?

Sollte das der Fall sein, so ist nicht einsichtig, warum es dann nicht Bestandteil der normativen Theorie ist, daß sich die tatsächlichen Präferenzen an der moralisch besten Präferenzrelation (über die Menge der kollektiven Entscheidungen) orientieren sollten. (Hier den A-Präferenzen) Damit hätten wir keine Interaktionssituation vom Typ des Gefangenendilemmas mehr vor uns, sondern eine ganz andere, in der das kollektiv Rationale mit dem individuell Rationalen übereinstimmt.

Oder ist der Gegenstand moralischer Beurteilung in den Ergebnissen der kollektiven Entscheidungen in Gestalt von z.B. Gefängnisjahren zu suchen? Damit hätten wir uns das hoffnungslose Problem einer unübersehbaren Vielfalt von Einzelfallbeurteilungen aufgehalst. Einer moralischen Metapräferenztheorie geht es nicht um die Verteilung von Gefängnisjahren, Geld und anderen Gütern, sondern um den moralisch gebotenen, den fairen Interessenausgleich. D.h., Gegenstand der moralischen Beurteilung ist nicht die Verteilung von Gütern, sondern die Verteilung der persönlichen Bedürfnisbefriedigung, oder wie immer man das nennen möchte. Die moralische Beurteilung der Verteilung von Gütern ist in diesem Sinne sekundär: Man muß den Zusammenhang von persönlichem Wohlergehen und der persön-

lichen Verfügung über einzelne Güter kennen, um beurteilen zu können, welche Güterverteilung gerecht ist. Üblicherweise werden die Interessen in Form von individuellen Nutzenfunktionen zusammengefaßt. Theorien kollektiver Wahl, besonders ihr Spezialfall, die moderne ökonomische Fairneßtheorie<sup>1</sup> machen davon Gebrauch. Wir haben im ersten Teil dieser Arbeit begründet, warum das nicht zulässig ist, und, wie die oben angeführten Zitate zeigen, ist offensichtlich auch Sen dieser Ansicht. Interessen müssen daher in speziellen Wohlfahrtsfunktionen dargestellt werden, die mit den v. Neumann-Morgensternschen Nutzenfunktionen außer der Kardinalität nicht viel gemein haben: Sie passen nicht in das Konzept der "revealed preferences", und selbst in Spielen gegen die Natur ist ihre Erwartungswertmaximierung kein Kriterium rationaler Entscheidung.

Und damit kommen wir zur zweiten Möglichkeit: Die normative Metapräferenztheorie als Ethik zweiter Ordnung.

- (1) Die moralische Metapräferenzrelation bezieht sich auf die resultierenden oder die subjektiv-moralischen Präferenzen.
- (2) Die individuellen resultierenden Präferenzen werden (normativ) beurteilt aufgrund der gegebenen Interessenkonstellation.<sup>2</sup>
- (3) Eine Interessenkonstellation ist - analog zur Normalform eines Spiel in der Spieltheorie - bestimmt durch die Anzahl der Spieler, ihre Strategienpotentiale und ihre Wohlfahrtsfunktionen (in der Bedeutung, wie wir sie in § 9 festgelegt haben).

---

<sup>1</sup> Vgl. u. a.:

Foley (1967), Braithwaite (1969), Kilgour (1974), Varian (1974), Varian (1975), Pazner (1976), Daniel (1975), Svensson (1977), Pazner u. Schmeidler (1978), Barry (1979), Wittmann (1979).

<sup>2</sup> I. a. durch die Vermittlung bestimmter normativer Kriterien, falls es sich nicht um *moralische Basisurteile* handelt.

Am Beispiel des Gefangenendilemmas sieht das folgendermaßen aus:

Die Entscheidungspotentiale  $\{\sigma_{11}, \sigma_{12}\}$ ,  $\{\sigma_{21}, \sigma_{22}\}$  und die Wohlfahrtsfunktionen  $\pi_1, \pi_2$  konstituieren das Gefangenendilemma

$$\pi_1(\langle \sigma_{11}, \sigma_{22} \rangle) > \pi_1(\langle \sigma_{12}, \sigma_{22} \rangle) > \pi_1(\langle \sigma_{11}, \sigma_{21} \rangle) > \pi_1(\langle \sigma_{12}, \sigma_{21} \rangle) \\ \wedge \pi_2(\langle \sigma_{12}, \sigma_{21} \rangle) > \pi_2(\langle \sigma_{12}, \sigma_{22} \rangle) > \pi_2(\langle \sigma_{11}, \sigma_{21} \rangle) > \pi_2(\langle \sigma_{11}, \sigma_{22} \rangle)$$

Zwar haben wir volle interpersonelle Vergleichbarkeit der individuellen Wohlfahrtsfunktionen angenommen<sup>1</sup>, aber hier wollen wir die Analyse zunächst möglichst allgemein halten, und es daher bei einer ordinalen (und interpersonell nicht vergleichbaren) Skala belassen.

Die Matrix der Interaktionssituation hat damit folgende Gestalt:

B

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 2, 2 \rangle$	$\langle 4, 1 \rangle$
$\sigma_{12}$	$\langle 1, 4 \rangle$	$\langle 3, 3 \rangle$

A

Die von Sen geforderte Dichotomie von "revealed preferences and welfare" ist hier durchgeführt. Das heißt aber auch, daß es nicht erlaubt ist, aus dem tatsächlichen Entscheidungsverhalten Rückschlüsse auf die individuellen Wohlfahrtsbewertungen der Konsequenzen zu ziehen. Diese Matrix sagt nichts darüber aus, welche kollektive Entscheidung von den Beteiligten für die beste, zweitbeste, ... gehalten wird, sondern sie stellt nur dar, welcher Zustand, welche kollektive Entscheidung den jeweiligen persönlichen Interessen mehr entspricht.

Aus  $\pi_1(\langle \sigma_{11}, \sigma_{22} \rangle) = 4$   $\pi_1(\langle \sigma_{12}, \sigma_{22} \rangle) = 3$  folgt nicht, daß A die kollektive Entscheidung  $\langle \sigma_{11}, \sigma_{22} \rangle$  für besser hält als die kollektive Entscheidung  $\langle \sigma_{12}, \sigma_{22} \rangle$ , sondern bloß daß A durch  $\langle \sigma_{11}, \sigma_{22} \rangle$  besser gestellt ist als durch  $\langle \sigma_{12}, \sigma_{22} \rangle$ . Im Falle  $\langle \sigma_{11}, \sigma_{22} \rangle$  sitzt A eben weniger Jahre im Gefängnis als im Fall  $\langle \sigma_{11}, \sigma_{21} \rangle$ .

<sup>1</sup> vgl. §9

Damit ist nicht unvereinbar, daß A  $\langle \sigma_{11}, \sigma_{21} \rangle$  (die kooperative Entscheidung) für die moralisch bessere -  $\langle \langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{11}, \sigma_{22} \rangle \rangle \in Rm_1$  - und u.U. auch für die resultierend (overall-)bessere -  $\langle \langle \sigma_{11}, \sigma_{21} \rangle, \langle \sigma_{11}, \sigma_{22} \rangle \rangle \in Rres_1$  kollektive Entscheidung hält. Die moralische Metapräferenzrelation kann sowohl die (subjektive) moralische Präferenzrelation wie auch die resultierenden Präferenzrelationen der Beteiligten beurteilen, denn resultierende Präferenzen können genauso moralisch besser oder schlechter sein wie subjektive moralische Präferenzen;

Die subjektiven moralischen Präferenzen ( über die Menge der Konsequenzen, bzw. kollektiven Entscheidungen ) geben an, welche Wohlfahrtsverteilung die betreffende Person (insgesamt gesehen) 'resultierend' vorzieht, hier können moralische Aspekte und persönliche Interessen in unterschiedlichem Ausmaß eine Rolle spielen. Die resultierenden wie die subjektiv-moralischen Präferenzen einer Person können moralisch beurteilt werden. Nur die resultierenden Präferenzen sind mit individuellen Entscheidungen eng verknüpft (wie, das versucht die rationale Entscheidungs- und Spieltheorie zu ergründen). Die moralisch gebotene individuelle Entscheidung und damit die (resultierende) Präferenzrelation  $Rres_i$  einerseits und die moralisch gebotene (moralische) Präferenzrelation  $Rm_i$  andererseits hängen dabei nicht so eng zusammen wie manche ethischen Theorien offensichtlich annehmen. Typische Fragen an die Präferenzrelation  $Rm_i$  lauten: Welche Güterverteilung (die sich aus den kollektiven Entscheidungen ergibt) ist gerecht? Was ist bei dieser Interaktionssituation ein fairer Interessenausgleich? etc. Diese Fragen können unabhängig davon beantwortet werden, ob die moralisch richtigen Entscheidungen bekannt sind oder nicht. Die moralischen Präferenzen über die Menge der kollektiven Entscheidungen bzw. über die Menge der Konsequenzen konstituieren eine moralische Bewertungsfunktion. Und so wie es einer eigenen wissenschaftlichen Disziplin, der Spieltheorie, bedarf, um bei gegebenen individuellen

Nutzenfunktionen die individuell rationalen Entscheidungen in der jeweiligen Interaktionssituation zu bestimmen, so ist die Bestimmung der jeweils moralisch gebotenen individuellen Entscheidungen aufgrund gegebener moralischer Bewertungen ein Problem für sich.



## 17. KOLLEKTIVE RATIONALITÄT ALS MINIMALMORAL

Eine Mindestbedingung moralischer Beurteilung ist, daß ein kollektiv irrationaler Zustand nicht höher bewertet wird als ein kollektiv rationaler. Die moralische distributive Beurteilung individueller Präferenzen und Entscheidungen sollte dieser Minimalmoral entsprechen.

Wir haben im letzten Abschnitt den Ansatz einer normativen Metapräferenztheorie erster Ordnung kritisiert. Nun wollen wir prüfen, was eine normative Metapräferenztheorie zweiter Ordnung zur Klärung dieses Problems beitragen kann. Auch in diesem Abschnitt bleiben wir bei dem konkreten Beispiel des Gefangenendilemmas in Wohlfahrtspräferenzen.

Das Gefangenendilemma ist eine Interaktionssituation mit zwei Personen und je zwei Strategien, also vier kollektiven Entscheidungen:  $\langle \sigma_{11}, \sigma_{21} \rangle$ ,  $\langle \sigma_{12}, \sigma_{21} \rangle$ ,  $\langle \sigma_{11}, \sigma_{22} \rangle$ ,  $\langle \sigma_{12}, \sigma_{22} \rangle$ . Wir stellen uns vor, daß die Konsequenzen bzw. die für die subjektiven Bewertungen relevanten Merkmale der Zustände (kurz: die Zustände) eindeutig durch die Strategienkombinationen (= die kollektiven Entscheidungen) der Beteiligten bestimmt sind. Jedes Individuum hat  $4! = 24$  Möglichkeiten, vier Zustände bzw. vier kollektive Entscheidungen in eine konnexe Präferenzordnung zu bringen. Insgesamt ergeben sich also  $24 \times 24 = 576$  logisch mögliche Kombinationen individueller Präferenzordnungen.

Würden wir das Gefangenendilemma nicht durch eine bestimmte Präferenzstruktur,

$$\begin{array}{ccccccc} \langle \sigma_{11}, \sigma_{22} \rangle & \underset{1}{>} & \langle \sigma_{12}, \sigma_{22} \rangle & \underset{1}{>} & \langle \sigma_{11}, \sigma_{21} \rangle & \underset{1}{>} & \langle \sigma_{12}, \sigma_{22} \rangle \\ \langle \sigma_{12}, \sigma_{21} \rangle & \underset{2}{>} & \langle \sigma_{12}, \sigma_{22} \rangle & \underset{2}{>} & \langle \sigma_{11}, \sigma_{21} \rangle & \underset{2}{>} & \langle \sigma_{11}, \sigma_{22} \rangle \end{array} ,$$

sondern durch die Ergebnisse der Strategiekombinationen, etwa in Form von Gefängnisjahren definieren,

	B gesteht	B gesteht nicht
A gesteht	<10,10>	<0,20>
A gesteht nicht	<20,0>	<5,5>

so ergäben sich  $576^3 = 191\ 102\ 976$  logisch mögliche Kombinationen individueller Präferenzordnungen der drei Arten  $\langle \langle R_{w_A}, R_{m_A}, R_A \rangle, \langle R_{w_B}, R_{m_B}, R_B \rangle \rangle$ .

Von einem Großteil dieser Präferenzstrukturen ist anzunehmen, daß sie aufgrund psychologischer Sachverhalte empirisch nicht vorkommen. Die oben diskutierten A-Präferenzen z.B. gehören als  $R_w$ -Präferenzen mit Sicherheit, als  $R_{res}$ -Präferenzen vermutlich und als  $R_m$ -Präferenzen vermutlich nicht zu diesen empirisch auszuschließenden Präferenzordnungen. Die Untersuchung der restlichen Präferenzstrukturen würde jedoch ausreichen, um uns für die nächsten Jahre zu beschäftigen. Deshalb beschränken wir uns hier zunächst auf einige interessante Spezialfälle zur Konkretisierung.

Im vorangegangenen Abschnitt haben wir den *Senschen* Ansatz einer normativen Metapräferenztheorie behandelt. Definieren wir das Gefangenendilemma durch die obige Matrix, so kann sich eine normative Metapräferenztheorie grundsätzlich auf alle drei Arten individueller Präferenzordnungen beziehen. In allen drei Fällen kann eine solche normative Metapräferenztheorie eine Ethik erster Ordnung sein<sup>1</sup>, nämlich dann, wenn sie ihre moralische Ordnung nicht von gegebenen Präferenzen (Wünschen, Bedürfnissen) abhängig macht.

- (1) Bezieht sich eine normative Metapräferenztheorie auf die Präferenzrelationen persönlichen Wohlergehens  $\langle R_{w_i} \rangle^2$ , so werden bestimmte psychologische Disposi-

---

<sup>1</sup> vgl. §11

<sup>2</sup> vgl. §9

tionen, die das persönliche Wohlbefinden der Personen in Abhängigkeit von gesellschaftlichen Zuständen, Güterverteilungen und damit auch in Abhängigkeit von kollektiven Entscheidungen bestimmen, für moralisch mehr oder weniger wünschenswert erklärt. Ohne Zweifel gehören viele alltägliche moralische Diskussionen heute, aber auch in der Vergangenheit - man denke nur an den alten Cato - zu dieser Art normativer Metapräferenztheorie (als Ethik erster Ordnung). Dieser Bereich normativ-ethischer Theorienbildung steht auch einem Vertreter des psychologischen Egoismus offen (vermutlich ist das der einzige Bereich).

- (2) Bezieht sich die normative Metapräferenztheorie (als Ethik erster Ordnung) auf die individuellen moralischen Präferenzen  $\langle Rm_i \rangle$ , so werden bestimmt subjektive moralische Überzeugungen, die sich in entsprechende moralische Präferenzen über die Menge der kollektiven Entscheidungen fassen lassen, in eine moralische Rangordnung gebracht.
- (3) Bezieht sich die normative Metapräferenztheorie auf  $\langle Rres_i \rangle$ , so beurteilt sie die overall-Präferenzen, die sich in das revealed-preference -Konzept der ökonomischen Theorie einfügen und damit eine enge Verknüpfung zum tatsächlichen Entscheidungsverhalten haben.<sup>1</sup>

Die Alternative zu einer Metapräferenztheorie als normativer Ethik erster Ordnung ist eine Metapräferenztheorie, die Relationen individueller resultierender (und u.U. zusätzlich subjektiv-moralischer) Präferenzen aufgrund der Interessenstruktur der jeweiligen Interaktionssituation in eine moralische Rangordnung bringt, also eine Ethik zweiter Ordnung.<sup>2</sup>

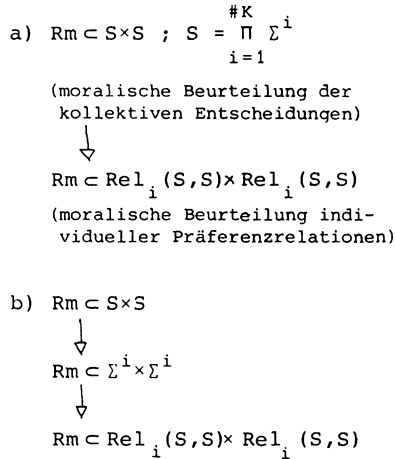
---

<sup>1</sup> vgl. §9

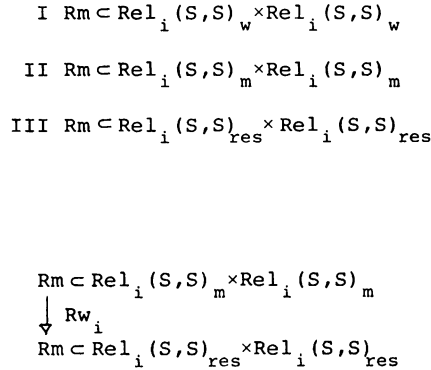
<sup>2</sup> vgl. §11

Beuründungsstrukturen

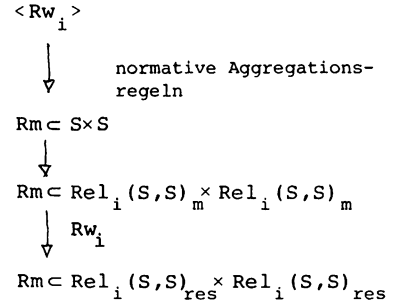
A) Sens 'quasiordering'



B) Normative Metapreferenztheorie als Ethik erster Ordnung



C) Normative Metapreferenztheorie als Ethik zweiter Ordnung



Wir wollen uns zunächst ausschließlich mit der Zuordnung von resultierenden individuellen Präferenzordnungen zu Strukturen persönlicher Wohlfahrtspräferenzen befassen. Als Beispiel dient uns erneut das Gefangenendilemma, das jetzt durch die Struktur der persönlichen Wohlfahrtspräferenzen konstituiert sei. Wir gehen dabei vorerst von einer äußerst schwachen Informationsbasis aus: Die Funktionen persönlicher Wohlfahrtsbewertung seien ordinal und interpersonell nicht vergleichbar. Anschließend erweitern wir diese Informationsbasis und diskutieren die Interaktionssituation bei vollständiger kardinaler interpersoneller Vergleichbarkeit der individuellen Wohlfahrtsfunktionen.<sup>1</sup>

Wenn wir Symmetrie voraussetzen, reduziert sich die Anzahl der logisch möglichen Kombinationen  $\langle Rres_1, Rres_2 \rangle$  von 576 auf 24. Erfüllen die resultierenden Präferenzen der Beteiligten die Rationalitätsbedingung  $\forall j \in K: [(\forall i \in K: \langle \xi, \xi' \rangle \in R w_i) \Rightarrow \langle \xi, \xi' \rangle \in R res_j]$ , so reduziert sich die Zahl weiter von 24 auf 12.

Diese Bedingung ist ein intuitiv sehr plausibles interpräferenzielles Rationalitätskriterium für die Beziehung von Wohlfahrtspräferenzen und resultierenden Präferenzen. In einer interpersonellen Interpretation kann diese Bedingung als normatives Kriterium gefordert werden. In einer intrapersonellen Interpretation kann sie sogar als eine vernünftige empirische Annahme gelten. "Interpersonelle Interpretation" bedeutet: Wenn die Personen eines Kollektivs Interessen haben, die wir durch  $\langle R w_i \rangle$  darstellen, dann sollten sich die resultierenden individuellen Präferenzen einer Person aus dieser Gruppe an dieses Kriterium halten. "Intrapersonelle Interpretation" bedeutet: Wenn einer Person  $i$  seine eigenen und die Interessen der anderen bekannt sind,

---

<sup>1</sup> Vgl. §6

dann sind dessen resultierende Präferenzen in Übereinstimmung mit diesem Kriterium.

Wir wollen dieses interpräferenzuelle Rationalitätskriterium als Paretoinklusivität der individuellen resultierenden Präferenzen bzgl. individueller Wohlfahrtspräferenzen" - kurz PIRPIW - nennen. Es ist wichtig zu sehen, daß die übliche Interpretation sozialwahltheoretischer Aggregationsregeln eine weit stärkere Voraussetzung hat, nämlich  $\forall i \in K: \langle \xi, \xi' \rangle \in R_{w_i} \Leftrightarrow \langle \xi, \xi' \rangle \in R_{res_i}$ , ohne diese je explizit zu machen. PIRPIW fordert, daß  $\langle \xi, \xi' \rangle \in R_{w_i} \Rightarrow \langle \xi, \xi' \rangle \in R_{res_i}$  gilt, wenn für alle anderen ebenfalls  $\xi$  persönlich günstiger als  $\xi'$  ist:  
 $\forall j \in K \setminus \{i\}: \langle \xi, \xi' \rangle \in R_{w_j}$ .

Diese Regel könnte man daher auch so formulieren:  
 Folge deinen eigenen Interessen, wenn keine fremden Interessen entgegenstehen. Oder: Maximiere getrost dein eigenes Wohlergehen, wenn das nicht das Wohlergehen anderer Personen beeinträchtigt.

Führen wir einige Fälle auf:

PD	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	$\langle 2, 2 \rangle$	$\langle 4, 1 \rangle$
$\sigma_{12}$	$\langle 1, 4 \rangle$	$\langle 3, 3 \rangle$

Um die Darstellung übersichtlicher zu machen, kennzeichnen wir die Konsequenzen bzw. die kollektiven Entscheidungen durch die entsprechenden Wohlfahrtsniveaus der Individuen.

- (1) Die Personen richten ihre resultierenden Präferenzen ausschließlich an ihrem persönlichen Wohlergehen aus. Solche resultierenden Präferenzen könnte man egoistische Präferenzen nennen. Sie hätten folgende Gestalt:  
 Für A:  $\langle 4, 1 \rangle \underset{1}{>} \langle 3, 3 \rangle \underset{1}{>} \langle 2, 2 \rangle \underset{1}{>} \langle 1, 4 \rangle$   
 Für B:  $\langle 1, 4 \rangle \underset{2}{>} \langle 3, 3 \rangle \underset{2}{>} \langle 2, 2 \rangle \underset{2}{>} \langle 4, 1 \rangle$

Diese Präferenzen sind pareto-inklusiv bzgl. der Wohlfahrtsbewertungen der Situation (PIRPIW).

'Tragik' der Egoisten (und Inhalt des Gefangenendilemmas) ist es, daß sie mit diesen Präferenzen bei  $\langle 2, 2 \rangle$  landen und daher ihren eigenen Interessen

gerade dadurch schaden, daß sie diese zum einzigen Kriterium ihrer (resultierenden) Präferenzen machen.

- (2) Die resultierenden Präferenzen sind geprägt von einem starken Gerechtigkeitsempfinden. Die von den Beteiligten an die gleiche Stelle ihrer Wohlfahrtspräferenzordnung gesetzten Zustände werden von den resultierenden Präferenzen vorgezogen, soweit das nicht mit PIRPIW in Konflikt kommt: Zustände, die bezüglich der Wohlfahrtspräferenzen pareto-besser sind, werden resultierend vorgezogen.

Für vier Zustände, in denen drei Personen folgende Wohlfahrtsniveaus erreichen  $\langle 2,2,2 \rangle$ ,  $\langle 3,3,3 \rangle$ ,  $\langle 9,9,1 \rangle$ ,  $\langle 4,4,2 \rangle$  ergäbe sich die resultierende Präferenzrelation:  $\langle 3,3,3 \rangle > \langle 4,4,2 \rangle > \langle 2,2,2 \rangle > \langle 9,9,1 \rangle$ .

Im Gefangenendilemma:

Für A:  $\langle 3,3 \rangle > \langle 2,2 \rangle > \langle 4,1 \rangle > \langle 1,4 \rangle$

Für B:  $\langle 3,3 \rangle \begin{matrix} 1 \\ > \\ 2 \end{matrix} \langle 2,2 \rangle \begin{matrix} 1 \\ > \\ 2 \end{matrix} \langle 1,4 \rangle \begin{matrix} 1 \\ > \\ 2 \end{matrix} \langle 4,1 \rangle$

Bei diesen Präferenzen wird angenommen, daß die Personen, falls sie unter moralischen Gesichtspunkten keinen Unterschied zwischen zwei Zuständen machen (wie zwischen  $\langle 4,1 \rangle$  und  $\langle 1,4 \rangle$ ), den Zustand resultierend vorziehen, in dem es ihnen persönlich besser geht. Diese resultierenden Präferenzen garantieren allerdings nicht kollektive Rationalität, denn bei diesem starken Gerechtigkeitsempfinden entscheidet sich A für  $\sigma_{11}$ , falls er annimmt, daß B sich für  $\sigma_{21}$  entscheidet (und das ist nicht ausgeschlossen, da  $\sigma_{22}$  für B nicht dominant ist).  $\langle \sigma_{11}, \sigma_{21} \rangle$  ist bezüglich der resultierenden Präferenzen genauso wie bezüglich der Wohlfahrtspräferenzen ein Gleichgewichtspunkt der Interaktionssituation, allerdings kommt in der durch die resultierenden Präferenzen bestimmte Interaktionssituation noch als zweiter Gleichgewichtspunkt die kollektiv rationale Entscheidung  $\langle \sigma_{12}, \sigma_{22} \rangle$  hinzu.

- (3) Auch wenn sich der Gerechtigkeitssinn mit Altruismus verbindet und sich folgende resultierenden Präferenzen ergeben:

$$A: \langle 3,3 \rangle \underset{1}{>} \langle 2,2 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle$$

$$B: \langle 3,3 \rangle \underset{2}{>} \langle 2,2 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 1,4 \rangle$$

ist die kollektiv rationale Entscheidung aus den gleichen Gründen wie unter (2) nicht garantiert.

- (4) Den Gegenpol zu diesen an 'Gerechtigkeit' orientierten Präferenzen bilden solche, die man als Ausdruck eines stark ausgeprägten Anti-Egalitarismus interpretieren könnte:

$$A: \langle 4,1 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 2,2 \rangle$$

$$B: \langle 1,4 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 2,2 \rangle$$

Diese anti-egalitären resultierenden Präferenzen haben (wie die egoistischen) einen Gleichgewichtspunkt in dominanten Strategien bei  $\langle \sigma_{11}, \sigma_{21} \rangle$ , garantieren also kollektive Irrationalität, obwohl die resultierenden Präferenzen pareto-inklusiv sind (PIRPIW).

- (5) Verdrängt jedoch der Altruismus den Gerechtigkeitssinn in der Gestalt folgender resultierender Präferenzen (A-Präferenzen)

$$A: \langle 3,3 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 2,2 \rangle$$

$$B: \langle 3,3 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 2,2 \rangle$$

so ist endlich kollektive Rationalität garantiert: Die kollektiv rationale Entscheidung ist Gleichgewichtspunkt in dominanten Strategien. Allerdings kommen einem gewisse Zweifel an der inneren Konsistenz dieser Präferenzrelationen, denn  $\langle 4,1 \rangle \underset{1}{>} \langle 2,2 \rangle$  ist weder altruistisch noch gerecht, sondern egoistisch, während  $\langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle$  extrem altruistisch ist. Bei  $\langle 3,3 \rangle \underset{1}{>} \langle 1,4 \rangle$  und  $\langle 3,3 \rangle \underset{2}{>} \langle 4,1 \rangle$  führt Egoismus zur Gerechtigkeit, oder man könnte mit der gleichen Berechtigung sagen, ist Gerechtigkeit im persönlichen Interesse.



- (6) Aber auch ein konsequenter (reiner) Altruismus garantiert kollektive Rationalität in Interaktionssituationen von der Art des Gefangenendilemmas.

A:  $\langle 1,4 \rangle \succ_1 \langle 3,3 \rangle \succ_1 \langle 2,2 \rangle \succ_1 \langle 4,1 \rangle$

B:  $\langle 4,1 \rangle \succ_2 \langle 3,3 \rangle \succ_2 \langle 2,2 \rangle \succ_2 \langle 1,4 \rangle$

Für diese rein altruistischen (resultierenden) Präferenzen ist die kollektiv rationale Lösung die einzige Gleichgewichtsentscheidung, und  $\sigma_{12}$  und  $\sigma_{22}$  ist jeweils die dominante individuelle Entscheidung.

- (7) A:  $\langle 3,3 \rangle \succ_1 \langle 1,4 \rangle \succ_1 \langle 2,2 \rangle \succ_1 \langle 4,1 \rangle$

B:  $\langle 3,3 \rangle \succ_2 \langle 4,1 \rangle \succ_2 \langle 2,2 \rangle \succ_2 \langle 1,4 \rangle$

Diese symmetrische Präferenzstruktur könnte bei kardinaler Interpretation der Wohlfahrtsbewertungen als Ausdruck eines abgeschwächten Altruismus interpretiert werden: Wenn die Erhöhung des Wohlfahrtsniveaus der anderen Person gleich hoch wie die Absenkung des eigenen Wohlfahrtsniveaus ist, wird die kollektive Entscheidung zugunsten des anderen vorgezogen. Ist die Veränderung des eigenen Wohlfahrtsniveaus jedoch doppelt so groß wie die des anderen, neigen sich die resultierenden Präferenzen zu den eigenen Gunsten. Diese Präferenzen haben einen Gleichgewichtspunkt in  $\langle \sigma_{12}, \sigma_{22} \rangle$ , garantieren also immer noch kollektive Rationalität.<sup>2</sup>

- (8) 'Masochistische' resultierende Präferenzen

A:  $\langle 1,4 \rangle \succ_1 \langle 2,2 \rangle \succ_1 \langle 3,3 \rangle \succ_1 \langle 4,1 \rangle$

B:  $\langle 4,1 \rangle \succ_2 \langle 2,2 \rangle \succ_2 \langle 3,3 \rangle \succ_2 \langle 1,4 \rangle$

haben eine besonders interessante Konsequenz.<sup>1</sup>

Selbsthaß garantiert zusammen mit (reinem und abgeschwächtem) Altruismus (6,7) und den etwas seltsam motivierten A-Präferenzen, die Sen "other regarding" nennt, und drei weiteren recht unplausiblen Präferenzstrukturen die kollektiv rationale Lösung.

<sup>1</sup> Diese Präferenzen entsprechen Baiers Self-hating-Präferenzen, vgl. Baier (1977), S.208

<sup>2</sup>  $\sigma_{12}$  und  $\sigma_{22}$  sind ja dominant

Kollektiv rational natürlich nur bezüglich der Wohlfahrtspräferenzen, also im engeren Sinne. Bezüglich der resultierenden Präferenzen selbst jedoch können die Handelnden mit dem Ergebnis nicht zufrieden sein: Obwohl sie den Zustand, in dem es ihnen persönlich am schlechtesten geht an die Spitze und den Zustand, in dem es ihnen am besten geht ans Ende ihrer (resultierenden) Präferenzordnungen gesetzt haben, geraten sie beide unausweichlich in einen Zustand, in dem es ihnen sehr gut geht:  $\langle 3,3 \rangle$ . Wir sagten etwas leichtfertig: Sie können mit dem Ergebnis nicht zufrieden sein. Aber andererseits ist es analytisch wahr, daß sie mit  $\langle 3,3 \rangle$  recht zufrieden sind, denn dieser Zustand rangiert in beiden Präferenzordnungen ihres persönlichen Wohlergehens an zweiter Stelle. Das Gefangenendilemma reproduziert sich gewissermaßen für Masochisten auf der Ebene der resultierenden Präferenzen. Man hat in diesem Fall tatsächlich Interpretationsprobleme: Kann jemand wollen, daß es ihm persönlich schlecht geht? Und ist er dann unglücklich, wenn es ihm gut geht, oder geht es ihm schlecht, wenn er glücklich ist? Oder will er nur, daß es ihm schlecht geht? Dann ist er zwar glücklich, wenn es ihm gut geht, aber glücklich will er eben nicht sein .... Das Problem liegt wohl darin, daß sich diese resultierenden Präferenzen ausschließlich durch die Berücksichtigung der eigenen Wohlfahrtspräferenzen ergeben. In den anderen Fällen war es - mehr oder weniger intuitiv überzeugend - möglich, sich die Differenz von individueller Wohlfahrts- und resultierenden Präferenzen durch die Berücksichtigung der Interessen des anderen Beteiligten plausibel zu machen, was hier nicht möglich ist.

- (9) Das Gegenstück ('sadistische' resultierende Präferenzen)

$$A: \langle 4, 1 \rangle \underset{1}{>} \langle 2, 2 \rangle \underset{1}{>} \langle 3, 3 \rangle \underset{1}{>} \langle 1, 4 \rangle$$

$$B: \langle 1, 4 \rangle \underset{2}{>} \langle 2, 2 \rangle \underset{2}{>} \langle 3, 3 \rangle \underset{2}{>} \langle 4, 1 \rangle$$

verstößt wie Masochismus gegen die interpräferenzelle Bedingung der Pareto-Inklusivität. Obwohl

$$\forall i \in \{1, 2\}: \pi_i(\langle \sigma_{12}, \sigma_{22} \rangle) > \pi_i(\langle \sigma_{11}, \sigma_{21} \rangle) \text{ ist}$$

$$U_i(\langle \sigma_{12}, \sigma_{22} \rangle) < U_i(\langle \sigma_{11}, \sigma_{21} \rangle) \text{ für } i \in \{1, 2\}.$$

Während jedoch Egoismus im Gefangenendilemma nicht erfolgreich war (die Personen erhielten jeweils weniger vom Erstrebten als wenn sie es nicht erstrebt hätten, vgl. (1,5,6,7,8)), ist das wechselseitige Bestreben, dem anderen zu schaden soweit erfolgreich, wie es die Symmetrie der Situation zuläßt:  $\sigma_{11}$  und  $\sigma_{21}$  sind die dominanten Strategien. Der Gleichgewichtspunkt  $\langle \sigma_{11}, \sigma_{21} \rangle$  ist kollektiv irrational. Nun ist man versucht, in diesem Fall die kollektive Irrationalität der nicht-kooperativen kollektiven Entscheidung zu bezweifeln: 'Immerhin haben sie ja erreicht, was sie wollten.' Dieses Argument übersieht jedoch, daß die nicht-kooperative Entscheidung auch für den 'Sadisten' schlechter ist als die kooperative, unbeschadet der Tatsache, daß es seinem Kompagnon in der nicht-kooperativen Entscheidung, wie erwünscht, ebenfalls schlecht geht. Andernfalls wären die Wohlfahrtsbewertungen kollektiver Entscheidungen in diesem Fall abzuändern, und wir hätten es nicht mehr mit einer Interaktionssituation vom Typ des Gefangenendilemmas zu tun.

- (10) Die resultierenden Präferenzen, die dem *Senschen* Assurance game entsprechen, haben folgende Struktur:

$$A: \langle 3, 3 \rangle \underset{1}{>} \langle 4, 1 \rangle \underset{1}{>} \langle 2, 2 \rangle \underset{1}{>} \langle 1, 4 \rangle$$

$$B: \langle 3, 3 \rangle \underset{2}{>} \langle 1, 4 \rangle \underset{2}{>} \langle 2, 2 \rangle \underset{2}{>} \langle 4, 1 \rangle$$

Die Personen mit V-Präferenzen sind dann bereit, sich kooperativ zu verhalten, wenn sie annehmen, daß sich die andere Person ebenfalls kooperativ verhält. Damit wird die kollektiv rationale Entscheidung  $\langle 3, 3 \rangle$  ein

Gleichgewichtspunkt der Interaktionssituation. Allerdings ist die kollektiv irrationale Entscheidung  $\langle 2,2 \rangle$  ebenfalls ein Gleichgewichtspunkt. Kein Beteiligter hat eine dominante individuelle Strategie. Hier ist wechselseitiges Vertrauen Bedingung kollektiver Rationalität.

Von den zwölf logisch möglichen PIRPIW-rationalen symmetrischen resultierenden Präferenzstrukturen in der  $2 \times 2$ -Interaktionssituation des PD-Typs (in persönlichen Wohlfahrtspräferenzen) haben wir bisher acht auf ihre Beziehung zur kollektiven Rationalität untersucht. Sadismus und Masochismus sind nicht PIRPIW-rational. Die restlichen vier PIRPIW-rationalen Präferenzstrukturen sind schwer zu interpretieren:

$$(11) \text{ A: } \langle 3,3 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 2,2 \rangle$$

$$\text{B: } \langle 3,3 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 2,2 \rangle$$

Das sind fast utilitaristische Präferenzen, sie garantieren kollektive Rationalität (Gleichgewichtspunkt in dominanten Strategien  $\langle 3,3 \rangle$ ).

$$(12) \text{ A: } \langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 2,2 \rangle$$

$$\text{B: } \langle 4,1 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 2,2 \rangle$$

Hier gibt es zwei Gleichgewichtspunkte:  $\langle \sigma_{11}, \sigma_{21} \rangle$  und  $\langle \sigma_{12}, \sigma_{22} \rangle$ .

Die letzten beiden Präferenzstrukturen garantieren wiederum kollektive Rationalität i.e.S.:

$$(13) \text{ A: } \langle 4,1 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 2,2 \rangle$$

$$\text{B: } \langle 1,4 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 2,2 \rangle$$

$$(14) \text{ A: } \langle 1,4 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 2,2 \rangle$$

$$\text{B: } \langle 4,1 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 2,2 \rangle$$

Die meisten der PIRPIW-irrationalen Präferenzrelationen werden wohl empirisch kaum vorkommen. Dennoch ist es interessant, daß von den zwölf logisch möglichen PIRPIW-irrationalen symmetrischen Präferenzstrukturen immerhin zwei kollektive Rationalität garantieren. M.a.W.: Selbst

wenn beide Personen eine (für beide) schlechtere Situation einer (für beide) besseren Situation vorziehen, kann individuell rationales Verhalten zum besseren Ergebnis führen. Das gilt zunächst für die schon genannte Präferenzstruktur (8) (Masochismus) und für folgende PIRPIW-irrationale symmetrische Präferenzstruktur:

$$(15) \text{ A: } \langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 2,2 \rangle \underset{1}{>} \langle 3,3 \rangle$$

$$\text{ B: } \langle 4,1 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 2,2 \rangle \underset{2}{>} \langle 3,3 \rangle$$

Von den PIRPIW-irrationalen Präferenzstrukturen ermöglichen immerhin sechs kollektive Rationalität, nämlich

$$(16) \text{ A: } \langle 2,2 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 1,4 \rangle$$

$$\text{ B: } \langle 2,2 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 4,1 \rangle$$

$$(17) \text{ A: } \langle 2,2 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle$$

$$\text{ B: } \langle 2,2 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 1,4 \rangle$$

$$(18) \text{ A: } \langle 2,2 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 1,4 \rangle$$

$$\text{ B: } \langle 2,2 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 4,1 \rangle$$

$$(19) \text{ A: } \langle 2,2 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 3,3 \rangle \underset{1}{>} \langle 4,1 \rangle$$

$$\text{ B: } \langle 2,2 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 3,3 \rangle \underset{2}{>} \langle 1,4 \rangle$$

$$(20) \text{ A: } \langle 4,1 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 2,2 \rangle \underset{1}{>} \langle 3,3 \rangle$$

$$\text{ B: } \langle 1,4 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 2,2 \rangle \underset{2}{>} \langle 3,3 \rangle$$

$$(21) \text{ A: } \langle 1,4 \rangle \underset{1}{>} \langle 2,2 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 3,3 \rangle$$

$$\text{ B: } \langle 4,1 \rangle \underset{2}{>} \langle 2,2 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 3,3 \rangle$$

Vier der PIRPIW-irrationalen Präferenzstrukturen garantieren kollektive Irrationalität. Außer sadistischen Präferenzen (9) noch:

$$(22) \text{ A: } \langle 4,1 \rangle \underset{1}{>} \langle 2,2 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 3,3 \rangle$$

$$\text{ B: } \langle 1,4 \rangle \underset{2}{>} \langle 2,2 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 3,3 \rangle$$

$$(23) \text{ A: } \langle 2,2 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 3,3 \rangle$$

$$\text{ B: } \langle 2,2 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 3,3 \rangle$$

$$(24) \text{ A: } \langle 2,2 \rangle \underset{1}{>} \langle 1,4 \rangle \underset{1}{>} \langle 4,1 \rangle \underset{1}{>} \langle 3,3 \rangle$$

$$\text{ B: } \langle 2,2 \rangle \underset{2}{>} \langle 4,1 \rangle \underset{2}{>} \langle 1,4 \rangle \underset{2}{>} \langle 3,3 \rangle$$

	PIRPIW -		kollektive Rationalität			intuitiv plausibel	
	rational	irrational	garan- tiert	mög- lich	nicht möglich	ja	nei
1. Egoismus	x				x	x	
2. Reine Gerechtigkeit	x			x		x	
3. Altruistische Gerechtigkeit	x			x		x	
4. Antiegalitarismus	x				x	x	
5. A-Präferenzen	x		x				x
6. Reiner Altruismus	x		x			x	
7. Abgeschwächter Altruismus	x		x			x	
8. Masochismus		x	x			x	
9. Sadismus		x			x	x	
10. V-Präferenzen	x			x		x	
11.	x		x				x
12.	x			x			x
13.	x		x				x
14.	x		x				x
15.		x	x				x
16.		x		x			x
17.		x		x			x
18.		x		x			x
19.		x		x			x
20.		x		x			x
21.		x		x			x
22.		x			x		x
23.		x			x		x
24.		x			x		x

Gegenstand der normativen Metapräferenztheorie (zweiter Ordnung) ist die moralische "besser"-Relation auf der Menge der individuellen resultierenden Präferenzen. Für Interaktionssituationen vom PD-Typ mit bloßer ordinaler (interpersoneller) Vergleichbarkeit der Wohlfahrtspräferenzen ergeben sich also drei Indifferenzklassen. Zur ersten Indifferenz ( $I_1$ ) gehören diejenigen Präferenzrelationen, die kollektive Rationalität garantieren; zur zweiten ( $I_2$ ) diejenigen, die kollektive Rationalität ermöglichen und zur dritten ( $I_3$ ) diejenigen, die kollektive Irrationalität garantieren.

Innerhalb dieser drei Klassen gibt es jeweils PIRPIW-rationale und PIRPIW-irrationale Präferenzstrukturen.

Hier sind drei Ergebnisse (individuell) rationalen Handelns bezüglich der Strukturen resultierender Präferenzen möglich, und eine 'ergebnisorientierte' normative Metapräferenztheorie (wie sie Sen offenbar im Blick hatte) kann innerhalb dieser drei Indifferenzklassen keine moralischen Unterschiede feststellen.

Zwei individuelle resultierende Präferenzrelationen sind moralisch gleich gut genau dann, wenn sie zur gleichen Indifferenzklasse gehören.

Alle individuellen resultierenden Präferenzrelationen der ersten Indifferenzklasse sind besser als jede beliebige resultierende Präferenzrelation der zweiten oder der dritten Indifferenzklasse, und die der zweiten Indifferenzklasse sind moralisch besser als die der dritten.

Insoweit könnte man versucht sein zu meinen, das Sen'sche Programm einer moralischen Ordnung individueller Präferenzordnungen sei zumindest in dieser Interpretation einer normativen Metapräferenztheorie zweiter Ordnung durchführbar: Wir haben eine Methode strikt ergebnisorientierter moralischer Beurteilung individueller Präferenzrelationen.

Diese Euphorie vergeht jedoch, sobald man die so entstandene moralische Ordnung resultierender Präferenz-

strukturen genauer betrachtet.

An erster Stelle rangieren nicht nur so ehrenwerte Präferenzen wie (6) und (7), sondern auch zwei PIRPIW-irrationale Präferenzstrukturen (8) und (15), nur mit Selbsthaß zu erklärende Präferenzen (8), und eine ganze Reihe recht unplausibler Präferenzstrukturen wie (11), (13) und (14). Einen mittleren Rang nehmen einige Präferenzstrukturen ein, die man als moralischer empfindet als manche andere, die von der moralischen Metapräferenzrelation am höchsten bewertet werden, wie (2), (3); daneben solche Präferenzen, die die eigenen Interessen gehörig berücksichtigen (10), und dann besonders viele Präferenzstrukturen (sieben von zehn insgesamt), die man guten Gewissens niemandem empfehlen kann. Letztere verstoßen einmal gegen intuitive Desiderata der Bewertungs-rationalität, die wir jedoch nicht explizit zu machen versucht haben, und zum zweiten sind von diesen sieben unplausiblen Präferenzstrukturen alle bis auf eine (12) PIRPIW-irrational.

In der von der moralischen Metapräferenzrelation am schlechtesten bewerteten Gruppe von Präferenzstrukturen rangieren erwartungsgemäß die beiden PIRPIW-rationalen Strukturen resultierender Präferenzen, Egoismus und Anti-Egalitarismus, der PIRPIW-irrationale Sadismus und zusätzlich drei weitere völlig unplausible PIRPIW-irrationale Präferenzstrukturen (22), (23), (24).

Da Sen sich auf die Analyse von Präferenzstrukturen beschränkt hatte, konnte das Konzept einer ergebnisorientierten moralischen Metapräferenzrelation dort noch einigermaßen überzeugend erscheinen (wenn es auch in der *Senschen* Darstellung Interpretationsprobleme aufwarf, die wir im letzten Abschnitt diskutiert haben), schon Baiers Erweiterung der Untersuchung um weitere drei Strukturen<sup>1</sup> zeigt einige Defizite dieses Ansatzes. Betrachtet

---

<sup>1</sup> Vgl. Baier (1977)



man jedoch die Einzelergebnisse im Zusammenhang aller logisch möglichen symmetrischen Strukturen resultierender Präferenzen, bleibt kaum noch Hoffnung, auf diese Weise kollektive Rationalität als Ergebnis individuell rationalen Verhaltens bezüglich moralisch gebotener resultierender Präferenzen sicherstellen zu können.

Der Konflikt von individueller Rationalität und kollektiver Rationalität bleibt also weiterhin bestehen. Der Fortschritt bestand bisher in erster Linie darin, mit Hilfe entsprechender Begriffsdifferenzierungen wenigstens das Problem adäquat analysieren zu können.

Nun haben wir von den in §9 eingeführten Präferenzbegriffen nur zum Teil Gebrauch gemacht. Aber auch intrinsische Präferenzen und (individuelle) Präferenzen höherer Ordnung bringen nur eine weitere Verfeinerung der Analyse mit sich, auf die wir jedoch nicht verzichten wollen.

Schlagen wir den von Baier vorgezeichneten Weg<sup>1</sup> ein: Die Interaktionsstruktur selbst ist definiert durch die individuellen intrinsischen Präferenzen erster Ordnung: Beide Betroffenen ziehen es vor, weniger Gefängnisjahre absitzen zu müssen als mehr. Vermutlich ziehen sie es deshalb vor, weil sie meinen, daß es in ihrem persönlichen Interesse ist, so wenig Jahre wie möglich hinter Schloß und Riegel zu verbringen. Insofern ergibt sich daraus allein noch kein wesentlicher Unterschied gegenüber zuvor, als wir die Interaktionssituation selbst durch die entsprechenden individuellen Wohlfahrtspräferenzen definiert haben.

Wir hatten die Präferenzstrukturen (1) - (24) als resultierende Präferenzen interpretiert, jetzt seien sie verstanden als Darstellung von Präferenzen zweiter Ordnung.

---

<sup>1</sup> Baier bietet tatsächlich eine 'Lösung' an, die jedoch zum Bereich der Rechtfertigung etablierter Konfliktregelungen gehört, den wir hier aussparen müssen.

Diese Präferenzen bringen zum Ausdruck, in welchem Ausmaß die Person wünscht, daß ihre intrinsischen Präferenzen erster Ordnung und die anderer Personen befriedigt werden. Ein Beispiel:

Reine Gerechtigkeitspräferenzen (zweiter Ordnung) (2) bevorzugen an erster Stelle, daß diejenige kollektive Entscheidung getroffen wird, die die eigenen intrinsischen Präferenzen erster Ordnung genauso wie (nach eigener Einschätzung) die intrinsischen Präferenzen erster Ordnung der anderen Person an die zweite Stelle setzen.

Reine Gerechtigkeitspräferenzen (zweiter Ordnung) bevorzugen an zweiter Stelle, daß diejenige kollektive Entscheidung getroffen wird, die die eigenen intrinsischen Präferenzen erster Ordnung, wie die des anderen Gefangenen an die dritte Stelle setzen.

Dann erst folgt diejenige kollektive Entscheidung, die den eigenen intrinsischen Präferenzen erster Ordnung am liebsten, aber den intrinsischen Präferenzen erster Ordnung des anderen am allerwenigsten lieb ist.

Und an letzter Stelle rangiert in dieser Präferenzordnung zweiter Ordnung die Situation, die die eigenen intrinsischen Präferenzen erster Ordnung an letzte und diejenigen der anderen Person an erste Stelle setzen:

$$\begin{array}{cccc} \langle 3,3 \rangle & \rangle & \langle 2,2 \rangle & \rangle & \langle 4,1 \rangle & \rangle & \langle 1,4 \rangle \\ & & 1 & & 1 & & 1 \\ \langle 3,3 \rangle & \rangle & \langle 2,2 \rangle & \rangle & \langle 1,4 \rangle & \rangle & \langle 4,1 \rangle \\ & & 2 & & 2 & & 2 \end{array}$$

Egoistische Präferenzen zweiter Ordnung richten sich dagegen vollständig nach den eigenen intrinsischen Präferenzen erster Ordnung, masochistische Präferenzen zweiter Ordnung sind den eigenen intrinsischen Präferenzen erster Ordnung genau entgegengesetzt, sadistische Präferenzen zweiter Ordnung sind den intrinsischen Präferenzen erster Ordnung des anderen genau entgegengesetzt etc..

Distributive Präferenzen zweiter Ordnung beziehen sich darauf, welche intrinsische Präferenzen erster Ordnung von welchen Personen in welchem Ausmaß befriedigt werden sollen. Es gibt auch nicht-distributive Präferenzen zweiter Ordnung, z.B. solche, die sich ausschließlich auf meine eigenen Präferenzen erster Ordnung beziehen: Ich mag es als Präferenz zweiter Ordnung wünschen, daß meine intrinsischen Präferenzen für Alkohol weniger und meine intrinsischen Präferenzen für sportliche Betätigung mehr befriedigt werden (und z.B. meine Abende entsprechend

planen ...). Ein unparteiischer Beobachter hat distributive Präferenzen zweiter Ordnung über intrinsische Präferenzen erster Ordnung anderer Personen; seine eigenen intrinsischen Präferenzen erster Ordnung sind von seinen distributiven Präferenzen zweiter Ordnung nicht betroffen.

Nun können Präferenzen zweiter Ordnung intrinsisch oder abgeleitet sein ( abgeleitete Präferenzen sind nicht-intrinsische Präferenzen). Das was wir sadistische Präferenzen zweiter Ordnung genannt haben, muß nicht unbedingt mit Sadismus etwas zu tun haben: Wenn ich z.B. der Ansicht bin, daß eine Bestrafung von Verbrechern notwendig ist, oder daß es der betreffenden Person nur gut tut, einige Jahre im Gefängnis zu sitzen, so habe ich keine intrinsische Präferenz zweiter Ordnung dafür, daß sie möglichst lange im Gefängnis sitzt, sondern nur eine abgeleitete. Ebenso halten wir 'masochistische' Präferenzen (8) als intrinsische Präferenzen zweiter Ordnung i.a. für irrational oder pathologisch, während solche Präferenzen als abgeleitete Präferenzen zweiter Ordnung völlig vernünftig sein können.

Die Uminterpretation der oben diskutierten Präferenzstrukturen zeigt hier also zum ersten Mal Wirkung: Als Präferenzen zweiter Ordnung können sie nicht mehr in der gleichen Weise wie zuvor gedeutet werden, denn die Strukturen (11)-(24) sind nur als intrinsische Präferenzen (zweiter Ordnung) unplausibel, als abgeleitete Präferenzen keineswegs.

Nehmen wir an, die Personen hätten intrinsische Präferenzen erster Ordnung, wie sie sich in der Interaktionsstruktur des Gefangenendilemmas zeigen und distributive intrinsische Präferenzen zweiter Ordnung gemäß der Präferenzstruktur der 'reinen Gerechtigkeit' (2). Nehmen wir weiter an, daß die distributiven intrinsischen Präferenzen zweiter Ordnung so stark sind, daß sie die

intrinsischen erster Ordnung dominieren, d.h. die resultierenden Präferenzen allein bestimmen. Die Personen sind also der Ansicht, daß die kollektiv rationale Entscheidung auch die individuell beste sei. Selbst in diesem Fall ist jedoch nicht garantiert, daß individuell rationales Verhalten zur kollektiv rationalen Entscheidung führt.

Auch diese Uminterpretation kann also unsere Probleme nicht lösen: Wir suchen nach Präferenzen (welcher Art auch immer), bezüglich derer individuell rationales Verhalten kollektive Rationalität (als sozialethische Minimalbedingung) garantiert. Bei diesem Bemühen ist es uns zwar gelungen, die Analyse der Interaktionssituation so zu verfeinern, daß sich die moralphilosophische Problematik deutlich zeigt<sup>1</sup>, aber dennoch kann das Ergebnis nicht befriedigen, denn wir haben feststellen müssen:

1. Eine ergebnisorientierte Metapräferenztheorie führt zu inadäquaten moralischen Beurteilungen individueller Präferenzen.
2. Kollektive Rationalität kann durch individuell rationales (präferenzenorientiertes) Handeln nicht überzeugend sichergestellt werden.

Die Konsequenz muß m.E. sein, den rein teleologischen Ansatz in der Sozialethik aufzugeben und eine gemischt deontologisch-teleologische Theorie zu entwickeln:

Auch eine subjektivistische Theorie der Sozialethik kann sich nicht darauf beschränken, moralisch gebotene resultierende Präferenzen zu bestimmen und es dann den Kriterien individuell rationalen Verhaltens in Interaktionssituationen überlassen, die individuelle moralische Entscheidung als diejenige zu bestimmen, die bezüglich der moralisch gebotenen resultierenden Präferenzen individuell rational ist. Es gibt zwei Gegenstände moralischer Beurteilung:

---

<sup>1</sup> und wir haben auch zwei symmetrische Präferenzstrukturen entdeckt, die kollektive Rationalität aufgrund individueller Rationalität garantieren, ohne selbst (intrinsisch interpretiert) irrational zu sein.

Entscheidungen und Präferenzen. Und die normative Theorie der einen Ebene ist, so scheint es, nicht reduzierbar auf die normative Theorie der anderen: Kollektive Rationalität ist neben individueller Rationalität ein eigenständiges Entscheidungsprinzip.

Wären die Schlußfolgerungen für die Unzulänglichkeiten eines teleologischen Ansatzes im ordinalen Rahmen wohlbegründet, sollten sie a fortiori bei einer Verstärkung der Informationsbasis, etwa zu voller kardinaler Vergleichbarkeit, gelten. Dennoch werfen wir einen Blick auf die kardinale Analyse unserer Ausgangssituation - immerhin könnten sich ja dabei ganz neue Aspekte ergeben, die die Tragweite der bisherigen Ergebnisse einschränken.

Rescher hat vorgeschlagen<sup>1</sup>, die wechselseitigen Mitgeföhle durch eine Art Nutzeninterdependenz darzustellen. Wir machen uns dieses Modell hier zunutze, ohne diese Interpretation zu übernehmen. Die Nutzeninterdependenz könnte auch zur bloßen (teleologischen) Rekonstruktion moralischer Normen dienen.

$p_{ij}$  ist der Prozentsatz, den  $i$  von der Wohlfahrt der Person  $j$  'internalisiert'.

$t_{ij}$  ist der Wohlfahrtstransfer, den  $i$  von  $j$  erhält.

$N$  sei die Norm, die als Maßstab für das Wohlergehen einer Person gilt.

Damit wird die resultierende subjektive Bewertungsfunktion folgendermaßen bestimmt:

$$u_i = p_0 \pi_i + \sum_{\substack{k \in K \\ k \neq i}} t_{ik} = p_0 \pi_i + \sum_{\substack{k \in K \\ k \neq i}} (\pi_k - N) \cdot p_{ik}$$

Nehmen wir an,  $p_{ij}$  wäre für beide Personen 0,5 und  $p_0 = 0$ .

Dann ist der Interaktionssituation in den subjektiven Wohlfahrtpräferenzen

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	(2,2)	(4,1)
$\sigma_{12}$	(1,4)	(3,3)

---

<sup>1</sup> vgl. N.Rescher, Unselfishness. The role of the vicarious affects in moral philosophy and social theory, Pittsburgh 1975, bes. Kap.2

folgende Interaktionssituation  
in den resultierenden Bewertungen  
zugeordnet ( mit  $N = 2,5$  )

(1,75/1,75)	(3,25/1,75)
(1,75/3,25)	(3,35/3,25)

Gerade für diese Transferrate geht das Gefangenendilemma in den resultierenden Präferenzen in eine Interaktionssituation über, in der individuelle und kollektive Rationalität nicht mehr auseinanderfallen.

Tatsächlich läßt sich zeigen, daß es für beliebige kardinal repräsentierte Interaktionssituationen vom PD-Typ eine Transferrate  $p_{ij}^*$  gibt, so daß für alle Transferraten, die größer als  $p_{ij}^*$  sind, gilt, daß das 'Dilemma behoben' ist. Auf der Basis kardinaler Vergleichbarkeit scheint also wider Erwarten doch eine 'Versöhnung' individueller und kollektiver Rationalität möglich. Bei genauerer Betrachtung entpuppt sich dies jedoch rasch als ein Irrtum. Die Mindestgrenzen der Transferraten sind für jede PD-Situation verschieden und es gibt keine Obergrenze für diese Mindestgrößen. Man muß unbegrenzt große Transferraten zulassen, um jeweils kooperatives Verhalten in PD-Situationen zur dominanten Entscheidung bezüglich der resultierenden Bewertungsfunktion zu machen.

Nun könnte man sich auf den Standpunkt stellen, eine Transferrate bestimmter Größe sei hinreichend für die teleologische Rekonstruktion, die unkooperatives Verhalten in allen übrigen Fällen, in denen diese Transferrate nicht ausreicht, möglicherweise moralisch zulässig sei.

Dagegen spricht jedoch folgende Überlegung. Nehmen wir an, die moralisch gebotene Transferrate sei einheitlich oder zumindest nicht von der Präferenz- bzw. Bewertungsstruktur der Interaktionssituation abhängig. PD\* sei eine PD-Situation, bei der kooperatives Verhalten moralisch geboten ist und PD\* erfordere die höchste Transferrate  $t^*$  von allen Situationen, in

denen kooperatives Verhalten geboten ist. Dann kann  $t^*$  schon deshalb nicht eine angemessene Transferrate sein, weil in diesem Falle aus Nicht-PD-Situationen nach dem Transfer PD-Situationen würden, d.h. Situationen, in denen kollektive Rationalität und individuelle Rationalität übereinstimmen, verwandelten sich in solche, für die die kollektiv irrationale Entscheidung individuell rational ist.<sup>1</sup>

Beispiel: Sei  $p_0 = 0$ ;  $p_{12} = 1$ ;  $N=0$ ;

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	-100, -100	-101, 6
$\sigma_{12}$	6, -101	5, 5

geht über in die resultierenden Bewertungen:

	$\sigma_{21}$	$\sigma_{22}$
$\sigma_{11}$	-100, -100	6, -101
$\sigma_{12}$	-101, 6	5, 5

für diese resultierenden Bewertungen ist jedoch die kollektiv irrationale Entscheidung Konsequenz individueller Rationalität. Eine besonders starke Berücksichtigung der Interessen des anderen verbunden mit völliger Selbstlosigkeit verwandelt

---

<sup>1</sup> Rescher war der Ansicht, daß "in every case of the sort at issue where the (uniform) operation of the positive vicarious affects produce a prisoner's dilemma situation, the original situation from which a prisoner's dilemma is to arise after internalization must itself already represent a prisoner's dilemma" (S.51), worin er sich jedoch leider irrt.

Rescher hat offensichtlich nicht bemerkt, daß das von ihm selbst angeführte 'saint's dilemma' eine Widerlegung dieser These ist. Weymark (1978) hat zwar entdeckt, daß Reschers These falsch ist, aber nicht, daß das saint's dilemma als Beweis genügt.

nicht nur eine Situation in der individuelle und kollektive Rationalität in den Wohlfahrtspräferenzen übereinstimmen in eine Interaktionssituation vom PD-Typ, sondern bei diesen Wohlfahrtsbewertungen führt individuell rationales Verhalten relativ zu diesen selbstlosen resultierenden Bewertungen gerade zu in die Katastrophe  $\langle \sigma_{11}, \sigma_{21} \rangle$ . (Wäre nur einer von beiden selbstlos, käme es zur kollektiven Entscheidung  $\langle \sigma_{11}, \sigma_{22} \rangle$  bzw.  $\langle \sigma_{12}, \sigma_{21} \rangle$ .)

Die Transferraten müssten sich also in Abhängigkeit von der jeweils vorliegenden Präferenzstruktur (nicht etwa den Wohlfahrtsniveaus der Beteiligten) ergeben. Damit wird jedoch die Wahl der Transferrate Funktion der Moralität von Handlungen und nicht Basis zur Bestimmung der Moralität von Handlungen. Der Versuch, den teleologischen Ansatz in der Sozialethik auf kardinaler Grundlage zu retten, kann damit ebenfalls als gescheitert gelten. Die Erweiterung der Informationsbasis bringt also für die Beziehung individueller und kollektiver Rationalität, wie sie sich anhand der ordinalen Darstellung gezeigt hat, keine grundsätzlich neuen Ergebnisse.<sup>1</sup>

---

<sup>1</sup> Das Verhältnis von individueller und kollektiver Rationalität entspricht weitgehend der Beziehung von Partikularinteressen und Gemeinwohl (vgl. dazu u.a. Runciman/Sen, *Games, justice and the general will*, *Mind* 74 (1965) 554-562, und G.O.Allen, *Le volonté de tous and le volonté général. A distinction and its significance*, *Ethics* 71 (1960/61) 263-274. Diese Ausweitung sollte jedoch im Zusammenhang der sozialetischen Rechtfertigung von *Institutionen* diskutiert werden, was den Rahmen dieser Arbeit sprengen würde.



## SCHLUSSBEMERKUNG

Für teleologische Theorien ist der moralische Wert einer Handlung allein durch den intrinsischen außermoralischen Wert ihrer Handlungsfolgen bestimmt. Für rein deontologische Theorien spielen die Handlungsfolgen für die moralische Bewertung einer Handlung keine Rolle. Häufig wird die Einteilung normativ-ethischer Theorien in deontologische und teleologische so verstanden, daß nicht-teleologische Theorien deontologisch sind.<sup>1</sup> Theorien, die die moralische Beurteilung von Entscheidungen nicht nur vom außermoralischen Wert der Handlungsfolgen, sondern noch von anderen Kriterien abhängig machen, haben wir jedoch gemischt deontologisch-teleologisch genannt.

Teleologische Theorien haben zwei Bestandteile.

Erstens eine Theorie des außermoralisch Wertvollen und zweitens eine Theorie der Maximierung des außermoralisch Wertvollen.

Zur eindeutigen Abgrenzung teleologischer Theorien ist es sinnvoll, als Maximierungstheorie das entscheidungstheoretische Rationalitätsmodell zu wählen: Für teleologische Theorien ist diejenige Handlung moralisch geboten, die den Erwartungswert der Funktion des außermoralischen Wertes ihrer Konsequenzen maximiert. Theorien, die den Wert einer Handlung nicht von einem außermoralischen, sondern (ausschließlich) vom moralischen Wert ihrer Folgen abhängig machen, sind, nach dem üblichen philosophischen Sprachgebrauch, zwar nicht teleologisch, aber sie haben mit teleologischen Theorien

eines gemeinsam, sie sind konsequentialistisch: Der moralische Wert einer Handlung bemißt sich ausschließlich nach dem Wert ihrer Folgen. Jede teleologische Theorie ist konsequentialistisch, die Umkehrung gilt nicht.

Wenn die Entscheidungstheorie einige notwendige Merkmale rationalen Verhaltens beschreibt, so ist moralisches Verhalten entweder irrational, oder es maximiert den Erwartungswert einer subjektiven Wertfunktion. Es wurde gezeigt, daß moralisches Verhalten nicht immer in Einklang ist mit subjektiven Konsequenzenbewertungen. Aber das ist nicht empirisch gemeint - empirisch gilt es ohnehin -, sondern analytisch. Auch die ideale moralische Person verhält sich nicht immer so, daß sie ihre Konsequenzenbewertungen (Erwartungswert-) maximiert. Es mag Konsequenzenbewertungen geben, die sie maximiert, aber es sind nicht 'ihre', es sind nicht diejenigen, die die propositionalen Einstellungen der Person wiedergeben.

Wenn die subjektiven Wertfunktionen der Entscheidungstheorie propositionale Einstellungen wiedergeben, wenn man die Entscheidungstheorie konsequentialistisch interpretiert, so gilt der Schluß:

- (1) Entscheidungstheoretische Rekonstruierbarkeit ist eine notwendige Bedingung von Entscheidungs-rationalität.
- (2) Moralische Entscheidungen sind rationale Entscheidungen.

- 
- (3) Eine adäquate Theorie normativer Ethik ist konsequentialistisch.

Wenn die Analysen in diesem Buch richtig waren, so gilt non-(3). An (2), denke ich, sollte man festhalten.

Die Entscheidungstheorie kann daher in ihrer konsequentialistischen Standardinterpretation nicht den Anspruch erheben, eine generelle Theorie rationalen Verhaltens zu sein. Das analytische Instrumentarium der Entscheidungstheorie ist jedoch dennoch ein wichtiges Hilfsmittel für die Klärung moralphilosophischer Fragestellungen.

ANHANG: MATHEMATISCHE UND LOGISCHE HILFSMITTEL

Die verwendeten mathematischen und logischen Hilfsmittel gehen über Schulwissen kaum hinaus. Um jedoch auch dem Leser, der mit formalen Darstellungsweisen nicht vertraut ist, die Lektüre zu ermöglichen, werden die verwendeten logischen und mathematischen Symbole hier kurz erläutert.

$\{a, b, c\}$	ist diejenige Menge, deren Elemente $a$ , $b$ und $c$ sind.
$\{x   Px\}$	vereint die <u>Menge</u> derjenigen Elemente $x$ , für die $Px$ wahr ist.
$x \in M$	$x$ ist ein <u>Element</u> der Menge $M$ . (Menge und Klasse verwenden wir bedeutungsgleich)
$M \subset M'$	Die Menge $M$ ist eine <u>Teilmenge</u> der Menge $M'$ , d.h. alle Elemente aus $M$ sind auch <u>Elemente</u> aus $M'$ .
$M_1 \neq M_2$	In (mindestens) einer der beiden Mengen $M_1$ und $M_2$ gibt es Elemente, die nicht in der anderen enthalten sind.
$M = M_1 \cap M_2$	$M$ ist die <u>Schnittmenge</u> von $M_1$ und $M_2$ , d.h. alle Elemente aus $M$ sind sowohl Elemente von $M_1$ wie von $M_2$ .
$M = M_1 \cup M_2$	$M$ ist die <u>Vereinigungsmenge</u> von $M_1$ und $M_2$ , d.h. alle Elemente aus $M$ sind Elemente von $M_1$ und/oder Elemente von $M_2$ .
$M = M_1 \setminus M_2$	Die Menge $M$ enthält genau diejenigen Elemente von $M_1$ , die nicht in $M_2$ enthalten sind.
$\text{Pot}(M)$	ist die Menge aller Teilmengen von $M$ .
$A, B, C$	seien Aussagen, also sprachliche Gebilde, die entweder wahr oder falsch sind. Dann sind auch $\sim A$ , $A \wedge B$ , $A \vee B$ , $A \Rightarrow B$ , $A \Leftrightarrow B$ Aussagen.
$\sim A$	ist genau dann wahr, wenn $A$ falsch ist. ("Nicht $A$ ")
$A \wedge B$	ist genau dann wahr, wenn sowohl $A$ wie $B$ wahr ist. (" $A$ und $B$ ")
$A \vee B$	ist genau dann wahr, wenn $A$ wahr ist oder wenn $B$ wahr ist, oder wenn $A$ und $B$ wahr sind. (" $A$ oder $B$ ")
$A \Rightarrow B$	ist genau dann falsch, wenn $A$ wahr ist und $B$ falsch. (" $A$ daraus folgt $B$ ")
$A \Leftrightarrow B$	ist genau dann wahr, wenn $A$ und $B$ wahr oder $A$ und $B$ falsch ist. (" $A$ genau dann, wenn $B$ ")

Um Klammern zu sparen, vereinbaren wir, daß die 'Bindungsstärke' der logischen Verknüpfungen in folgender Reihenfolge abnimmt:

$\sim \wedge \vee \Rightarrow \Leftrightarrow$  .

Statt  $(A \wedge B) \Rightarrow C$  können wir also schreiben  $A \wedge B \Rightarrow C$ , jedoch können wir bei  $A \wedge (B \Rightarrow C)$  die Klammer nicht weglassen.

$\forall x \in M: P_x$  Für alle Elemente  $x$  aus  $M$  gilt die Aussage  $P_x$ .

$\exists x \in M: P_x$  Es gibt mindestens ein Element  $x$  aus  $M$ , für das  $P_x$  wahr ist.

Beispiele: -  $M_1 \subset M_2$  ist genau dann wahr, wenn  $\forall x \in M_1: x \in M_2$

bzw. :  $\forall x [x \in M_1 \Rightarrow x \in M_2]$  .

-  $M = M_1 \cap M_2$  ist genau dann wahr, wenn

$\forall x \in M : [x \in M_1 \wedge x \in M_2]$  .

$\langle a, b \rangle$  ist ein (geordnetes) Paar.  $\langle a, b \rangle := \{\{a\}, \{a, b\}\}$  .

$\langle a_1, a_2, a_3 \rangle$  ist ein Tripel.

$\langle a_1, a_2, a_3, \dots, a_n \rangle$  ist ein  $n$ -Tupel.

Wir schreiben oft  $\langle a_i \rangle$  für  $\langle a_1, a_2, \dots, a_n \rangle$ , wenn aus dem Kontext klar ist, welches  $n$ -Tupel gemeint ist.

$X \times Y$  ist die Menge aller geordneten Paare  $\langle x, y \rangle$ , bei denen  $x$  aus  $X$  und  $y$  aus  $Y$  ist.

Eine zweistellige Relation  $R$  ist ein Tripel von Mengen

$\tilde{R} = \langle X, Y, R \rangle$ , so daß  $\tilde{R} \subset X \times Y$ .

Eine zweistellige Relation  $\tilde{R}$  auf  $X$  ist ein Tripel von Mengen  $\{X, X, R\}$  so daß  $R \subset X \times X$ .  $R$  und  $\tilde{R}$  benennen wir sonst nicht unterschiedlich.

Beispiel: Die Relation "größer als"  $>$  auf der Menge  $\{2, 3, 5\}$  besteht aus folgenden Elementen:  $\langle 3, 2 \rangle$ ,  $\langle 5, 2 \rangle$ ,  $\langle 5, 3 \rangle$

Die Relation  $>$  auf  $\{2, 3, 5\}$  ist also

die Menge  $\{\langle 3, 2 \rangle, \langle 5, 2 \rangle, \langle 5, 3 \rangle\}$ .

Eine Funktion ist eine zweistellige Relation  $\langle X, Y, R \rangle$  der Art, daß

zu jedem  $x$  aus  $X$  genau ein  $y$  aus  $Y$  mit  $\langle x, y \rangle \in R$  existiert.

Ist  $f = \langle X, Y, R \rangle$  eine Funktion, schreibt man für  $\langle x, y \rangle \in R$  auch  $f(x) = y$ , oder  $f: x \mapsto y$ .

$X$  nennt man den Definitionsbereich von  $f$ ,

$W = \{y \mid \exists x \in X : \langle x, y \rangle \in R\}$  den Wertebereich =  $Bi(f)$ .

Haben zwei Funktionen

$f_1 = \langle X_1, Y_2, R_1 \rangle$  und  $f_2 = \langle X_2, Y_2, R_2 \rangle$  die Eigenschaft, daß  $Y_1 = Y_2$ ,

so kann man die Hintereinanderausführungen der beiden Funktionen definieren:

$f_2 \circ f_1 = \langle X, Y, R \rangle$  mit  $X = X_1$ ,  $Y = Y_2$  und

$R = \{ \langle x, f_2(f_1(x)) \rangle \mid x \in X_1 \}$ .

Eine Hintereinanderausführung zweier Funktionen ist wieder eine Funktion.

$$(f_2 \circ f_1)(x) = f_2(f_1(x)).$$

### Weitere Symbole

Die Anzahl der Elemente einer Menge  $M$  ist  $\#M$ .

$\prod_{i \in K} M_i = M_1 \times M_2 \times M_3 \times \dots \times M_n$ , wobei  $n = \#K$ .

$\sum_{i=1}^n a_i = a_1 + a_2 + a_3 + \dots + a_n$

$\bigcup_{i \in K} M_i = M_1 \cup M_2 \cup M_3 \cup \dots \cup M_n$ , wobei  $n = \#K$

$\min(M)$  : kleinstes Element in  $M$  :  $\forall x \in M : \min(M) \leq x$

$\max(M)$  : größtes Element in  $M$  :  $\forall x \in M : \max(M) \geq x$

$\min(\{a \cdot b\}_{b \in B})$  : kleinstes Element der Menge  $\{a \cdot b \mid b \in B\}$ .

Darüberhinaus kommen noch weitere mathematische Hilfsmittel vor:  
Ableitungen von Funktionen,  $\sup(M)$ , ... Aber diese spielen für das  
Verständnis des Textes kaum eine Rolle.

ANHANG: ALLAIS' KRITIK AM BAYESIANISMUS

Allais<sup>3</sup> hat gezeigt, daß das tatsächliche Verhalten in bestimmten Wahlexperimenten dem Bayesschen Kriterium nicht entspricht.<sup>4</sup>

So entscheiden sich die meisten Versuchspersonen, vor die Wahl zwischen den Lotterien

$l_1 = \langle 1 | 1 \text{ Mio.F.} \rangle$ , d.h. sicheren 1 Mio.F. und

$l_2 = \langle 0,1 | 5 \text{ Mio.F.}, 0,89 | 1 \text{ Mio.F.}, 0,01 | 0 \text{ F.} \rangle$

gestellt, für  $l_1$ ; aber zugleich, vor die Wahl zwischen den Lotterien

$l_3 = \langle 0,11 | 1 \text{ Mio.F.}, 0,89 | 0 \text{ F.} \rangle$  und

$l_4 = \langle 0,1 | 5 \text{ Mio.F.}, 0,9 | 0 \text{ F.} \rangle$

gestellt, für  $l_4$ , was nicht im Einklang mit den eigenen

---

<sup>3</sup> Vgl. M. Allais, *Le comportement de l'homme rationnel devant le risque: Critiques des Postulats et axiomes de l'école Américaine*, *Econometrica* 21 (1953) 503-546; zur Diskussion von Allais' Kritik vgl. Allais/Hagen (1979).

<sup>4</sup> Vgl. a. Schneeweiß (1967), S.79.

Präferenzen sein kann, falls wir voraussetzen dürfen, daß die Axiome des Bernoullinutzens erfüllt sind.

Denn es gibt keine Nutzenfunktion des Geldes, die beide Entscheidungen zugleich als rational auszeichnet.

Es gibt keine Nutzenfunktion  $U$  des Geldes, so daß:

$$(1) U(1) > 0,1 \cdot U(5) + 0,89 \cdot U(1) + 0,01 \cdot U(0)$$

$$(2) 0,1 \cdot U(5) + 0,9 \cdot U(0) > 0,11 \cdot U(1) + 0,89 \cdot U(0)$$

denn aus (1)  $\wedge$  (2) folgt

$$U(1) + 0,1 \cdot U(5) + 0,9 \cdot U(0) >$$

$$0,1 \cdot U(5) + 0,89 \cdot U(1) + 0,01 \cdot U(0) + 0,11 \cdot U(1) + 0,89 \cdot U(0)$$

bzw.

$U(1) + 0,1 \cdot U(5) + 0,9 \cdot U(0) > U(1) + 0,1 \cdot U(5) + 0,9 \cdot U(0)$ ,  
was ein logischer Widerspruch ist.

Das Beispiel ist so gewählt, daß für beliebige Nutzenfunktionen gilt:

$$l_1 \succ_r l_2 \Rightarrow l_3 \succ_r l_4$$

$$l_3 \succ_r l_4 \Rightarrow l_1 \succ_r l_2$$

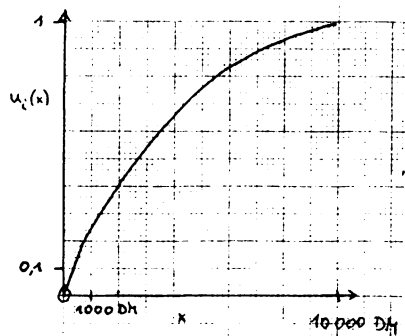
$$l_1 \prec_r l_2 \Rightarrow l_3 \prec_r l_4$$

$$l_3 \prec_r l_4 \Rightarrow l_1 \prec_r l_2$$

Der logische Widerspruch ist also unvermeidlich.

#### BRLT-INTERPRETATION SUBJEKTIVER BEWERTUNGSFUNKTIONEN UND RISIKOBEREITSCHAFT

Der Graph der nebenstehenden Abbildung zeigt die Nutzenfunktion des Geldes einer Person  $\iota$ . Dieser Graph ist die Kurzform für eine Menge von unendlich vielen Dispositionsaussagen über die Person  $\iota$ . Welcher Art sind diese?



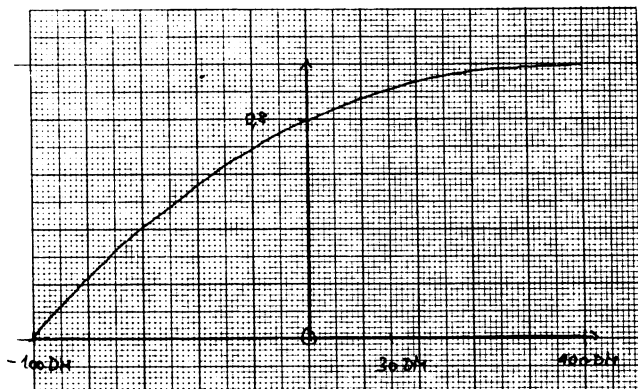


Üblicherweise wird der Graph der Abbildung von  $S$  als Indifferenzkurve zwischen jeweils einer Lotterie - etwa mit 40 % Wahrscheinlichkeit 10.000 DM zu gewinnen und mit 60 % Wahrscheinlichkeit leer auszugehen - und einem sicheren Betrag - hier 2000 DM - interpretiert.<sup>1</sup> Man spricht dann von einer  $\rho$ -BRLT-Zuordnung.<sup>2</sup>

Der Graph stellt also u.a. folgende Indifferenzen dar:

- 2000 DM und  $\langle 0,4 \mid 10.000 \text{ DM}, 0,6 \mid 0 \text{ DM} \rangle$
- 4000 DM und  $\langle 0,7 \mid 10.000 \text{ DM}, 0,3 \mid 0 \text{ DM} \rangle$
- 6000 DM und  $\langle 0,84 \mid 10.000 \text{ DM}, 0,16 \mid 0 \text{ DM} \rangle$
- 8000 DM und  $\langle 0,94 \mid 10.000 \text{ DM}, 0,6 \mid 0 \text{ DM} \rangle$

Sehen wir uns noch eine zweite Bewertungsfunktion an:



Hier ist die Person z.B. indifferent zwischen der Lotterie  $\langle 0,8 \mid 100 \text{ DM}, 0,2 \mid -100 \text{ DM} \rangle$  und 0 DM. M.a.W.: Wenn ihr eine Lotterie angeboten wird, die bei einem Einsatz von 100 DM mit höherer Wahrscheinlichkeit als mit 80% 200 DM auswirft, ist die Person zu dem Spiel bereit. Die Person ist indifferent zwischen 30 DM auf die Hand und einem Spiel, bei dem sie mit 90% Wahrscheinlichkeit 100 DM gewinnt und mit 10% 100 DM verliert, bzw. einem Spiel, das bei einem Einsatz von 100 DM mit 90% Wahrscheinlichkeit 200 DM auswirft.

<sup>1</sup> Vgl. H. Raiffa, *Decision analysis*, London 1978, S.57f und S.66-70.

<sup>2</sup> BRLT steht für basic reference lottery ticket.

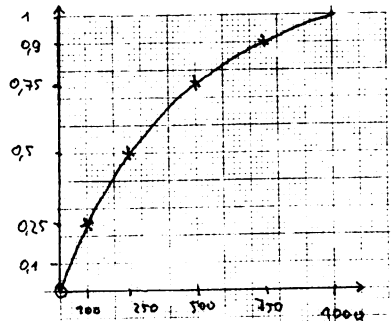
In der ökonomischen Literatur hat es sich eingebürgert, konvexe Nutzenfunktionen (des Geldes) als Ausdruck einer Risikoscheu des betreffenden Individuums zu interpretieren.

Bei genauerer Betrachtung zeigt sich jedoch, daß das ein irreführender Sprachgebrauch ist. Nehmen wir die Nutzenfunktion von Kühlschränken. Ein Kühlschrank ist mir sehr viel wert, dafür bin ich bereit 1000 DM zu bezahlen; ein zweiter Kühlschrank bringt mir nicht mehr viel, aber wenn ich ihn für 300 DM kaufen könnte, wäre er mir das wert. Für einen dritten zahle ich keine müde Mark mehr, er nimmt bloß unnötig Platz weg.<sup>1</sup> So ähnlich verhält es sich mit den allermeisten Gütern: ein zusätzliches Stück davon wird mit zunehmendem Besitz immer weniger wert: sein Grenznutzen fällt. Daher ist es nur natürlich, daß auch der Grenznutzen des Geldes abnimmt. Diese Überlegung macht an keiner Stelle von Entscheidungen unter Risiko Gebrauch: Konkave Nutzenfunktionen des Geldes sind - wie hier skizziert - erklärbar ohne Rückgriff auf die psychologische Eigenschaft der Risikoscheu.

Dennoch verbirgt sich hinter diesem scheinbar harmlosen Sprachgebrauch ein ernsthaftes Problem der ökonomischen Präferenz- bzw. Nutzentheorie.

Nehmen wir an, eine Person ist indifferent zwischen

1.  $\langle 0,25 | 1000, 0,75 | 0 \rangle$  und 100 DM
2.  $\langle 0,5 | 1000, 0,5 | 0 \rangle$  und 250 DM
3.  $\langle 0,75 | 1000, 0,25 | 0 \rangle$  und 500 DM
4.  $\langle 0,9 | 1000, 0,1 | 0 \rangle$  und 750 DM



<sup>1</sup> Wir wollen annehmen, daß ein Wiederverkauf nicht möglich ist.

Danach müßte diese Person indifferent zwischen der Lotterie  $\langle 0,5|1000, 0,5|250 \rangle$  und 500 DM sein, denn  
 $\langle 0,5|1000, 0,5|250 \rangle =$   
 $\langle 0,5|1000, 0,5|\langle 0,5|1000, 0,5|0 \rangle \rangle =$   
 $\langle 0,5|1000, 0,25|1000, 0,25|0 \rangle =$   
 $\langle 0,75|1000, 0,25|0 \rangle = 500 \text{ DM} , \text{ wegen (3)} .$

Wenn sie jedoch indifferent zwischen der Lotterie  $\langle 0,5|1000, 0,5|250 \rangle$  und 450 DM ist, so hätte sie nach der  $\rho$ -BRLT-Nutzentheorie inkonsistente Präferenzen. Das ist nicht unbedingt überzeugend, denn die Lotterie  $\langle 0,9|1000, 0,1|0 \rangle$  stellt insofern ein geringeres Risiko als  $\langle 0,5|1000, 0,5|250 \rangle$  dar, als die Wahrscheinlichkeiten näher bei 1 bzw. 0 liegen. Streben die Wahrscheinlichkeiten gegen den Sicherheitsfall, so sollte auch die Nutzenbewertung der Lotterie gegen die Nutzenbewertung dieser Geldsumme streben. Unter diesem Aspekt könnte eine risikoscheue Person zwischen  $\langle 0,5|1000, 5|250 \rangle$  und z.B. 450 DM indifferent sein.

#### EINE MODIFIKATION DES BAYESSCHEN ENTSCHEIDUNGSKRITERIUMS

Stellen wir uns vor, eine Person habe in einem Bereich 0 - 10 DM eine (annähernd) lineare Nutzenfunktion der monetären Auszahlung. Es braucht uns hier nicht zu interessieren, wie sich das feststellen läßt.

Nehmen wir weiter an, daß es sich um eine risikoscheue Person handelt, was sich dadurch äußert, daß ihr Indifferenzverhalten zwischen Lotterien (und sicheren Auszahlungen) durch das Entscheidungskriterium  $\Lambda$

$$\tilde{u}(\langle \rho|\alpha, (1-\rho)|\beta \rangle) = \rho \cdot \tilde{u}(\alpha) + (1-\rho) \tilde{u}(\beta) - \rho(1-\rho) (\tilde{u}(\alpha) - \tilde{u}(\beta)) \lambda$$

mit  $\lambda > 0$  (Risikoscheu); mit NEW(1) Korrekturterm T  
mit  $\lambda < 0$  (Risikofreudigkeit)

gekennzeichnet ist, o.B.d.A. sei dabei  $\tilde{u}(\alpha) > \tilde{u}(\beta)$ .

Nehmen wir an,  $l = \langle \rho|\alpha, (1-\rho)|\beta \rangle$  und

$l' = \langle \rho'|\alpha', (1-\rho')|\beta' \rangle$  seien Lotterien mit dem gleichen Nutzenerwartungswert. Nehmen wir weiterhin an, daß die

Auszahlungen der Lotterien gleich sind. In diesem Fall würde eine risikoscheue Person  $l$  genau dann vorziehen, wenn die Wahrscheinlichkeiten von  $l$  näher bei 1 (bzw.) 0 sind als die Wahrscheinlichkeiten von  $l'$ . Unterscheiden sich jedoch  $l$  und  $l'$  weder in ihrem Nutzwert noch in ihren Wahrscheinlichkeiten, würde eine risikoscheue Person  $l$  genau dann vorziehen, wenn sich die subjektiven Bewertungen der Auszahlungen von  $l$  weniger unterscheiden als von  $l'$ .

$$NEW(l) = NEW(l') \Leftrightarrow$$

$$NEW(\langle \rho | \alpha, (1-\rho) | \beta \rangle) = NEW(\langle \rho' | \alpha', (1-\rho') | \beta' \rangle) \Leftrightarrow$$

$\rho \cdot \hat{0}(\alpha) + (1-\rho) \cdot \hat{0}(\beta) = \rho' \cdot \hat{0}(\alpha') + (1-\rho') \cdot \hat{0}(\beta')$ , das schließt nicht aus, daß (1)  $|\hat{0}(\alpha) - \hat{0}(\beta)| \leq |\hat{0}(\alpha') - \hat{0}(\beta')|$

$$\text{und} \quad (2) \quad \rho(1-\rho) \leq \rho'(1-\rho').$$

Ist  $\hat{0}(\alpha) = \hat{0}(\alpha')$  und  $\hat{0}(\beta) = \hat{0}(\beta')$ , gilt:

$$I \quad l \succ_l l' \Leftrightarrow (2), \text{ falls } l \text{ risikoscheu}$$

$$II \quad l \succ_l l' \Leftrightarrow \sim(2), \text{ falls } l \text{ risikofreudig.}$$

Ist  $\rho = \rho'$ , gilt:

$$III \quad l \succ_l l' \Leftrightarrow (1), \text{ falls } l \text{ risikoscheu}$$

$$IV \quad l \succ_l l' \Leftrightarrow \sim(1), \text{ falls } l \text{ risikofreudig}$$

Diese vier Forderungen I - IV werden vom Bayesschen Entscheidungskriterium nicht erfüllt, während  $\hat{\Lambda}$  (für Risikoscheu  $\lambda > 0$ ; für Risikofreudigkeit  $\lambda < 0$ ) diesen Forderungen entspricht.

Für sehr kleine und sehr große  $\rho$  nähert sich der Korrekturterm  $T$  dem Wert Null, womit wir das intuitiv erwünschte Ergebnis erhalten:

$$\lim_{\rho \rightarrow 1} \tilde{u}(\langle \rho | \alpha, (1-\rho) | \beta \rangle) = \hat{0}(\alpha) \quad \text{und}$$

$$\lim_{\rho \rightarrow 0} \tilde{u}(\langle \rho | \alpha, (1-\rho) | \beta \rangle) = \hat{0}(\beta).$$

Für risikoscheue Personen ist  $\lambda > 0$ , für risikofreudige  $\lambda < 0$ . Ein differenzierteres Modell könnte  $\lambda$  als 3-stellige Funktion von  $\rho$ ,  $\alpha$  und  $\beta$  einführen. Es ist hier nicht

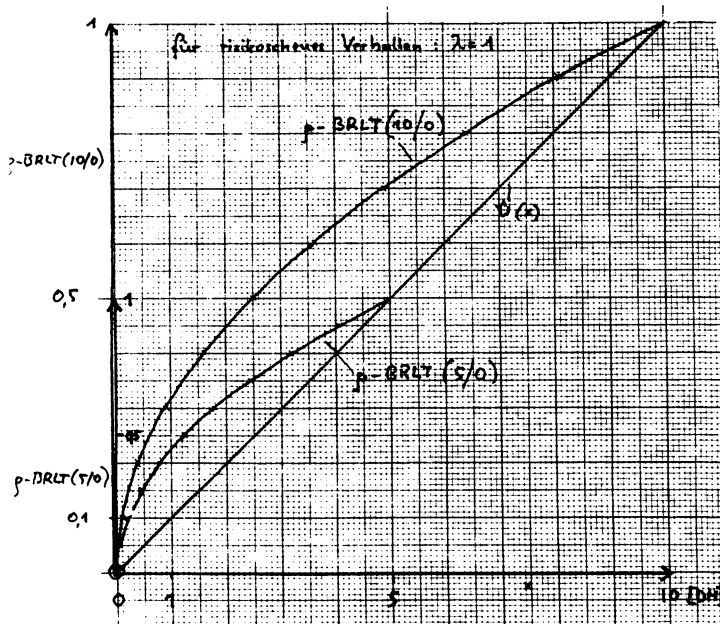
er Ort, das zu vertiefen. Es genügt, die Konsequenz eines solchen Modells der Risikoscheu an einem einfachen Fall aufzuzeigen:

Sei  $\lambda = 1$ ;

$\alpha = 10 \text{ DM}$ ;  $\beta = 0 \text{ DM}$ ;

$L = \langle \rho | \alpha, (1-\rho) | \beta \rangle$ ;

0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	0,4	0,9	1,6	2,5	3,6	4,9	6,4	8,1



Unter der Annahme einer linearen Nutzenfunktion  $U_1$  des Geldes im Bereich  $[0 \text{ DM}, 10 \text{ DM}]$ , erhalten wir eine zweite Funktion, die die Indifferenzen der Person 1 zwischen Lotterien  $\langle \rho | \alpha, (1-\rho) | \beta \rangle$  (wobei  $\alpha = 10 \text{ DM}$  und  $\beta = 0 \text{ DM}$ ) und festen Geldbeträgen darstellt.

Die üblichen durch  $\rho\text{-BRLT}$ -Funktionen dargestellten Nutzenbewertungen sind in dem Sinne unabhängig von der Auszahlung  $\alpha$  und  $\beta$  der Lotterien, als die Graphen sich

als Teilstück einer potentiell den gesamten Bereich durchlaufenden Nutzenfunktion ergeben.

Das ist hier nicht der Fall: die entsprechende BRLT-Bewertung ergibt für  $\alpha = 5$  DM und  $\beta = 0$  DM die Indifferenzkurve  $\rho$ -BRLT(5/0), die mit  $\rho$ -BRLT(10/0) nicht konsistent ist.

$\alpha = 5$  DM,  $\beta = 0$  DM

$\rho$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$u(1)$	0,05	0,2	0,45	0,8	1,25	1,8	2,45	3,2	4,0

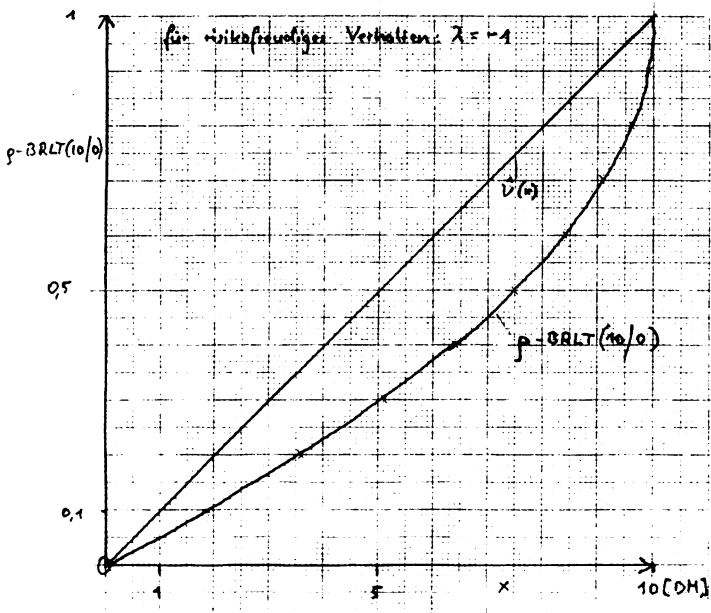
Für eine risikofreudige Person sei

$$\lambda = -1$$

für  $\alpha = 10$  DM,  $\beta = 0$  DM

$$l = \langle \rho | \alpha, (1-\rho) | \beta \rangle$$

$\rho$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$u(1)$	1,9	3,6	5,1	6,4	7,5	8,4	9,1	9,6	9,9



Unter der Annahme einer linearen Nutzenfunktion des Geldes erhalten wir also eine zweite Funktion, die die Risikoscheu der betreffenden Person darstellt.

Die Grundidee wird an diesem Beispiel deutlich:

Eine Person ist risikoscheu, wenn für sie der Nutzen einer Lotterie geringer ist als ihr Nutzenerwartungswert.

Eine Person ist risikofreudig, wenn für sie der Nutzen einer Lotterie größer ist als ihr Nutzenerwartungswert.

Was kann aus diesen Überlegungen gefolgert werden?

(1) Die Lotteriesubstituierbarkeit muß aufgegeben werden. Scheinbar inkonsistentes Wahlverhalten, wie die Indifferenz zwischen

1.  $\langle 0,75|1000, 0,75|0 \rangle$  und 500 DM
2.  $\langle 0,5|1000, 0,5|0 \rangle$  und 250 DM und
3.  $\langle 0,5|1000, 0,5|250 \rangle$  und 450 DM

läßt sich als konsistentes, aber risikoscheues Verhalten interpretieren.

Das Entscheidungskriterium  $\Lambda$  würde auch erlauben, diese Präferenzen als risikofreudiges Verhalten zu interpretieren. Das setzt allerdings eine ziemlich seltsame Geldnutzenfunktion voraus. I.a. ist es natürlich, eine konvexe Gestalt des Graphen anzunehmen. Während jedoch unter besonderen Umständen die Konvexität aufgegeben werden kann, was besonders bei positionellen Gütern nahe liegt,<sup>1</sup> scheint mir positive Monotonie der Geldnutzenfunktion in jedem Fall unverzichtbar.<sup>2</sup>

Die Eingrenzung 'zulässiger Werte für  $\lambda$ ' und die Bestimmung der 'zulässigen' Nutzenfunktionen und gegebenenfalls  $\lambda$ -Funktionen für rationales Präferenz- und Entscheidungsverhalten müßte von solchen Postulaten Gebrauch

---

<sup>1</sup> Zur Theorie der positionellen Güter vgl. F. Hirsch, *Social limits to growth*, Cambridge/Mass. 1976.

<sup>2</sup> Und i.a. auch die positive Monotonie der Nutzenfunktionen anderer Güter in einer Marktwirtschaft.

machen, wobei sich vermutlich eine unabhängige Festlegung als inadäquat erweisen würde, so daß jeweils das geordnete Paar  $\langle \lambda_1, \hat{0}_1 \rangle$  beurteilt werden müßte.

Setzen wir in unserem Fall o.B.d.A.  $\hat{0}(0 \text{ DM}) = 0$  und  $\hat{0}(1000 \text{ DM}) = 1$ , so ergeben sich folgende Nutzenbewertungen:

$$\hat{0}(500 \text{ DM}) = 0,75, \hat{0}(250 \text{ DM}) = 0,5.$$

Mit diesen Nutzenbewertungen steht

$\hat{u}(\langle 0,5 | 1000, 0,5 | 250 \rangle) = \hat{0}(450 \text{ DM})$  in Widerspruch, wenn der Nutzen einer Lotterie  $\hat{u}(l)$  dem Nutzenerwartungswert ihrer Auszahlungen - ohne Berücksichtigung des Risikos - gleichgesetzt wird. (Beweis s.o.)

Das Gleichungssystem

$$(1) \hat{u}(\langle 0,75 | 1000 \text{ DM}, 0,25 | 0 \text{ DM} \rangle) = \hat{0}(500 \text{ DM})$$

$$(2) \hat{u}(\langle 0,5 | 1000 \text{ DM}, 0,5 | 0 \text{ DM} \rangle) = \hat{0}(250 \text{ DM})$$

$$(3) \hat{u}(\langle 0,5 | 1000 \text{ DM}, 0,5 | 250 \text{ DM} \rangle) = \hat{0}(x \text{ DM})$$

hat keine nicht-triviale Lösung in  $\{\hat{0}(0), \hat{0}(250), \hat{0}(500), \hat{0}(1000)\}$ ,

wenn  $x \neq 500$  und

$$\hat{u}(\langle \rho | \alpha, (1-\rho) | \beta \rangle) = \rho \cdot \hat{0}(\alpha) + (1-\rho) \cdot \hat{0}(\beta) ; \rho \in [0,1].$$

Dagegen hat das Gleichungssystem  $(1') \wedge (2') \wedge (3')$  eine nicht-triviale Lösung auch für  $x = 450$ , wenn

$$\hat{u}(\langle \rho | \alpha, (1-\rho) | \beta \rangle) = \rho \cdot \hat{0}(\alpha) + (1-\rho) \cdot \hat{0}(\beta) -$$

$$- \rho \cdot (1-\rho) \cdot (\hat{0}(\alpha) - \hat{0}(\beta)) \cdot \lambda :$$

$$(1') 0,75 \cdot \hat{0}(1000 \text{ DM}) + 0,25 \cdot \hat{0}(0 \text{ DM}) - 0,75 \cdot 0,25 \cdot 1 \cdot \lambda = \hat{0}(500 \text{ DM})$$

$$(2') 0,5 \cdot \hat{0}(1000 \text{ DM}) + 0,5 \cdot \hat{0}(0 \text{ DM}) - 0,5 \cdot 0,5 \cdot 1 \cdot \lambda = \hat{0}(250 \text{ DM})$$

$$(3') 0,5 \cdot \hat{0}(1000 \text{ DM}) + 0,5 \cdot \hat{0}(250 \text{ DM}) - 0,5 \cdot 0,5 \cdot (1 - \hat{0}(250 \text{ DM})) \cdot \lambda = \hat{0}(450 \text{ DM})$$

hat für eine leicht risikoscheue Person mit  $\lambda = 0,5$  eine Lösung für folgende Nutzenbewertungen des Geldes:

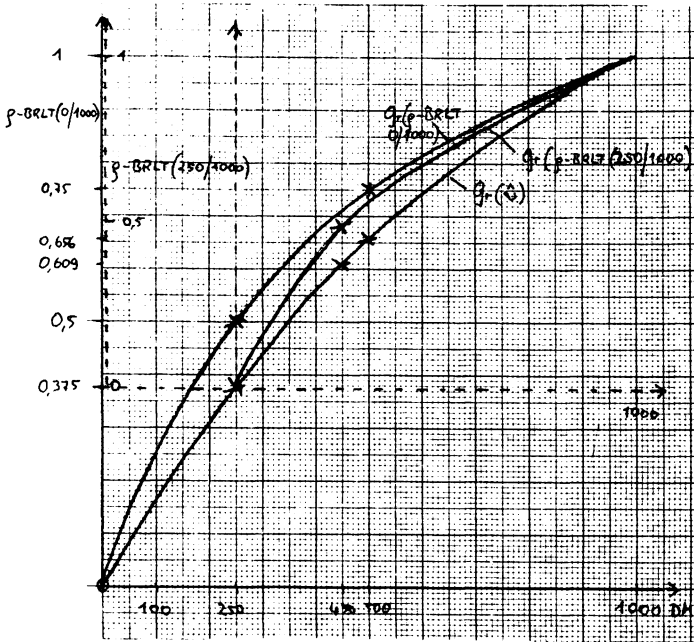
$$\hat{0}(500 \text{ DM}) = 0,656; \hat{0}(250 \text{ DM}) = 0,375; \hat{0}(450 \text{ DM}) = 0,609.$$

Damit können die Präferenzen (1), (2), (3) als Folge einer leicht konvexen Geldnutzenfunktion und leichter Risikoscheue interpretiert werden.  $(1) \wedge (2) \wedge (3)$  ist entscheidungs-



konsistent bezüglich

$$\langle \lambda, 0(250), 0(450), 0(500) \rangle = \langle 0,5, 0,375, 0,609, 0,656 \rangle.$$



(2) Eine weitere Konsequenz ist, daß die Gestalt der Nutzenfunktion allein über die Risikobereitschaft der betreffenden Person nichts aussagt.

Eine konvexe  $\rho$ -BRLT-Indifferenzkurve ist dann Ausdruck von Risikoscheu ( $\lambda > 0$ ), wenn die Nutzenfunktion  $0(x)$  konkav ist. Das ist aber keine notwendige Bedingung, der Satz ist nicht umkehrbar.

Eine konvexe  $\rho$ -BRLT-Indifferenzkurve ist dann Ausdruck eines abnehmenden Grenznutzens, wenn die Person risikofreudig ist. Der Satz ist ebenfalls nicht umkehrbar.

Eine konvexe  $\rho$ -BRLT-Indifferenzfunktion kann Ausdruck von Risikoscheu sein, muß aber nicht. Ebenso kann eine konvexe  $\rho$ -BRLT-Indifferenzfunktion Ausdruck eines abneh-

menden Grenznutzens sein, sie kann auch Folge sowohl von abnehmendem Grenznutzen wie von Risikoscheu sein.

Entscheidend für die Risikobereitschaft ist, ob die  $\rho$ -BRLT-Indifferenzfunktion nach oben ( $\lambda > 0$ ) oder unten ( $\lambda < 0$ ) von der Nutzenfunktion abweicht.

Die neue Definition für konsistentes Entscheidungsverhalten würde nach dieser Modifikation lauten:

Eine Klasse von Entscheidungen bzw. eine Präferenzrelation von Handlungen einer Person  $i$  ist konsistent genau dann, wenn es eine subjektive Wahrscheinlichkeitsfunktion  $\rho_i$ , eine Bernoulli-Nutzenfunktion  $\hat{U}_i$  und ein  $\lambda$  gibt, so daß eine Entscheidung genau dann präferiert wird, wenn ihr Wert gemäß unserem Entscheidungskriterium  $\Lambda$  höher ist.

Ist  $\lambda$  eine Funktion der subjektiven Wahrscheinlichkeit und der Auszahlungen, so lautet die entsprechende Konsistenzdefinition:

"Eine Klasse von Entscheidungen bzw. eine Präferenzrelation von Handlungen einer Person  $i$  ist konsistent genau dann, wenn es eine subjektive Wahrscheinlichkeitsfunktion  $\rho_i$ , eine Bernoulli-Nutzenfunktion  $\hat{U}_i$  und eine Funktion  $\lambda: \langle \rho_i, \hat{U}_i(\alpha), \beta \rangle \rightarrow \mathbb{R} \in [0, 1]$  gibt, so daß eine Entscheidung genau dann präferiert wird, wenn ihr Wert  $\hat{U}_i(1)$  gemäß  $\Lambda$  höher ist."

Diese Definition wäre allerdings zu weit, denn eine hinreichend 'bizarre' Funktion  $\lambda$  macht so ziemlich jedes (empirisch) denkbare Entscheidungsverhalten 'konsistent'. Zusätzliche Bedingungen an die Funktion  $\lambda$  sind daher für eine adäquate Definition unumgänglich.

Diese Möglichkeit einer nicht-bayesianischen Konsistenzdefinition individuellen Entscheidungsverhaltens in Risikosituationen ohne Interaktion wurde hier angeführt, um deutlich zu machen, daß manche Mängel der v. Neumann/Morgensternschen Nutzentheorie<sup>1</sup> durchaus 'im Geist der modernen Entscheidungstheorie' behoben werden können.

---

<sup>1</sup> Vgl. dazu auch Dreyfus/Dreyfus (1978), Arnaszus (1974), Höffe (1975)

ANHANG: ZUR INTERDEPENDENZPROBLEMATIK

Sei  $(v_1^{(n)})$ ,  $n \in \mathbb{N}$  die Folge der subjektiven Bewertungen der ersten und  $(v_2^{(n)})$ ,  $n \in \mathbb{N}$  diejenige der zweiten Person; die anfänglichen Bewertungen seien  $v_1^{(1)} = a_1$  und  $v_2^{(2)} = a_2$ ; dann gilt für  $n \in 2\mathbb{N}$ :

$$v_1^{(n)} = f_1(f_2(f_1 \dots (f_1(a_2)) \dots))$$

$$v_2^{(n)} = f_2(f_1(f_2 \dots (f_2(a_1)) \dots)) \quad \text{jeweils } n-2 \text{ Hintereinanderausführungen}$$

und für  $n \in 2\mathbb{N} + 1$ :

$$v_1^{(n)} = f_1(f_2(f_1 \dots (f_2(a_1)) \dots))$$

$$v_2^{(n)} = f_2(f_1(f_2 \dots (f_1(a_2)) \dots)) \quad \text{jeweils } n-2 \text{ Hintereinanderausführungen}$$

Die ersten Glieder einer Teilfolge der Nutzenbewertungsfolge der ersten Person seien hier für eine Anfangsbeurteilung von

$v_1^{(1)}(z) = a_1 = 0,5$  angeführt:

$(v_1^{(n)})$ ,  $n \in 2\mathbb{N} + 1$ ,  $a_1 = 0,5$ .

$$v_1^{(3)} = f_1(f_2(a_1)) = 1 - 0,25 = 0,75$$

$$v_1^{(5)} = f_1(f_2(f_1(f_2(a_1)))) = 1 - 0,75^2 = 0,475$$

$$v_1^{(7)} = f_1(f_2(f_1(f_2(f_1(f_2(a_1))))) = 1 - 0,475^2 = 0,8086$$

$$v_1^{(9)} = f_1(f_2(f_1(f_2(f_1(f_2(f_1(f_2(a_1)))))))) = 1 - 0,8086^2 = 0,3462$$

$$v_1^{(11)} = \dots = 1 - 0,3462^2 = 0,8801$$

$$v_1^{(13)} = \dots = 1 - 0,8801^2 = 0,2253$$

$$v_1^{(15)} = \dots = 1 - 0,2253^2 = 0,9494$$

$$v_1^{(17)} = \dots = 1 - 0,9494^2 = 0,0989$$

$$v_1^{(19)} = \dots = 1 - 0,0989^2 = 0,9902$$

$$v_1^{(21)} = \dots = 1 - 0,9902^2 = 0,0195$$

$$v_1^{(23)} = \dots = 1 - 0,0195^2 = 0,9996$$

$$v_1^{(25)} = \dots = 1 - 0,9996^2 = 0,0007$$

$$\dots = \dots = \dots$$

Die Nutzenwerte nähern sich also sehr rasch den beiden Extremen (und zwar intrapersonell).

Für die Nutzeninterdependenzen unseres Beispiels konvergiert die Teilfolge von  $(v_1^{(n)})$ ,  $n \in \mathbb{N}$ , die aus dem 5., dem 9., dem 13. ... Glied aus  $(v_1^{(n)})$  besteht, also die Nutzenbewertungsfolge  $(v_1^{(n)})$ ,  $n \in 4\mathbb{N}+1$  für  $a_1=0,5$  gegen Null, während die Teilfolge von  $(v_1^{(n)})$ ,  $n \in \mathbb{N}$ , die aus dem 3., 7., 11., 15. ... Glied aus  $(v_1^{(n)})$  besteht, also die Nutzenbewertungsfolge  $(v_1^{(n)})$ ,  $n \in 4\mathbb{N}+3$  für  $a_1=0,5$  gegen Eins konvergiert:

$$\text{für } a_1=0,5 : - \lim_{\substack{n \in 4\mathbb{N}+1 \\ n \rightarrow \infty}} v_1^{(n)} = 0 ;$$

$$- \lim_{\substack{n \in 4\mathbb{N}+3 \\ n \rightarrow \infty}} v_1^{(n)} = 1 ;$$

Die Teilfolge von  $(v_1^{(n)})$ , die aus dem 1., 3., 5., 7. ... Glied aus  $(v_1^{(n)})$  besteht, also die Nutzenbewertungsfolge  $(v_1^{(n)})$ ,  $n \in 2\mathbb{N}+1$  ist damit für  $a_1=0,5$  divergent und nimmt hintereinander allein aufgrund iterierter Offenlegungen immer extremere Werte an, die sich einer abwechselnden Bewertung des Zustandes  $z$  als summum bonum bzw. summum malum der ersten Person zunehmend annähern.

Nun läßt sich zeigen, daß sich diese unangenehme Konsequenz der Nutzeninterdependenz nicht nur für diese, sondern für alle anfänglichen Bewertungen ergibt. Die Bewertungen streben nicht nur keiner Gleichgewichtsbewertung zu, sondern wechselseitige Information verschlimmert nur jedesmal die Lage: Die (intrapersonellen) Änderungen der subjektiven Bewertung von  $z$  werden immer heftiger, um am Ende zwischen den beiden äußersten Bewertungen zu schwanken.

Es ist, denke ich, zweckmäßig, es an diesem Beispiel einmal durchzurechnen, weil das den Blick für die auch in der Realität häufig recht verwickelte Situation wechselseitiger Bewertungsinterdependenz schärft.

Die Nachfolger-Funktion  $F: [0,1] \ni v_1^{(n)} \mapsto v_1^{(n+4)} \in [0,1]$   
 beider Teilfolgen  $(v_1^{(n)})_{n \in 4\mathbb{N}_0 + 3}$  und  $(v_1^{(n)})_{n \in 4\mathbb{N} + 1}$

ist die Funktion  $F: [0,1] \ni x \mapsto f_1(f_2(f_1(f_2(x))))$

bzw. die Funktion  $F: [0,1] \ni x \mapsto 1 - (1 - x^2)^2 \in [0,1]$

Der Graph von  $F$  verläuft oberhalb der Winkelhalbierenden für alle Werte von  $x$ , für die gilt:

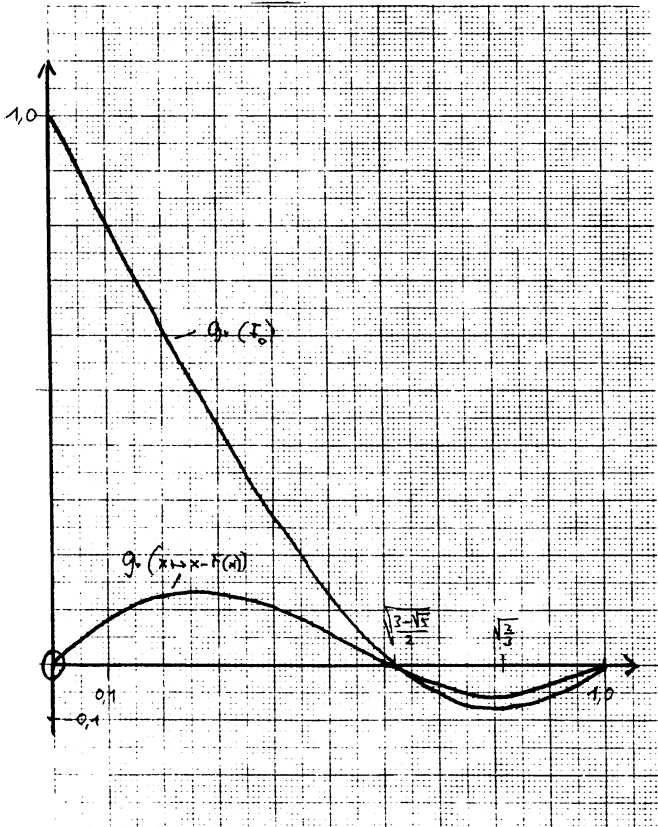
$$x < 1 - (1 - x^2)^2 \quad ; \quad \text{bzw.} :$$

$$0 > x^4 - 2x^2 + x \quad ; \quad \text{bzw. (da } x \in [0,1]) :$$

$$0 > x^3 - 2x + 1 \wedge x \in (0,1) \quad ;$$

In  $[0,1]$  hat der Graph der Funktion  $F: [0,1] \ni x^3 \mapsto x - 2x + 1$

bei  $x = \sqrt{\frac{2}{3}} \approx 0,8165$ , da  $F_0''(\sqrt{\frac{2}{3}}) > 0$ , liegt hier ein lokales, aber wie genauere Analyse zeigt, auch das absolute Minimum des Graphen im Bereich  $[0,1]$ .



Für  $v_1^{(n)} \in (0, \sqrt{\frac{3-\sqrt{5}}{2}})$  ist das nachfolgende Glied der Bewertungsfolgen  $(v_1^{(n)})_{n \in 4N_0+3}$  und  $(v_1^{(n)})_{n \in 4N+1}$  jeweils kleiner und für

$v_1^{(n)} \in (\sqrt{\frac{3-\sqrt{5}}{2}}, 1)$  ist es jeweils größer als  $v_1^{(n)}$ .

Damit sind diese beiden Teilfolgen für ein Anfangsglied  $v_1^{(3)}$  bzw.  $v_1^{(5)} \in (0, \sqrt{\frac{3-\sqrt{5}}{2}})$  strikt monoton fallend und für ein Anfangsglied

$v_1^{(3)}$  bzw.  $v_1^{(5)} \in (\sqrt{\frac{3-\sqrt{5}}{2}}, 1)$  strikt monoton steigend.

Da sich zeigen läßt, daß sich für die Nutzeninterdependenzen dieses Beispiels, außer den beiden Grenzbewertungen der Nutzenskala keine weiteren Häufungspunkte ergeben, erhalten wir für alle Anfangsbewertungen  $a_1$ , für die folgende Bedingung (\*) gilt:

$$f_1(f_2(f_1(f_2(a_1)))) > a_1 \quad \wedge$$

$$f_1(f_2(f_1(f_2(f_1(f_2(a_1)))))) < f_1(f_2(a_1)) \quad \vee$$

$$f_1(f_2(f_1(f_2(a_1)))) < a_1 \quad \wedge$$

$$f_1(f_2(f_1(f_2(f_1(f_2(a_1)))))) > f_1(f_2(a_1)) \quad ;$$

oder, was dazu äquivalent ist

$$a_1^3 - 2a_1 + 1 < 0 \quad \wedge$$

$$[f_1(f_2(a_1))]^3 - 2f_1(f_2(a_1)) + 1 > 0 \quad \vee$$

$$a_1^3 - 2a_1 + 1 > 0 \quad \wedge$$

$$[f_1(f_2(a_1))]^3 - 2f_1(f_2(a_1)) + 1 < 0$$

eine intrapersonell divergente Bewertungsfolge, deren Häufungspunkte das summum malum und das summum bonum der Person darstellen, denn für die hier nicht behandelten Glieder der Folge, also für die Glieder der Teilfolgen  $(v_1^{(n)})_{n \in 4N_0+2}$  und  $(v_1^{(n)})_{n \in 4N}$  gilt natürlich die Argumentation ganz analog - hier dann für  $a_2$ .<sup>1</sup>

---

<sup>1</sup> Allerdings strebt die Teilfolge  $(v_1^{(n)})_{n \in 4N_0+2}$  nicht für  $a_2 \in (0, \sqrt{\frac{3-\sqrt{5}}{2}})$ , sondern für  $f_1(a_2) \in (0, \sqrt{\frac{3-\sqrt{5}}{2}})$  gegen Null bzw. für  $f_1(a_2) \in (\sqrt{\frac{3-\sqrt{5}}{2}}, 1)$  gegen Eins, und die Teilfolge  $(v_1^{(n)})_{n \in 4N}$  für  $f_1(f_2(f_1(a_2))) \in (0, \sqrt{\frac{3-\sqrt{5}}{2}})$  gegen Null bzw. für  $f_1(f_2(f_1(a_2))) \in (\sqrt{\frac{3-\sqrt{5}}{2}}, 1)$  gegen Eins.

Für welche anfänglichen Nutzenbewertungen der ersten Person  $a_1$  ist die obige Bedingung (\*) erfüllt?

$$f_1(f_2(f_1(f_2(f_1(f_2(a_1))))) < f_1(f_2(a_1)) \text{ gdw.}$$

$$F(1-a_1^2) < 1-a_1^2 \text{ gdw. } a_1 \in (0,1) \wedge (1-a_1^2)^3 - 2(1-a_1^2) + 1 > 0$$

$$\text{gdw. } 1-3a_1^2+3a_1^4-a_1^6-2+2a_1^2+1 > 0$$

$$-a_1^6+3a_1^4-a_1^2 > 0$$

$$(a_1^2)^3-3(a_1^2)^2-a_1 < 0$$

$$\text{Für } a_1 \in (0,1) \text{ ist } (a_1^2)^3-3(a_1^2)^2-a_1 = 0 \text{ nur für } a_1 = \sqrt{\frac{3-\sqrt{5}}{2}}.$$

Damit gilt für  $a_1 \in (0, \sqrt{\frac{3-\sqrt{5}}{2}})$  :

Die Teilfolge  $(v_1^{(n)})_{4N+1}$  fällt monoton und strebt gegen Null, während die Teilfolge  $(v_1^{(n)})_{4N_0+3}$  monoton steigt und gegen Eins strebt.

Andererseits gilt für  $a_1 \in (\sqrt{\frac{3-\sqrt{5}}{2}}, 1)$  :

Die Teilfolge  $(v_1^{(n)})_{4N+1}$  steigt monoton und strebt gegen Eins, während die Teilfolge  $(v_1^{(n)})_{4N_0+3}$  monoton fällt und gegen Null strebt.

Die intrapersonelle Divergenz der Nutzenbewertungen eines Zustandes bei iterierten Offenlegungen ist also nicht ein Merkmal speziell gewählter Anfangsbewertungen. In diesem Beispiel gibt es keine einzige Gleichgewichtsbewertung für beide Personen, d.h. mindestens eine Person hat bei dieser Interdependenz keine bei Iteration stabile Bewertung. Stabile Bewertungen gibt es nur bezüglich bestimmter Teilfolgen.

Denn für eine für beide Personen stabile Bewertung müßte gelten:

$$(1) a_1 = f_1(a_2) = 1-a_2$$

$$(2) a_2 = f_2(a_1) = a_1^2$$

$$(3) a_1 = f_1(f_2(a_1)) = 1-a_1^2$$

$$(4) a_2 = f_2(f_1(a_2)) = (1-a_2)^2$$

---

<sup>1</sup>  $[a_1^0, a_2^0]$  stellt eine Gleichgewichtsbewertung für die erste Person dar  $\Leftrightarrow \forall n \in \mathbb{N} : v_1^{(n)} = a_1^0$ . Im allgemeinen ist jedoch eine Gleichgewichtsbewertung für die Nutzenbewertungsfolge der ersten Person nicht zugleich eine Gleichgewichtsbewertung für die Nutzenbewertungsfolge der zweiten Person.

$$(3) \Rightarrow a_1 = \frac{-1+\sqrt{5}}{2} \vee a_1 = \frac{-1-\sqrt{5}}{2};$$

$$(4) \Rightarrow a_2 = \frac{3+\sqrt{5}}{2} \quad a_2 = \frac{3-\sqrt{5}}{2};$$

Von den vier möglichen Bewertungskombinationen ist nur  $a_1, a_2'$  mit (1) vereinbar. Aber auch  $a_1, a_2'$  ist nicht mit (2) vereinbar.

Man kann auch den weniger anschaulichen, aber mathematisch eleganteren Weg wählen, und jeweils für eine gegebene Nutzeninterdependenz die damit vereinbaren 'Gleichgewichtspunkte' von Bewertungskombinationen bestimmen. Die Tatsache, daß aufgrund einer bestimmten Interdependenz nur einzelne oder sogar keine Bewertungskombinationen möglich sind, spricht in gleicher Weise gegen einen uniformen Präferenzbegriff wie die Existenz von seltsamen Grenzwerten und intrapersonell divergenten Bewertungsfolgen.



ANHANG: DIE AGGREGATION INDIVIDUELLER URTEILE  
ZU EINEM GRUPPENURTEIL

Man könnte die Hoffnung haben, daß sich auf beiden Ebenen - der der Urteils- und der der Interessenaggregation - die gleichen 'Aggregationsregeln' als adäquat erweisen. Tatsächlich kann es zweckmäßig sein, sich für den Fall eines theoretischen Dissens Kriterien zu überlegen, an die sich eine Gruppe von Personen halten kann, wenn sie zu einem kollektiven Urteil gezwungen ist. Damit ist der theoretische Dissens natürlich nicht behoben, das Gruppenurteil nicht als richtig erwiesen. Bei genauerer Betrachtung zeigt es sich allerdings, daß die Aggregationsregeln im theoretischen Bereich ganz anderen Kriterien entsprechen sollten als im praktischen Bereich.

Im folgenden gehen wir auf die Aggregationsproblematik im theoretischen Bereich kurz ein, um die Unterschiede über das methodologische Argument hinaus durch die Beurteilung von konkreten Aggregationsregeln zu verdeutlichen.

Obwohl sich einige empirische Untersuchungen mit der Thematik der 'Aggregation' individueller Urteile beschäftigt haben,<sup>1</sup> beschränken wir uns dabei auf eine logische Analyse.

Wir gehen dabei von einem äußerst einfachen Modell aus: Meinungsänderungen aufgrund von Kommunikation und interpersonell verschiedene Irrtumswahrscheinlichkeiten bleiben unberücksichtigt.<sup>2</sup>

---

<sup>1</sup> Vgl. z.B. R.L. Thorndike, *The effect of discussion upon the correctness decisions, when the factor of majority influence is allowed for*, Journal of social psychology 9 (1938) 343-363; D.C. Barnlund, *A comparative study of individual, majority, and group judgement*, Journal of abnormal and social psychology 58 (1959) 55-60.

<sup>2</sup> Keith Lehrer hat sich in den 70er Jahren eingehend mit dem Prozeß der sozialen Konsensusbildung befaßt. Wir machen hier von seinen mathematischen Modellen keinen Gebrauch, da ihr methodischer

Stellen wir uns zunächst vor, eine Gruppe von Personen  $i$  (aus  $K$ ) stehe angesichts einer Reihe von Alternativen vor der Notwendigkeit einer kollektiven Entscheidung, stellen wir uns nun weiter vor, die Personen seien alle- samt vollkommen moralisch, nicht in dem Sinne, daß sie im Besitz der richtigen normativ-ethischen Theorie sind, sondern in dem Sinne, daß ihre individuellen Präferenzurteile ausschließlich aufgrund ihrer jeweiligen moralischen Überzeugung, welche kollektive Entscheidung die für die Gruppe richtige sei, erfolge. Dies ist ohne Zweifel ein Grenzfall der normalen Ausgangssituation normativer Sozialwahltheorie. In diesem Fall geht es aber interes- santerweise ausschließlich um die Frage: Wie komme ich von individuellen Urteilen, die sich mit einer gewissen Wahrscheinlichkeit irren, zu einem guten Gruppenurteil? Diese Frage stellt sich für normative Urteile in der gleichen Weise wie für deskriptive und sie hat mit der gerechten Aggregation individueller Interessen nichts zu tun.

Die individuellen Urteile seien 'binär', d.h. so zer- legt, daß sie für eine endliche Anzahl von Propositionen in "Ja"/-"Nein"-Antworten dargestellt werden können.<sup>1</sup> Die Aggregation individueller Urteile  $d_i$  zu einem Gruppenurteil  $D$  erfolge aufgrund einer Regel kollektiven Ur- teils  $RkU$ .  $RkU$  ist eine Funktion, die jeder Urteils-

---

Status ungeklärt ist und sie sich deshalb zur Klärung der Unterschiede beider Aggregationstypen nicht eignen.

Vgl. K. Lehrer, *Social consensus and rational agniology*, Syn- these 31 (1975) 141-160; ders., *Consensus and comparison: A theory of social rationality*, in: Hooker/Leych/McClennen, Bd.1 (1978) 283-309; ders., *When rational disagreement is impossible*, *Nous* 10 (1976) 327-332 und ders., *Rational consensus in science and society*, London 1981.

<sup>1</sup> Die Aggregation individueller moralischer Präferenzordnungen würde in diesem Modell keine Schwierigkeit machen. Sind z.B. in einer Gruppe von 6 Personen 4 der Ansicht, daß die gesellschaftliche Situation  $s$  besser ist als  $s'$  und wird die Mehrheitsregel  $M$  zur Bestimmung des Gruppenurteils verwandt, ist die Meinung der Gruppe, daß  $s$  besser als  $s'$  ist. Mit  $p_i = "$  $s$  ist sozialetisch besser als  $s'$ ", ordnet die  $RkU$   $M$  der Urteilsstruktur  $d(p) = \langle d_i(p) \rangle = \langle w, w, f, w, f, w \rangle$  das kollektive Urteil  $w(p)$  zu:  $M: \langle d_i(p) \rangle = \langle w, w, f, w, f, w \rangle \rightarrow w(p)$ .

struktur  $d$  in  $K$  ein ('binäres') Gruppenurteil  $D(K)$  zugeordnet. Die individuellen Urteile der gleichen Urteilsstruktur beziehen sich auf jeweils eine einzige (interpersonell invariante) Proposition ebenso wie das zugeordnete Gruppenurteil.

$RkU: \{w, f\}^{\#K} \ni \langle d_i(p) \rangle \mapsto D(p) \in \{w, f, u\}$ .

Das Gruppenurteil kann drei Werte ("ja": $w$ ; "nein": $f$ ; "unbestimmt": $u$ ) annehmen, während die individuellen Urteile jeweils bestimmt sein sollen ("ja": $w$ ; "nein": $f$ ). Es soll nämlich nicht vorausgesetzt werden, daß es bezüglich jeder  $RkU$  zu einer beliebigen Urteilsstruktur ein Gruppenurteil gibt. Manche Regeln kollektiven Urteils mögen nur unter bestimmten Bedingungen, die von der Urteilsstruktur erfüllt sein müssen, ein kollektives Urteil erlauben. Um das in unserem Modell zu erfassen, ordnen wir solchen Urteilsstrukturen, denen bezüglich einer Regel kollektiven Urteils kein kollektives Urteil zugeordnet wird, den Wert  $u$  zu, so daß trotzdem jede  $RkU$  eine Funktion ist. Nur wenn  $RkU(d) \neq u$ , sprechen wir von einem Gruppenurteil oder kollektiven Urteil.

Sei  $\rho(d_i)$  die Wahrscheinlichkeit dafür, daß das individuelle Urteil der Person  $i$  richtig ist. Nehmen wir weiter, stark vereinfachend, an, wir hätten keinen Grund für unterschiedliche Richtigkeitswahrscheinlichkeiten a priori:  $\forall i, j \in K : \rho(d_i) = \rho(d_j) = \rho_0$ . Je nach Wahl der Regel kollektiven Urteils ergeben sich danach unterschiedliche Wahrscheinlichkeiten für die Richtigkeit des Gruppenurteils  $D$  bezüglich der Wahrscheinlichkeit  $\rho_0$  für die Richtigkeit der individuellen Urteile.

$\rho_D (= \rho(D))$  ist Funktion von  $RkU$  und  $\rho_0$  (bzw. allgemein von  $\rho_i$ ). Sucht man nun nach einem Bewertungskriterium für die unterschiedlichen Aggregationsregeln im theoretischen Bereich, ist es naheliegend, die Irrtumswahrscheinlichkeit des Gruppenurteils mit der Irrtumswahrscheinlichkeit der individuellen Urteile zu vergleichen, was schon im Ansatz zu völlig anderen Ergebnissen führen muß als die Anwendung sozialethischer Anonymitäts-, Symmetrie-, Demokratie- und Effizienzkriterien.

Als ein allgemeines Gütekriterium für eine Regel kollektiven Urteils bietet sich an:

$$G(X) = \int_0^1 (\rho_D(X, \rho_0) - \rho_0) \cdot d\rho_0$$

und als spezielles Gütekriterium für eine gegebene Wahrscheinlichkeit  $\rho_0$ :

$$G_s(X, \rho_0) = \rho_D(X, \rho_0) - \rho_0.$$

Um die Güte einer Regel kollektiven Urteils für unterschiedliche individuelle Irrtumswahrscheinlichkeiten

vergleichen zu können, führen wir noch das relative spezifische Gütekriterium  $G_{s/rel}(X, \rho_0) = \frac{\rho_D(X, \rho_0) - \rho_0}{(1 - \rho_D) \cdot \rho_0}$  ein.

Der Übergang vom spezifischen zum relativen spezifischen Gütekriterium hat folgenden Grund:

Die absolute Veränderung der Wahrscheinlichkeiten durch die Anwendung einer RkU wäre ein verzerrter Maßstab. Eine Verbesserung der Wahrheitswahrscheinlichkeit um 0,09 von  $\rho_0 = 0,9$  auf  $\rho_D = 0,99$  stellt eine weit größere Leistung einer Regel kollektiven Urteils dar als eine Verbesserung der Wahrheitswahrscheinlichkeit um 0,1 von  $\rho_0 = 0,4$  auf  $\rho_D = 0,5$ . Ebenso stellt aber auch eine Verbesserung der Wahrheitswahrscheinlichkeit um 0,09 von  $\rho_0 = 0,01$  auf  $\rho_D = 0,1$  eine größere Leistung dar als eine Verbesserung der Wahrheitswahrscheinlichkeit um 0,1 von  $\rho_0 = 0,4$  auf  $\rho_D = 0,5$ . Dieser Überlegung werden die Gütekriterien  $G_{s/rel}$  und  $\bar{G}_{s/rel}$  gerecht.

Stellen wir die Ergebnisse der beiden Kriterien für diese drei Beispiele gegenüber:

$$\text{Für } \rho_0 = 0,9 \text{ und } \rho_D = 0,99 : G_{s/rel} = 10; \quad G_s = 0,09$$

$$\text{Für } \rho_0 = 0,4 \text{ und } \rho_D = 0,5 : G_{s/rel} = 0,5; \quad G_s = 0,1$$

$$\text{Für } \rho_0 = 0,01 \text{ und } \rho_D = 0,1 : G_{s/rel} = 10; \quad G_s = 0,09.$$

Eine Verbesserung der Wahrheitswahrscheinlichkeit von 0,4 auf 0,5 wird vom relativen spezifischen Gütekriterium nur halb so hoch bewertet wie eine Verbesserung von 0,9 auf 0,99 und von 0,01 auf 0,1 (bei der Anwendung einer bestimmten RkU), während das spezifische Gütekriterium die Veränderung von 0,4 auf 0,5 am höchsten bewertet.

Während die Wahrscheinlichkeit  $\bar{\rho}_0$  dafür, daß eine bestimmte Person  $i$  aus  $K$  ein falsches Urteil fällt, gleich  $1 - \rho_0$  ist, gilt nicht,  $\bar{\rho}_D = 1 - \rho_D$ , da ja nicht jeder Urteilsstruktur von  $K$  bei der Anwendung einer beliebigen Regel kollektiven Urteils  $X$  ein Gruppenurteil  $D$  zugeordnet wird. Die Wahrscheinlichkeit dafür, daß es aufgrund einer Regel kollektiven Urteils bei gegebenem  $\rho_0$  zu keinem Gruppenurteil kommt, ist  $\dot{\rho}_D(\rho_0) = 1 - (\rho_D(\rho_0) + \bar{\rho}_D(\rho_0))$ .

Analog wäre ein Gütekriterium für die Vermeidung falscher kollektiver Urteile:

$$\bar{G}(X) = \int_0^1 (\bar{\rho}_0 - \bar{\rho}_D) d\rho_0,$$

das dazugehörige spezifische Gütekriterium

$$\bar{G}_s(X, \rho_0) = \bar{\rho}_0 - \bar{\rho}_D(X, \rho_0)$$

und das relative spezifische Gütekriterium

$$\bar{G}_{s/rel}(X, \rho_0) = \frac{\bar{\rho}_0 - \bar{\rho}_D(X, \rho_0)}{\bar{\rho}_D(1 - \bar{\rho}_0)} .$$

Alle bisher genannten Kriterien sind epistemisch orientiert; sie berücksichtigen nur Glaubensgrade und nicht die Folgen falscher Vermutungen.

Geht es um die Auswahl einer Regel kollektiver Wahl in einer Entscheidungssituation, in der die Nachteile eines falschen Gruppenurteils, die Vorteile eines richtigen und die Konsequenzen eines nicht erfolgten kollektiven Urteils abgeschätzt werden können, so bietet sich als Gütekriterium einer Regel kollektiven Urteils der Nutzen-erwartungswert von  $X$  in dieser Entscheidungssituation an.

$$U(X) = \int_0^1 (R \cdot \rho_D(X, \rho_0) + F \cdot \bar{\rho}_D(X, \rho_0) + I \cdot \dot{\rho}_D(X, \rho_0)) d\rho_0,$$

wobei  $R$  der Nutzen eines richtigen kollektiven Urteils,  $F$  der Nutzen (bzw. Schaden) eines falschen und  $I$  der eines nicht erfolgten kollektiven Urteils ist.

Ist  $\rho_0$  bekannt, bietet sich das Gütekriterium

$$U_S(X, \rho_0) = R \cdot \rho_D(X, \rho_0) + F \cdot \bar{\rho}_D(X, \rho_0) + I \cdot \dot{\rho}_0(X, \rho_0) \text{ an.}^1$$

Damit läßt sich für eine gegebene individuelle in K konstante Irrtumswahrscheinlichkeit  $\bar{\rho}_0$  diejenige Regel kollektiven Urteils zur Grundlage des Gruppenurteils machen, die in der jeweiligen Entscheidungssituation - F, R, I sind ja Funktionen der Entscheidungssituationen - den höchsten Wert von  $U_S(X, \rho_0)$  hat.

Dieses Modell ist bei weitem nicht differenziert genug, um in allen Situationen die Anwendung bestimmter Regeln kollektiven Urteils begründen zu können, immerhin aber kann es die Anwendung einzelner RkUs, wie sie in vielen Organisationen üblich ist, plausibel machen. Es macht u.a. deutlich, daß pauschale Urteile darüber, welche Regeln kollektiven Urteils allgemein vorzuziehen sind, auf einer mangelnden Differenzierung in der Analyse der Entscheidungssituationen beruhen.<sup>2</sup>

An einigen Aggregationsregeln, deren Analoga auch in der sozialwahltheoretischen Literatur diskutiert werden, wollen wir die Brauchbarkeit der vorgeschlagenen Gütekriterien prüfen, um damit zu verdeutlichen, wie notwendig die Trennung der theoretischen von der praktischen Ebene der Aggregationsproblematik ist.

#### 1. Beispiel: Meinungsdiktatur

Bei der Meinungsdiktatur lautet die Regel kollektiven Urteils A :  $D = d_i$ , i fest und beliebig. Hier gilt:

---

<sup>1</sup>  $U(X)$  und  $U_S(X, \rho_0)$  nennen wir im Gegensatz zu  $G(X)$ ,  $G_S(X, \rho_0)$  und  $G_{s/rel}(X, \rho_0)$  entscheidungsabhängige Kriterien.

<sup>2</sup> Wie z.B. "(...) the best kind of group decision is the ordinary majority vote, at least in circumstances in which each member of the group is as likely as another to make the "correct" response." Vgl. N. Smoke/R. Zajonc, *On the reliability of group judgements and decisions*, in: Criswell/Solomon/Suppes (1962) 322-333, S.333. Hier unterbleibt in zweierlei Hinsicht eine notwendige Differenzierung: Interessen- und Urteilsaggregation wird nicht unterschieden, und die Relevanz der Entscheidungssituationen für die Auswahl der besten Regel kollektiven Urteils bleibt unberücksichtigt.

$$\rho_D(A, \rho_0) = \rho_0 \text{ für } \rho_0 \in [0, 1] ;$$

$$G(A) = 0;$$

$$G_S(A, \rho_0) = \rho_D(A, \rho_0) - \rho_0 = 0, \text{ für alle } \rho_0$$

$$G_{S/rel}(A, \rho_0) = 0;$$

$$\bar{G}(A) = 0;$$

$$\bar{G}_S(A, \rho_0) = 0;$$

$$\bar{G}_{S/rel}(A, \rho_0) = 0.$$

## 2. Beispiel: Meinungsanarchie

Es kommt nur dann zu einem Gruppenurteil D, wenn alle darin übereinstimmen.

$$B(d) = D \Leftrightarrow \forall i \in K : d_i = D;$$

Ist die Wahrscheinlichkeit dafür, daß eine Person i aus K mit ihrem Urteil recht hat,  $\rho_0$ , dann ist die Wahrscheinlichkeit dafür, daß das aufgrund der Einstimmigkeitsregel B erfolgte Gruppenurteil D richtig ist  $\rho_D(B, \rho_0) = \rho_0^n$ , mit  $\#K=n$ .

$$\begin{aligned} G(B) &= \int_0^1 (\rho_0^n - \rho_0) d\rho_0 = \left[ \frac{1}{n+1} \rho_0^{n+1} - \frac{1}{2} \rho_0^2 \right] \\ &= \frac{1}{n+1} - \frac{1}{2}; \end{aligned}$$

d.h. für  $\#K > 2$  liefert unser Gütekriterium für die Einstimmigkeitsregel kollektiven Urteils ein negatives Ergebnis. M.a.W., die Wahrscheinlichkeit mit dem eigenen Urteil richtig zu liegen, ist größer als die eines ausschließlich durch allgemeine Übereinstimmung erfolgten Gruppenurteils (und zwar für beliebige Werte von  $\rho_0$ ). Das heißt allerdings nicht, daß die Einstimmigkeitsregel für jede Situation, in der ein Gruppenurteil getroffen werden soll, unzumutbar ist. Die Wahrscheinlichkeit dafür, daß ein aufgrund der Einstimmigkeitsregel getroffenes Urteil falsch ist, ist ebenfalls klein:  $\bar{\rho}_D(B, \rho_0) = (1 - \rho_0)^n$ ; ist man also primär daran interessiert, ein falsches Gruppenurteil zu vermeiden, so bietet sich die Einstimmigkeitsregel besonders an.

$$\begin{aligned}\bar{G}(B) &= \int_0^1 (\bar{\rho}_0 - \bar{\rho}_D) \cdot d\rho_0 = \int_0^1 ((1-\rho_0) - (1-\rho_0)^n) \cdot d\rho_0 = \\ &= \left[ \rho_0 - \frac{1}{2}\rho_0^2 - \frac{1}{n+1} (1-\rho_0)^{n+1} (-1) \right]_0^1 = \\ &= \frac{1}{2} - \frac{1}{n+1} ;\end{aligned}$$

Das ausschließlich epistemisch orientierte Gütekriterium  $\bar{G}$  zur Vermeidung falscher Gruppenurteile ist damit positiv für  $\#K > 2$  (und wächst monoton mit  $n$ ).

Die absolute Veränderung der Irrtumswahrscheinlichkeit des kollektiven Urteils gegenüber der Irrtumswahrscheinlichkeit des individuellen Urteils zeigt das Gütekriterium  $G_S$  bzw.  $\bar{G}_S$  :

$$G_S(B, \rho_0) = \rho_D(B, \rho_0) - \rho_0 = \rho_0^n - \rho_0 ;$$

$$\bar{G}_S(B, \rho_0) = \bar{\rho}_0 - \bar{\rho}_D(B, \rho_0) = 1 - \rho_0 - (1-\rho_0)^n ;$$

- Für  $\#K=n=10$  und  $\rho_0=0,9$  :

$$G_S(B, 0,9) = -0,551$$

$$\bar{G}_S(B, 0,9) = 0,1 - 10^{-10} \approx 0,1$$

- Für  $n=10$  und  $\rho_0 = 0,5$  :

$$G_S(B, 0,5) = 9,8 \cdot 10^{-4} - 0,5 \approx -0,499$$

$$\bar{G}_S(B, 0,5) = 0,5 - 9,8 \cdot 10^{-4} \approx 0,499$$

- Für  $n=10$  und  $\rho_0 = 0,1$  :

$$G_S(B, 0,1) = 10^{-10} - 0,1 \approx -0,1$$

$$\bar{G}_S(B, 0,1) = 0,9 - 10^{-10} \approx 0,9$$

Diese Beispiele zeigen durchgängig ein Charakteristikum der Einstimmigkeitsregel: Für die Verringerung der Irrtumswahrscheinlichkeit positive Werte, für die Verbesserung der Wahrheitswahrscheinlichkeit negative Werte.

Die Anwendung der relativen spezifischen Gütekriterien geben eine bessere Vergleichsgrundlage ab. Damit ergibt sich für eine anarchistische Regel kollektiven Urteils



folgende Bewertung in Abhängigkeit von der individuellen Irrtumswahrscheinlichkeit:

$$G_{s/rel}(B, \rho_0) = \frac{\rho_D(B, \rho_0) - \rho_0}{(1 - \rho_D) \cdot \rho_0} = \frac{\rho_0^n - \rho_0}{(1 - \rho_0^n) \cdot \rho_0} ;$$
$$\bar{G}_{s/rel}(B, \rho_0) = \frac{\bar{\rho}_D - \bar{\rho}_D(B, \rho_0)}{\bar{\rho}_D \cdot (1 - \bar{\rho}_0)} = \frac{1 - \rho_0 - (1 - \rho_0)^n}{(1 - \rho_0)^n \cdot \rho_0} ;$$

Angewendet auf drei Beispiele für eine Gruppe von jeweils 10 Personen:

(1)  $G_s(B, 0,9) = -0,551$   
 $G_{s/rel}(B, 0,9) = -0,941$   
 $\bar{G}_s(B, 0,9) = 0,1$   
 $\bar{G}_{s/rel}(B, 0,9) = 1,1 \cdot 10^9 ;$

(2)  $G_s(B, 0,5) = -0,499$   
 $G_{s/rel}(B, 0,5) = -0,998$   
 $\bar{G}_s(B, 0,5) = 0,449$   
 $\bar{G}_{s/rel}(B, 0,5) = 1022 ;$

(3)  $G_s(B, 0,1) = 0,1$   
 $G_{s/rel}(B, 0,1) = -1$   
 $\bar{G}_s(B, 0,1) = 0,9$   
 $\bar{G}_{s/rel}(B, 0,1) = 15,8 ;$

Ist die Wahrscheinlichkeit richtiger individueller Urteile hoch, so liefert uns das relative spezifische Gütekriterium  $\bar{G}_{s/rel}$  bei der Einstimmigkeitsregel extrem hohe Werte. Aber auch für größere individuelle Irrtumswahrscheinlichkeiten schneidet die Einstimmigkeitsregel unter dem Kriterium  $\bar{G}_{s/rel}$  sehr gut ab (ganz im Gegensatz zu den Ergebnissen bei dem Kriterium  $G_{s/rel}$ ).

### 3. Beispiel: Meinungsoligarchie

Erfolgt das Gruppenurteil D aufgrund der Übereinstimmung einer festen Gruppe  $L \subset K$ ,  $C(d) = D \Leftrightarrow \forall i \in L : d_i = D$ , so gilt analog zu der Einstimmigkeitsregel kollektiven Urteils

$$\rho_D(\rho_0) = \rho_0^m \text{ mit } \#L=m < \#K=n,$$

und aus dem Gütekriterium G ergibt sich

$$G(C, \rho_0) = \int_0^1 (\rho_0^m - \rho_0) d\rho_0 = \frac{1}{m+1} - \frac{1}{2}.$$

Die weitere Diskussion der Meinungsoligarchie erübrigt sich, da sie sich aus dem letzten Beispiel ergibt, wenn man jeweils "n" durch "m" ersetzt.

### 4. Beispiel: Mindestzahlregel ("Quorum")

Erfolgt das Gruppenurteil aufgrund einer Übereinstimmung von mindestens m Personen in der Gruppe K,  $\#K=n > m$ ,  $M(d) = D \Leftrightarrow \exists k \in \{1, 2, \dots, n\} \exists \sigma \in \text{Bij}(\{1, 2, \dots, n\}) :$

$$\forall \sigma(i) \leq k : d_i = D.$$

Damit ist die Wahrscheinlichkeit dafür, daß aufgrund der Mindestzahlregel eine richtige Gruppenentscheidung D erfolgt,

$$\rho_D(d) = \sum_{i=m}^n \binom{n}{i} \rho_0^i \cdot (1-\rho_0)^{n-i};$$

die Wahrscheinlichkeit dafür, daß aufgrund von M ein falsches Gruppenurteil (bei der Urteilsstruktur d in der Gruppe K) erfolgt,

$$\bar{\rho}_D(d) = \sum_{i=m}^n \binom{n}{i} (1-\rho_0)^i \cdot \rho_0^{n-i}$$

und die Wahrscheinlichkeit dafür, daß aufgrund der Mindestzahlregel keine Gruppenentscheidung erfolgt,

$$\hat{\rho}(d) = 1 - (\rho_D(d) + \bar{\rho}_D(d)).$$

Das allgemeine Gütekriterium ergibt, angewendet auf die Mindestzahlregel kollektiven Urteils (die im Fall  $m > \frac{1}{2}n$  eine (qualifizierte) Mehrheitswahlregel darstellt) :

$$G(M) = \int_0^1 (\rho_D(M, \rho_0) - \rho_0) d\rho_0 =$$

$$= \int_0^1 \left( \sum_{i=m}^n \binom{n}{i} \rho_0^i (1-\rho_0)^{n-i} - \rho_0 \right) d\rho_0;$$

beispielsweise für #K=n=6 und m=2 :

$$G(M) = \int_0^1 \left( \sum_{i=2}^6 \binom{6}{i} \rho_0^i (1-\rho_0)^{6-i} - \rho_0 \right) d\rho_0 =$$

$$= \int_0^1 (15\rho_0^2 (1-\rho_0)^4 + 20\rho_0^3 (1-\rho_0)^3 + 15\rho_0^4 (1-\rho_0)^2 +$$

$$+ 6\rho_0^5 (1-\rho_0) + \rho_0^6 - \rho_0) d\rho_0 =$$

$$= \int_0^1 (15(\rho_0^2 - 4\rho_0^3 + 6\rho_0^4 - 4\rho_0^5 + \rho_0^6) +$$

$$+ 20(\rho_0^3 - 3\rho_0^4 + 3\rho_0^5 - \rho_0^6) +$$

$$+ 15(\rho_0^4 - 2\rho_0^5 + \rho_0^6) +$$

$$+ 6(\rho_0^5 - \rho_0^6) + \rho_0^6 - \rho_0) \cdot d\rho_0$$

$$= \left[ \frac{15}{3} \rho_0^3 \right]_0^1 - \left[ \frac{40}{4} \rho_0^4 \right]_0^1 + \left[ \frac{45}{5} \rho_0^5 \right]_0^1 - \left[ \frac{24}{6} \rho_0^6 \right]_0^1 +$$

$$\left[ \frac{5}{7} \rho_0^7 \right]_0^1 - \left[ \frac{1}{2} \rho_0^2 \right]_0^1 = 0,214.$$

Das allgemeine Gütekriterium für die Vermeidung falscher kollektiver Urteile ergibt:

$$G(M) = \int_0^1 (\bar{\rho}_0 - \bar{\rho}_D(M, \rho_0)) d\rho_0 =$$

$$= \int_0^1 \left( 1 - \rho_0 - \sum_{i=m}^n \binom{n}{i} (1-\rho_0)^i \rho_0^{n-i} \right) d\rho_0 ;$$

für n=6 und m=2 :  $\bar{G}(M) \cong -0,214.$

<sup>1</sup> n=6 ∧ m=2 :

$$\bar{G}(M) = \int_0^1 (\bar{\rho}_0 - \bar{\rho}_D(M, \rho_0)) d\rho_0 = \int_0^1 (1 - \rho_0 - \sum_{i=2}^6 \binom{6}{i} (1-\rho_0)^i \cdot \rho_0^{6-i}) d\rho_0 =$$

$$= \int_0^1 (1 - \rho_0 - [15(1-\rho_0)^2 \cdot \rho_0^4 + 20(1-\rho_0)^3 \cdot \rho_0^3 + 15(1-\rho_0)^4 \cdot \rho_0^2 +$$

$$+ 6(1-\rho_0)^5 \cdot \rho_0 + 1(1-\rho_0)^6 \cdot 1]) d\rho_0 =$$

$$= \int_0^1 (-\rho_0 + 6\rho_0^5 - 5\rho_0^6) d\rho_0 =$$

$$= \left[ -\frac{1}{2} \rho_0^2 \right]_0^1 + \left[ \frac{6}{6} \rho_0^6 \right]_0^1 - \left[ \frac{5}{7} \rho_0^7 \right]_0^1 \cong -0,214$$

Sind die individuellen Irrtumswahrscheinlichkeiten bekannt, kann man die relativen spezifischen Gütekriterien  $g_{s/rel}$  und  $\bar{G}_{s/rel}$  zur Beurteilung heranziehen. Wir nehmen den Fall  $n=5$  und  $m=3$ , womit wir zugleich die Mehrheitswahl als Regel kollektiver Urteilsfindung in diesem einfachen Beispiel beurteilen.

$$\begin{aligned}
 G_{s/rel}(M, \rho_0) &= \frac{\rho_D(M, \rho_0) - \rho_0}{(1 - \rho_D) \cdot \rho_0} = \frac{\sum_{i=m}^n \binom{n}{i} \rho_0^i (1 - \rho_0)^{n-i} - \rho_0}{(1 - \sum_{i=m}^n \binom{n}{i} \rho_0^i (1 - \rho_0)^{n-i}) \cdot \rho_0} = \\
 &= \frac{\binom{5}{3} \rho_0^3 (1 - \rho_0)^2 + \binom{5}{4} \rho_0^4 (1 - \rho_0) + \binom{5}{5} \rho_0^5 \cdot 1 - \rho_0}{(1 - [\binom{5}{3} \rho_0^3 (1 - \rho_0)^2 + \binom{5}{4} \rho_0^4 (1 - \rho_0) + \binom{5}{5} \rho_0^5]) \cdot \rho_0} = \\
 &= \frac{10(\rho_0^3 - 2\rho_0^4 + \rho_0^5) + 5(\rho_0^4 - \rho_0^5) + \rho_0^5 - \rho_0}{(1 - 10(\rho_0^3 - 2\rho_0^4 + \rho_0^5) - 5(\rho_0^4 - \rho_0^5) - \rho_0^5) \cdot \rho_0} = \\
 &= \frac{10\rho_0^3 - 15\rho_0^4 + 6\rho_0^5 - \rho_0}{\rho_0 - 10\rho_0^4 + 15\rho_0^5 - 6\rho_0^6} ; \\
 \bar{G}_{s/rel}(M, \rho_0) &= \frac{\bar{\rho}_0 - \bar{\rho}_D(M, \rho_0)}{\bar{\rho}_D(1 - \bar{\rho}_0)} = \frac{1 - \rho_0 - \sum_{i=m}^n \binom{n}{i} (1 - \rho_0)^i \cdot \rho_0^{n-i}}{(\sum_{i=m}^n \binom{n}{i} (1 - \rho_0)^i \cdot \rho_0^{n-i}) \cdot \rho_0} = \\
 &= \frac{1 - \rho_0 - [\binom{5}{3} (1 - \rho_0)^3 \cdot \rho_0^2 + \binom{5}{4} (1 - \rho_0)^4 \cdot \rho_0 + \binom{5}{5} (1 - \rho_0)^5]}{[\binom{5}{3} (1 - \rho_0)^3 \cdot \rho_0^2 + \binom{5}{4} (1 - \rho_0)^4 \cdot \rho_0 + \binom{5}{5} (1 - \rho_0)^5] \cdot \rho_0} = \\
 &= \frac{-\rho_0 + 10\rho_0^3 - 15\rho_0^4 + 6\rho_0^5}{\rho_0 - 10\rho_0^4 + 15\rho_0^5 - 6\rho_0^6} ;
 \end{aligned}$$

Bei kleiner Irrtumswahrscheinlichkeit, etwa  $\rho_0 = 0,9$  :

$$G_{s/rel}(M, 0,9) = \frac{0,09}{0,01} = 9 ;$$

bei mittlerer Irrtumswahrscheinlichkeit, etwa  $\rho_0 = 0,5$  :

$$G_{s/rel}(M, 0,5) = \frac{0}{0,249} = 0 ;$$

bei hoher Irrtumswahrscheinlichkeit, etwa  $\rho_0 = 0,1$  :

$$G_{s/rel}(M, 0,1) = \frac{-0,0915}{0,09915} = -0,92285 .$$

Die Mehrheitswahl ist also für niedrige individuelle Irrtumswahrscheinlichkeiten eine gute Regel kollektiven Urteils. Das gilt auch unter dem Aspekt der Vermeidung falscher kollektiver Urteile. In unserem Beispiel, also für eine Gruppe von fünf Personen, liefern die beiden relativen spezifischen Gütekriterien für beliebige individuelle Irrtumswahrscheinlichkeiten  $(1-\rho_0)$  jeweils gleiche Bewertungen der einfachen Mehrheitswahlregel:  
 $\forall \rho_0 \in [0,1] : G_{s/rel}(M, \rho_0) = \bar{G}_{s/rel}(M, \rho_0).$

Obwohl wir hier von einem äußerst simplen Ansatz Gebrauch gemacht haben, denke ich, ist deutlich geworden, daß die Aggregation individueller Urteile nicht mit der Aggregation individueller Interessen, dem normativen Problem speziell der Sozialwahltheorie, konfundiert werden darf. Regeln der Urteilsaggregation unterliegen ganz anderen Beurteilungskriterien als Regeln der Interessenaggregation. Nur letztere sind Gegenstand normativer Ethik.<sup>1</sup>

---

<sup>1</sup> Während Lehrer in der Tradition soziologischer Machtanalyse steht, wie sie unter anderen von John French entwickelt wurde (vgl. French (1956) und Thibaut/Kelly (1959)), steht Carl Wagners Modell der Konsensusbildung unter einem explizit normativen Aspekt, vgl. C. Wagner, *Consensus through respect: A model of rational group decision making*, *Philosophical studies* 34 (1978) 335-349, vgl. dazu a. Kreyszig (1978). Was diesen Ansatz interessant macht, ist unter anderem, daß er mit der häufig behaupteten Dichotomie kalkulatorischer Modelle der politischen Ökonomie und der englischsprachigen Moralphilosophie einerseits und kommunikativer Ansätze kontinentaler und speziell deutschsprachiger Diskussion normativer Problemstellungen andererseits über Kreuz liegt (vgl. etwa Höffe (1975), S. 201 ff).

## LITERATURVERZEICHNIS

- Acham, K.: Wissenschaftliche Politikberatung aus der Sicht der analytischen Philosophie. In: Maier/Ritter/Matz (1971) 53-138.
- Ackoff s. Minas/Ackoff
- Albert, H.: Ethik und Meta-Ethik. Das Dilemma der analytischen Moralphilosophie. In: Archiv für Philosophie 11 (1961) 28-63.
- Plädoyer für kritischen Rationalismus. München 1971.
- Konstruktion und Kritik. Hamburg 1972.
- Kritische Vernunft und menschliche Praxis. Stuttgart 1977.
- Albert, H./Topitsch, E. (Hrsg.): Werturteilsstreit. Darmstadt 1971.
- Aldrup, D.: Das Rationalitätsprinzip in der politischen Ökonomie. Methodenkritische Lösungsansätze. Tübingen 1971.
- Allais, M.: Le comportement de l'homme rationnel devant le risque. Critiques des postulats et axiomes de l'école américaine. In: Econometrica 21 (1953) 503-546.
- Allais, M./Hagen, O. (Hrsg.): Expected utility hypotheses and the Allais paradox. Dordrecht 1979.
- Allen, G.O.: Le volonté de tous and le volonté général. A distinction and its significance. In: Ethics 71 (1960/61) 263-274.
- Alt, F.: Über die Meßbarkeit des Nutzens. In: Zeitschrift für Nationalökonomie 7 (1936). Engl.: On the measurability of utility. In: Chipman u.a. (1971) 424-431.
- Anderson, A.: Logic, norms and rules. In: Criswell/Solomon/Suppes (1962) 11-23.
- Apel, K.-O.: Zum Problem einer rationalen Begründung der Ethik im Zeitalter der Wissenschaft. In: Riedel, Bd 2 (1972) 13-32.
- Archibald, G.C.: Welfare economics, ethics and essentialism. In: Economica 26 (1959) 316-327.
- Arkoff s. Vanacke/Arkoff
- Armbruster, W.: Spiele mit unbekanntem Nutzenfunktionen und ihre Gleichgewichtspunkte. Heidelberg 1978.
- Armbruster, W./Böge, W.: Bayesian game theory. In: Moeschlin/Pallaschke (1979) 17-28.
- Armstrong, W.E.: The determinateness of the utility function. In: Economic journal 49 (1939) 453-467.
- Utility and the theory of welfare. In: Oxford economics papers, New series 3 (1951) 259-271.
- Arnaszus, H.: Spieltheorie und Nutzenbegriff aus marxistischer Sicht. Eine Kritik aktueller ökonomischer Theorien. Frankfurt/M. 1974.

- Arrow, K.J.: Alternative approaches in the theory of choice in risk-taking situations. In: *Econometrica* 19 (1951) 404-437.
- Rational choice functions and orderings. In: *Econometrica* 26 (1959) 121-127.
- Social choice and individual values. New York, London, Sydney 1963.
- Exposition of the theory of choice under uncertainty. In: *Synthese* 16 (1966) 253-269.
- Public and private values. In: *Hook* (1967) 3-21. (1967a)
- Values and collective decision-making. In: *Laslett/Runciman*, Bd 3 (1967) 215-232. (1967b)
- Arrow, K.J./ Intriligator, M.P.: *Handbook of mathematical economics*, Bd 1. Amsterdam, New York, Oxford 1981.
- Arrow, K.J./ Karlin, S./ Suppes, P.: *Mathematical methods in the social sciences*. Proceedings of the 1<sup>st</sup> Stanford Symposium. Stanford 1966.
- d'Asprement, C./ Gevers, L.: Equity and the informational basis of collective choice. In: *Review of economic studies* 46 (1977) 199-210.
- d'Asprement, C./ Gevers, L.: Leximin and utilitarian rules. A joint characterization. In: *Journal of economic theory* 19 (1978) 143-163.
- Austin, J.L.: *How to do things with words* (1962). Dt. Bearb. von E. v. Savigny: *Zur Theorie der Sprechakte*. Stuttgart 1972.
- Bähr, J.K.: *Vorträge über Newtons und Goethes Farbenlehre*. Dresden 1863.
- Baier, K.: *The moral point of view. A rational basis of ethics*. New York 1958.
- Moral obligation. In: *American philosophical quarterly* 3 (1966) 210-226.
- Rationality and morality. In: *Erkenntnis* 11 (1977) 197-223.
- Maximization and the good life. In: *Morscher/Stranzinger* (1981) 33-42. (1981a)
- Defining morality without prejudice. In: *The monist* 64 (1981) 425-441. (1981b)
- Bar-Hillel, J. (Hrsg.): *Logic, methodology and philosophy of science*. Amsterdam 1965.
- Barnlund, D.C.: A comparative study of individual majority and group judgement. In: *Journal of abnormal and social psychology* 58 (1959) 55-60.
- Barry, B.: *The liberal theory of justice*. Oxford 1973.
- *Neue politische Ökonomie*. Frankfurt 1975.
- Don't shoot the trumpeter - he's doing his best. Reflections on a problem of fair division. In: *Theory and decision* 11 (1979) 153-180.

- Baumol, W.J.: The Neumann-Morgenstern utility index. An ordinalist view. In: Journal of political economy 59 (1951) 61-66.
- Bayes, Th.: An essay towards solving a problem in the doctrine of chances (1763). In: Biometrika 45 (1958) 296-315.
- Becker, G.M./ De Groot, M.H./ Marschak, J.: Measuring utility by a single-response sequential method. In: Behavioral science 9 (1964) 226-232.
- Beckerath, E.von/ Giersch, H. (Hrsg.): Problem der normativen Ökonomik und der wirtschaftlichen Beratung. Berlin 1963.
- Behrens, H.: Politische Entscheidungsprozesse. Opladen 1980.
- Bennett, J.: Rationalität. Versuch einer Analyse. Frankfurt/M. 1967.
- Bentham, J.: An introduction to the principles of morals and legislation (1789). Oxford 1907.
- Berg, E.: Democracy and self-determination. In: Birnbaum/Lively/ Geraint (1978) 149-172.
- Bergson, A.: A reformulation of certain aspects of welfare economics. In: Quarterly journal of economics 52 (1938) 310-334.
- Essays in normative economics. Cambridge/Mass. 1966.
- Bergström, L.: Reason in ethics. In: Ryle (1976) 177-187.
- Bernoulli, Daniel: Specimen theoriae novae de mensura sortis (1738). Dt. Übers. von A. Pringsheim: Versuch einer neuen Theorie der Wertbestimmung von Glücksfällen. Leipzig 1896.
- Bertrand, J.: Calcul de probabilités. Paris 1889.
- Birnbaum, P./ Lively, J./ Geraint, P. (Hrsg.): Democracy, consensus and social contract. London, Beverley Hills 1978.
- Bixenstine s. Wilson/Bixenstine
- Bjerring, A.K.: The 'tracing procedure' and a theory of rational interaction. In: Hooker/Leach/McClennen, Bd 1 (1978) 1-54.
- Black, D.: The theory of committees and elections. Cambridge 1938.
- Black, M. (Hrsg.): Problems of choice and decision. Ithaca/ N.Y., Aspen/Col. 1975.
- Blau, J.H.: Liberal values and independence. In: Review of economic studies 42 (1975) 395-402.
- Blau, U.: Die dreiwertige Logik der Sprache. Ihre Syntax, Semantik und Anwendung in der Sprachanalyse. Berlin, New York 1978.
- Blocker, H.G., Smith, E.H. (Hrsg.): John Rawls' theory of social justice. Athens/Ohio 1980.
- Bock, R.D./ Jones, L.V.: The measurement and prediction of judgment and choice. San Francisco 1968.
- Böge s. Armbruster/Böge
- Bohnen, A.: Die utilitaristische Ethik als Grundlage der modernen Wohlfahrtsökonomik. Göttingen 1964.



- Borch, K.: Ethics institutions and optimality. In: Gottinger/Leinfellner (1978) 237-249.
- Bordes, G.: Some more results on consistency, rationality and collective choice. In: Laffont (1979) 175-197.
- Boulding, K.: Some contributions of economics to the general theory of value. In: Philosophy of science 23 (1956) 1-14.
- Social justice in social dynamics. In: Brandt (1962) 73-92.
- Einführung in die Wohlfahrtsökonomie. In: Gäfgen (1966) 77-109.
- Ethics of the critique of preferences. In: Finnin/Smith (1979) 9-23.
- Braithwaite, R.B.: Why is it reasonable to base a betting rate upon an estimate of chance. In: Bar-Hillel (1965) 263-274.
- Theory of games as a tool for the moral philosopher. Cambridge 1969.
- Brams, S.J.: Game theory and politics. New York, London 1975.
- Brandt, R.B. (Hrsg.): Social justice. Englewood Cliffs 1962.
- Personal values and the justification of institutions. In: Hook (1967) 22-40.
- A theory of the good and the right. Oxford 1979.
- Bratman, M.: Intention and means-ends-rationality. In: Philosophical review 90 (1980) 252-265.
- Braungart, G.: Die Ethik und ihre außermoralische Rechtfertigung bei Hobbes. In: Ratio 31 (1981) 50-63.
- Braybrooke, D.: The choice between utilitarianisms. In: American philosophical quarterly 4 (1967) 28-38.
- Variety among hierarchies of preference. In: Hooker/Leach/McClennen, Bd 1 (1978) 55-65.
- Braybrooke, D./ Lindblom, D.: A strategy of decision. Policy evaluation as a social process. London 1963.
- Brehmer, K.: Rawls' 'original position' oder Kants 'ursprünglicher Kontrakt'. Königstein/Ts. 1980.
- Brenner, C.: Grundrisse der Psychoanalyse. Frankfurt/M. 1972.
- Brew, J.S.: An altruism parameter for prisoner's dilemma. In: Journal of conflict resolution 17 (1973) 351-367.
- Brock, H.W.: A new theory of social justice based on the mathematical theory of games. In: Ordeshook (1978) 563-627.
- Brock, H.W. (Hrsg.): Game theory, social choice, and ethics, Teil 1 u.2. In: Theory and decision 11 (1979) 143-337.
- A game theoretic account of social justice. In: Theory and decision 11 (1979) 239-266.
- Bryan s. Shelly/Bryan
- Buchanan, J.M.: Ethical rules, expected values and large numbers. In: Ethics 76 (1965/66) 1-13.

- Buchanan, J.M.: Freedom in constitutional contract. London 1977.
- From private preferences to public philosophy. The development of public choice. In: Buchanan u.a. (1978) 1-20.
- Buchanan, J.M./ Tullock, G.: The calculus of consent. Ann Arbor 1974.
- Buchanan, J.M. u.a.: The economics of politics. London 1978.
- Bühlmann, H./ Loeffel, H./ Nievergelt, E.: Einführung in die Theorie und Praxis der Entscheidung bei Unsicherheit. Berlin, Heidelberg, New York 1967 (Lecture notes in operations research and mathematical economics, 1).
- Entscheidungs- und Spieltheorie. Berlin, Heidelberg, New York 1975.
- Burger, E.: Einführung in die Theorie der Spiele. Berlin 1959.
- Burnet s. Platon
- Burns, T./ Meeker, D.: Structural properties and resolutions of the prisoners' dilemma game. In: Rapoport (1974) 35-62.
- Burros, R.H.: Complementary relations in the theory of preference. In: Theory and decision 7 (1976) 181-190.
- Bush s. Luce/Bush/Galanter
- Butler, J.: Fifteen sermons, preached at the Rolls Chapel (1726). London 1914.
- Butts s. Leach/Butts/Pearce
- Camacho, A.: Maximizing expected utility and the rule of long run success. In: Allais/Hagen (1979). (1979a)
- On cardinal utility. In: Theory and decision 10 (1979) 131-145. (1979b)
- Approaches to cardinal utility. In: Theory and decision 12 (1980) 359-379.
- Carnap, R.: Testability and meaning. In: Philosophy of science 3 (1936) 420-471; 4 (1937) 2-40.
- The methodological character of theoretical concepts. In: Minnesota studies in the philosophy of science 1 (1954) 38-76.
- Der logische Aufbau der Welt (1928). Hamburg 1961.
- Logical foundations of probability. Chicago 1962.
- Carnap, R./ Stegmüller, W.: Induktive Logik und Wahrscheinlichkeit. New York, Berlin 1958.
- Castaneda, H./ Nakhnikian, G. (Hrsg.): Morality and the language of conduct. Detroit 1963.
- Chamberlain, J.: A diagrammatic exposition of the logic of collective action. In: Hooker/Leach/McClennen, Bd 2 (1978) 17-27.
- Chammah s. Rapoport/Chammah
- Chapman, J.W.: Coercion in politics and strategy. In: Pennock/Chapman (1972) 289-319.
- Chernoff, H./ Moses, L.E.: Elementary decision theory. New York 1959.

- Chipman, J.S./ Hurwicz, L./ Richter, M.K./ Sonnenschein, H.F. (Hrsg.):  
Preferences, utility and demand. New York, Chicago, San Francisco,  
Atlanta 1971.
- Chisholm, R.M./ Sosa, E.: On the logic of 'intrinsically better'.  
In: American philosophical quarterly 3 (1966) 244-249.
- Chomsky, N.: American power and the new mandarines. Harmondsworth  
1969.
- Problems of knowledge and freedom. London 1972.
- Churchman, C.W.: Prediction and optimal decision. Philosophical  
issues of a science of values. Englewood Cliffs 1961.
- Coerner s. Krelle/Coerner
- Cohen, G.A.: Robert Nozick and Wilt Chamberlain: How patterns pre-  
serve liberty. In: Erkenntnis 11 (1977) 5-23.
- Coombs s. Thrall/Coombs/Davis
- Cramer, K.: Hypothetische Imperative? In: Riedel, Bd 1 (1972) 159-212.
- Criswell, J.H./ Solomon, H./ Suppes, P. (Hrsg.): Mathematical  
methods in small group processes. Stanford 1962.
- Craemer-Ruegenberg, I.: Über methodische Schwierigkeiten bei der  
Auslegung von moralischen Werturteilen. In: Riedel, Bd 1 (1972)  
133-158.
- Dahl, R.A./ Lindblom, D.: Politics, economics and welfare. New York  
1953.
- Daniel, T.: A revised concept of distributional equity. In: Journal  
of economic theory 11 (1975).
- Daniels, N. (Hrsg.): Reading Rawls. Critical studies of 'A theory  
of justice'. Oxford (o.J., ca. 1976).
- Davidson, D.: How ist weakness of the will possible? In: Feinberg  
(1970).
- Davidson, D./ Suppes, P./ Siegel, S.: Decision-making. An experi-  
mental approach. Stanford 1957.
- Davis s. Thrall/Coombs/Davis
- Debreu, G.: Representation of a preference ordering by a numerical  
function. In: Thrall/Coombs/Davis (1954) 159-165.
- Stochastic choice and cardinal utility. In: Econometrica  
(1958) 440-444.
- Theory of value. New York 1959.
- Topological methods in cardinal utility theory. In: Arrow/  
Karlin/Suppes (1966) 16-26.
- Werttheorie. Eine axiomatische Analyse ökonomischen Gleichge-  
wichts. Berlin, Heidelberg, New York 1976.
- De Groot, M.H.: Some comments on the experimental measurement of  
utility. In: Behavioural science 8 (1963) 146-149.

- Deschamps, R./ Gevers, L.: Leximin and utilitarian rules. A joint characterization. In: Journal of economic theory 17 (1978) 143-163.
- Diamond, P.: Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. A comment. In: Journal of political economy 75 (1967).
- Downs, A.: Ökonomische Theorie der Demokratie. Tübingen 1968.
- Dreyfus, H.L./ Dreyfus, S.E.: Inadequacies in the decision analysis model of rationality. In: Hooker/Leach/McClennen, Bd 1 (1978) 115-124.
- Dworkin, R.: The original position. In: University of Chicago law review 40 (1973), abgedruckt in: Daniels, 16-53.
- Taking rights seriously. London 1977.
- Edwards, W.: Utility, subjective probability, their interaction and variance preferences. In: Journal of conflict resolution 6 (1962) 42-51.
- Egle, K.: Entscheidungstheorie. Eine strukturtheoretische Darstellung. Basel 1975.
- Ehmann, R.: Rawls and Nozick. Justice without well-being. In: The journal of value inquiry 14 (1980) 7-21.
- Eisfeld, R.: Der ideologische und soziale Stellenwert der Pluralismustheorie. In: Politische Vierteljahresschrift 12 (1971) 332-366.
- Ells, J./ Sermat, V.: Motivational determinants of choice in chicken and prisoner's dilemma. In: Journal of conflict resolution 12 (1968) 374-380.
- Ellsberg, D.: Classical and current notions of measurable utility. In: Economic journal 64 (1954) 528-556.
- Ellsworth, L.: Decision-theoretic analysis of Rawls' original position. In: Hooker/Leach/McClennen, Bd 2 (1978) 29-45.
- Eschenburg, R.: Der ökonomische Ansatz zu einer Theorie der Verfassung. Die Entwicklung einer liberalen Verfassung im Spannungsverhältnis zwischen Produktivität und Effektivität der Kooperation. Tübingen 1977.
- Essler, W.K.: Induktive Logik. Grundlagen und Voraussetzungen. Freiburg, München 1970.
- Wissenschaftstheorie II. Theorie und Erfahrung. Freiburg, München 1971.
- Wissenschaftstheorie III. Wahrscheinlichkeit und Induktion. Freiburg, München 1973.
- Philosophie in unserer Zeit. In: Hakamies (1977) 31-41.
- Wissenschaftstheorie IV. Erklärung und Kausalität. Freiburg, München 1979.
- Wissenschaftstheorie I. Definition und Reduktion. Freiburg, München 1982.
- Ethanbolker: Simultaneous axiomatization of utility and subjective probability. In: Philosophical science 34 (1967) 333-340.

- Favre, P.: The rationality of the 'social treaty'. In: Birnbaum/  
Lively/Geraint (1978) 101-132.
- Feinberg, J.: Moral concepts. Oxford 1970.
- Ferejohn, J.: The distribution of rights and society. In: Gottinger/  
Leinfellner (1978) 119-131.
- Fetscher s. Hobbes
- Finetti, B. de: Initial probabilities. A prerequisite for any valid  
deduction. In: Synthese 20 (1969) 2-16.
- Finnin, W./ Smith, G.A. (Hrsg.): The morality of scarcity.  
Baton Rouge, London 1979.
- Fishburn, P.C.: Decision and value theory. New York, London, Sidney  
1964.
- Additivity in utility theory with denumerable product sets. In:  
Econometrica 34 (1966) 500-503.
- Utility theory for decision making. New York 1970.
- Mathematics of decision theory. Mouton, The Hague, Paris 1972.
- Foley, D.: Resource allocation and the public sector. In: Yale  
economic essays 7 (1967).
- Foot, P. (Hrsg.): Theories of ethics. Oxford 1967.
- Fox s. Guyer/Fox/Hamburger
- Frankena, W.K.: Analytische Ethik. München 1972.
- On saying the ethical thing. Proceedings and addresses of the  
American Philosophical Association 39 (1966), abgedruckt in:  
Goodpaster (1976) 107-124.
- Thinking about morality. University of Michigan Press 1980.
- French, J.: A formal theory of social power. In: Psychological review  
63 (1956) 181-194.
- Freud, S.: Abriß der Psychoanalyse. Frankfurt, Hamburg 1955.
- Frey, B.: Moderne politische Ökonomie. München 1977.
- Frey, B./ Schneider, F.: On the modeling of politico-economic inter-  
dependence. In: European journal of political research 1975, 339-  
360.
- Friedman, M./ Savage, L.J.: The expected utility, hypothesis and  
the measurability of utility. In: Journal of political economy  
60 (1952) 463-474.
- Frischeisen-Köhler s. Hobbes
- Funk s. Rapoport/Kahan/Funk/Horowitz
- Gäfgen, G.: Die Beziehungen zwischen Welfare Economics und politi-  
scher Theorie. In: Beckerath (1963) 159-163.

- Gäfigen, G. (Hrsg.): Grundlagen der Wirtschaftspolitik. Köln, Berlin 1966
- Formale Theorie des politischen Handelns. Wissenschaftliche Politik als rationale Wahl von Strategien. In: Maier/Ritter/Matz (1971) 209-258.
- Theorie der wirtschaftlichen Entscheidung. Tübingen 1974.
- Gärdenfors, P.: Forecasts, decisions and uncertain probabilities. In: Erkenntnis 14 (1979) 159-181.
- Gaertner, W.: Über Kenneth Arrows allgemeines Unmöglichkeitstheorem und einige Auswege aus dem Dilemma. In: Jahrbuch für Sozialwissenschaften 3 (1978) 288-310.
- Rawlsianism, utilitarianism and profiles of extended orderings. Bielefeld 1980.
- Gaertner, W./ Krüger, L.: From handcuffed paretians to self-consistent libertarians. A new possibility theorem. Bielefeld 1978.
- Galanter s. Luce/Bush/Galanter
- Gauthier, D.P.: The logic of Leviathan. The moral and political theory of Thomas Hobbes. Oxford 1969.
- (Hrsg.): Morality and rational self-interest. Englewood Cliffs 1970.
- The impossibility of rational egoism. In: Journal of philosophy 71 (1974) 439-456.
- Rational cooperation. In: Nous 8 (1974) 53-65. (1974a)
- Reason and maximization. In: Canadian journal of philosophy 4 (1975) 411-433.
- Morality and advantage. In: Philosophical review 76 (1976) 460-475.
- The social contract. Individual decision or collective bargain? In: Hooker/Leach/McClennen, Bd 2 (1978) 47-67.
- Gawlik s. Hobbes
- Georgescu-Roegen, N.: The pure theory of consumer's behavior. In: Quarterly journal of economy 1 (1936) 545-593.
- Geraint s. Birnbaum/Lively/Geraint
- Gevers s. d'Asprement/Gevers (1977 und 1978)
- s.a. Deschamps/Gevers
- Gibbard, A.: A pareto-consistent libertarian claim. In: Journal of economic theory 7 (1974) 388-410.
- Gibbard, A./ Harper, W.: Counterfactuals and two kinds of expected utility. In: Hooker/Leach/McClennen, Bd 1 (1978) 125-162.
- Giersch s. Beckerath/Giersch
- Gnedenko, B.W.: Lehrbuch der Wahrscheinlichkeitsrechnung. Frankfurt/M. 1978 .
- Goethe s. Bähr

- Goldman, H.S.: The 'collective' interpretation of utilitarian generalization. In: Philosophical studies 34 (1978) 207-209.
- Goodman L.A./ Markowitz, H.: Social welfare functions based on individual rankings. In: American journal of sociology 18 (1952) 257-262.
- Goodpaster, K.E. (Hrsg.): Perspectives on morality. Essays by William K. Frankena. London/Ind. 1976. s.a. Frankena
- Gordon s. Rapoport/Guyer/Gordon
- Gorman, W.M.: Preference, revealed preference and indifference. In: Chipman/Hurwicz/Richter/Sonnenschein (1971) 81-113.
- Gottinger, H.W.: Grundlagen der Entscheidungstheorie. Stuttgart 1974. (1974a)
- Subjektive Wahrscheinlichkeiten. Göttingen 1974. (1974b)
- Gottinger, H.W./ Leinfellner, W. (Hrsg.): Decision theory and social ethics (Int. Sympos. on decision theory and social ethics, Schloß Reisenburg 1976). Dordrecht 1978.
- Graaf, J. de: Theoretical welfare economics. Oxford 1957.
- Equity and efficiency as components of the general welfare. In: South African journal of economics 45 (1977) 362-375.
- Grau, G.-G. (Hrsg.): Probleme der Ethik. Freiburg, München 1972.
- Gray, J.N.: Social contract, community and ideology. In: Birnbaum/Lively/Geraint (1978) 225-244.
- Greenstein, F./ Polsby, N. (Hrsg.): Handbook of political science, Bd 3, s. Taylor
- Grewendorf, G./Meggle, G. (Hrsg.): Seminar: Sprache u. Ethik. Zur Entwicklung der Metaethik. Frankfurt/M. 1974.
- Gunderson, K. (Hrsg.): Language, mind and knowledge. Minneapolis 1975 (Minnesota studies in the philosophy of science, Bd 7).
- Guyer, M./ Fox, J./Hamburger, H.: Format effects in the prisoner's dilemma game. In: Journal of conflict resolution 17 (1973) 719-744.
- s. Rapoport/Guyer/Gordon
- Habermas, J.: Wahrheitstheorien. In: Schulz (1973) 211-265.
- Hagen s. Allais/Hagen
- Hakamies, A. (Hrsg.): Logik, Mathematik und Philosophie des Transzendenten. Paderborn 1977.
- Halldén, S.: On the logic of 'better'. Lund, Kopenhagen 1957.
- Hamburger s. Guyer/Fox/Hamburger
- Hammond, P.J.: Why ethical measures of inequality need interpersonal comparisons. In: Theory and decision 7 (1976) 263-274. (1976b)
- Equity, Arrow's conditions, and Rawls' difference principle. In: Econometrica 44 (1976) 793-804. (1976a)
- Hansson, B.: Group preferences. In: Econometrica 37 (1969) 50-54.
- The decision game - the conceptualisation of risk and utility. In: Morscher/Stranzinger (1981) 187-193.

- Hardin, R.: Collective action as an agreeable n-persons prisoners' dilemma game. In: Rapoport (1974) 35-62.
- Hare, R.M.: The language of morals. Oxford 1952. Dt.: Die Sprache der Moral. Frankfurt/M. 1972.
- Freedom and reason. Oxford 1963. Dt.: Freiheit und Vernunft. Düsseldorf 1973.
- Rawls' theory of justice. In: Philosophical quarterly 23 (1973); abgedruckt in: Daniels, 81-107.
- Harper s. Gibbard/Harper
- Harsanyi, J.C.: Cardinal utility in welfare economics and the theory of risk-taking. In: Journal of political economy 61 (1953) 434-435; abgedruckt in: Harsanyi (1976).
- Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. In: Journal of political economy 62 (1955) 309-321; abgedruckt in: Harsanyi (1976).
- Ethics in terms of hypothetical imperatives. In: Mind 67 (1958) 305-316; abgedruckt in: Harsanyi (1976),
- On the rationality postulates underlying the theory of cooperative games. In: Journal of conflict resolution 5 (1961) 179-196.
- Rationality postulates for bargaining solutions in cooperative and non-cooperative games. In: Management science 9 (1962) 141-153.
- A general theory of rational behavior in game situations. In: Econometrica 34 (1966) 613-634.
- Games with incomplete information played by Bayesian players. In: Management science 14 (1967/68) 159-182.
- The tracing procedure. A Bayesian approach to defining a solution for n-persons non-cooperative games. In: International journal of game theory 4 (1975) 61-94. (1975a)
- Can the maximin principle serve as a basis for morality? A critique of John Rawls' theory. In: American political science review 59 (1975) 594-606; abgedruckt in: Harsanyi (1976). (1975b)
- Nonlinear social welfare functions. Do welfare economists have a special exemption from Bayesian rationality? In: Theory and decision 6 (1975) 311-332; abgedruckt in: Harsanyi (1976). (1975c)
- Essays on ethics, social behavior and scientific explanation. Dordrecht 1976.
- Rule utilitarianism and decision theory. In: Erkenntnis 11 (1977) 25-53. (1977a)
- Morality and the theory of rational behavior. In: Social research 44 (1977) 623-656. (1977b)
- Rational behavior and bargaining equilibrium in games and social situations. Cambridge 1977. (1977c)
- Bayesian decision theory and utilitarian ethics. In: American economic review, Papers and proceedings 68 (1978) 223-228. (1978a)



- Harsanyi, J.C.: Sneed on Rawls' theory of social institutions. Some comments. In: Erkenntnis 13 (1978) 225-230. (1978b)
- Preferences and utilitarian theory. Some comments. In: Erkenntnis 13 (1978) 397-399. (1978c)
- A new solution concept for both cooperative and noncooperative games. Opladen 1979. (1979a)
- Bayesian decision theory, rule utilitarianism, and Arrow's impossibility theorem. In: Theory and decision 11 (1979) 289-317. (1979b)
- Rule utilitarianism, rights, obligations and the theory of rational behavior. In: Theory and decision 12 (1980) 115-133.
- Hart, H.L.A.: Are there any natural rights ? In: Philosophical review 64 (1955) 175-191.
- Law, liberty and morality. Stanford 1963.
- Recht und Moral. Göttingen 1971.
- Hax, H.: Die Koordination von Entscheidungen. Köln u.a. 1965.
- Hayek, F.A.: The constitution of liberty. London 1960. Dt.: Die Verfassung der Freiheit. Tübingen 1971.
- New studies in philosophy, politics, economics and the history of ideas. London 1978.
- Held, V.: Rationality and social value im Game-theoretical analysis. In: Ethics 76 (1965/66) 215-220.
- Hempel, C.G.: Fundamentals of concept formation in empirical science. Chicago 1952.
- Rational action. In: Proceedings and addresses of the APA 35 (1961/62) 5-23.
- Coherence and morality. In: The journal of philosophy 62 (1965) 539-542.
- Herstein, J.N./ Milner, J.: An axiomatic approach to measurable utility. In: Econometrica 21 (1953).
- Hilpinen, R. (Hrsg.): Rationality in science. Dordrecht, Boston, London 1980.
- Hintikka, J.: Possible-world semantics as a framework for comparative and critical philosophy. In: Ryle (1976) 57-69.
- Hirsch, F.: Social limits to growth. Cambridge/Mass. 1976.
- Hobbes, Th.: Elementa philosophica de cive. (Amsterdam 1647). Dt.: Vom Menschen. Vom Bürger. Elemente der Philosophie II.III. Eingel. u. hrsg. von G. Gawlik. Hamburg 1966.
- Leviathan or the matter from and authority of government (1651). Dt.: Leviathan oder Stoff, Form und Gewalt eines bürgerlichen und kirchlichen Staates. Hrsg. u. eingel. von J. Fetcher. Neu-wied, Berlin 1966.

- Hobbes, Th.: Elementorum philosophiae sectio prima de corpore (London 1655). Dt.: Vom Körper. Elemente der Philosophie, Bd 1. Ausgew. u. übers. von M. Frischeisen-Köhler. Hamburg 1967.
- Höffe, O.: Rationalität, Deziision oder praktische Vernunft. Zur Diskussion des Entscheidungsbegriffes in der Bundesrepublik. In: Philosophisches Jahrbuch 80 (1973) 340-368.
- Strategien der Humanität. Freiburg, München 1975. (1975a)
- (Hrsg.): Einführung in die utilitaristische Ethik. München 1975. (1975b)
- (Hrsg.): Über John Rawls' Theorie der Gerechtigkeit. Frankfurt/M. 1977.
- Ethik und Politik. Grundmodelle und Grundprobleme praktischer Philosophie. Frankfurt/M. 1979.
- Hoerster, N.: Utilitaristische Ethik und Verallgemeinerung. Freiburg, München 1977.
- Zur Begründung einer Minimalmoral. In: Morscher/Stranzinger (1981) 131-133.
- Hook, S. (Hrsg.): Human values and economic policy. New York 1967.
- Hooker, C.A./ Leach, J.J./ McClennen, E.F. (Hrsg.): Foundations and applications of decision theory, Bd 1: Theoretical foundations. Bd 2: Epistemic and social applications. Dordrecht, Boston 1978.
- Horowitz s. Rapoport, Amnon/Kahan/Funk/Horowitz
- Houthakker, H.S.: Revealed preference and the utility function. In: *Economica* 17 (1950) 159-174.
- Howard, N.: The theory of metagames. In: *General systems* 11 (1966) 167-186.
- Paradoxes of rationality. Cambridge, London 1971.
- 'General' metagames. An extension of the metagame concept. In: Rapoport (1974) 261-283.
- Hume, D.: A treatise of human nature. Oxford 1888.
- An enquiry concerning the principles of morals (Oxford 1894). Dt.: Eine Untersuchung über die Prinzipien der Moral. Berlin 1962.
- Hurst, M.P./ Siegel, S.: Prediction of decision from a metric scale of utility. In: *Journal of experimental psychology* 52 (1956) 128-134.
- Hurwitz s. Chipman/Hurwicz/Richter/Sonnenschein
- Hutchison, T.W.: Das Problem der systematischen Verknüpfung von Normen und Aussagen der positiven Ökonomik in grundsätzlicher Betrachtung. In: Beckerath (1963) 3-15.

- Inciarte, F.: Theoretische und praktische Wahrheit. In: Riedel, Bd 2 (1972) 155-170.
- Intriligator s. Arrow/Intriligator
- Isbell, J.R.: Absolute games. In: Tucker/Luce (1959)
- Iwand, T.: Ethical systems as order relations. In: Ratio 7 (1965) 145-163.
- Jeffrey, R.C.: Ethics and the logic of decision. In: The journal of philosophy 62 (1965) 528-539.
- Logik der Entscheidungen. Wien, München 1967.
- On interpersonal utility theory. In: The journal of philosophy 68 (1971) 647-656.
- Preferences among preferences. In: The journal of philosophy 71 (1974) 377-391.
- Axiomatizing the logic of decision. In: Hooker/Leach/McClennen, Bd 1 (1978) 227-231.
- Jehle, F.: Boolesche Algebra. München 1974.
- Jochimsen, R.: Grundlagen, Grenzen und Entwicklungsmöglichkeiten der Welfare Economics. In: Beckerath (1963) 129-153.
- Johnson, Karen: A note on the inapplicability of Olson's logic of collective action to the state. In: Ethics 85 (1974/75) 170-174.
- Jones s. Bock/Jones
- Junne, G.: Spieltheorie in der internationalen Politik. Düsseldorf 1972.
- Kästle, R.: Ordinale und kardinale Nutzenmessung in der Ökonomie. In: Jahrbuch für Sozialwissenschaften 1 (1978) 17-37.
- Kahan s. Rapoport, Amnon/.Kahan/Funk/Horowitz
- Kalin, J.: In defence of egoism. In: Gauthier (1970) 64-87.
- Kambartel, F.: Ethik und Mathematik. In: Riedel, Bd 1 (1972) 489-503.
- Kant, I.: Grundlagen zur Metaphysik der Sitten (Riga 1786). Stuttgart 1974.
- Kritik der praktischen Vernunft (1788). Stuttgart 1976.
- Zum ewigen Frieden (1795). Stuttgart 1979.
- Kaplan, M.A.: Means-ends rationality. In: Ethics (1976/77) 61-74.
- Karlin s. Arrow/Karlin/Supper
- Kasch, F./ Pareigis, B.: Grundbegriffe der Mathematik. München 1974.
- Katzner, L.: The original position and the veil of ignorance. In: Blocker/Smith (1980) 42-70.
- Kelly, H.H. s. Thibaut/Kelly
- Kelly, J.S.: The impossibility of a just liberal. In: Economica 43 (1976) 67-75.

- Kemp, M.C.: On the existence of social welfare functions, social orderings and social decisions. In: *Economica* 43 (1976) 59-66.
- Kern, L.: Comparative distributive ethics. An extension of Sen's examination of the pure distribution problem. In: *Gottinger/Leinfellner* (1978) 187-200.
- Sind demokratische Entscheidungsregeln verfälschungsfrei? Eine politische Interpretation einiger Ergebnisse der Theorie kollektiver Entscheidungen. In: *Politische Vierteljahresschrift* 20 (1979) 330-356.
- Zur axiomatischen Charakterisierung alternativer Vertragsprinzipien. In: *Erkenntnis* 15 (1980) 1-31. (1980a)
- Neutralität und Anonymität allgemeiner Vertragsprinzipien. In: *Zeitschrift für philosophische Forschung* 34 (1980) 226-247. (1980b)
- Neue Vertragstheorie. Zur rationalen Rekonstruktion politisch-ethischer Grundprinzipien. Königstein/Ts. 1980. (1980c)
- Keynes, J.M.: *A treatise on probability*. London 1921.
- Keyt, D.: The social contract as an analytic, justificatory and polemic device. In: *Canadian journal of philosophy* 4 (1974) 241-252.
- Kilgour, D.M.: On  $2 \times 2$ -Games and Braithwaite's arbitration scheme. In: *Rapport* (1974) 63-74.
- Kirman, A.: Measure <sup>theory</sup> with applications to economics. In: *Arrow/Intriligator* (1981) Kap. 5.
- Kirsch, G.: *Ökonomische Theorie der Politik*. Tübingen, Düsseldorf 1974.
- Kirsch, W.: *Einführung in die Theorie der Entscheidungsprozesse*, Bd 1-3. Wiesbaden 1977.
- Klaus, G.: *Spieltheorie in philosophischer Sicht*. Berlin 1968.
- Körner, S.: *Practical reason*. Oxford 1974.
- Krantz, D.H./ Luce, D.R./ Suppes, P./ Tversky, A.: *Foundations of measurement*. New York, London 1971.
- Krelle, W./Coerner, P.: *Präferenz- und Entscheidungstheorie*. Tübingen 1968.
- Kreyszig, E.: On a decision theoretic method for social decisions. In: *Hooker/Leach/McClennen* (1978) 275-281.
- Krueger s. Gaertner/Krüger
- Kuchemann, C.: *Rationality and moral obligation*. In: Kuchemann (Hrsg.): *The life and choice*. Boston 1978.
- Kuhn, A.: On relating individual and social decisions. In: *Hooker/Leach/McClennen*, Bd 2 (1978) 69-89.
- Kuhn, H.: *Politik und Wissenschaft*. In: *Maier/Ritter/Matz* (1971) 3-52.
- Kuhn, H.W./Tucker, A.W.: *Contributions to the theory of games*, Bd 2. Princeton 1953.
- Kutschera, F.v.: *Einführung in die Logik der Normen, Werte und Entscheidungen*. Freiburg, München 1973.

- Kutschera, F. v.: Grundlagen der Ethik. Berlin, New York 1982.
- Ladenson, R.F.: Nozick on law and the state. A critique. In: Philosophical studies 34 (1978) 437-444.
- Laffont, J.J. (Hrsg.): Aggregation and revelation of preferences. Amsterdam, New York, Oxford 1979.
- Laffont, J.J./ Maskin, E. : A differential approach to expected utility maximizing mechanisms. In: Laffont (1979) 289-307.
- Laslett, P./Runciman, W.G. (Hrsg.): Philosophy, politics and society, Bd 3. Oxford 1967.
- Leach, J./ Butts, R./ Pearce, G. (Hrsg.): Science, decision and value. (Proceedings of the 5th Univ. of Western Ontario Philos. Coll. 1969.) Dordrecht 1973.
- Leach s. Hooker/Leach/McClennen
- Lehning, F.B.: Social contract and property rights. A comparison between John Rawls and James Buchanan. In: Birnbaum/Lively/Geraint (1978) 279-294.
- Lehrer, K.: Social consensus and rational agnology. In: Synthese 31 (1975) 141-160.
- Rationality and science and society. A consensual theory. In: Ryle (1976) 14-29.
- When rational disagreement is impossible. In: Nous 10 (1976) 327-332. (1976a)
- Consensus and comparison. A theory in social rationality. In: Hooker/Leach/McClennen, Bd 1 (1978) 283-309.
- A model of rational consensus of science. In: Hilpinen (1980).
- Rational consensus in science and society. London 1981.
- Leinfellner s. Gottinger/Leinfellner
- Lewis, D.: Konventionen. Eine sprachphilosophische Abhandlung. Berlin, New York 1975.
- Languages and language. In: Gunderson (1975) 3-35. Dt. in: Meggle (1979) 197-240.
- Liebermann, B.: Experimental studies of conflict in some two-person and three-person games. In: Criswell (1962) 204-220.
- Lindblom s. Braybrooke/Lindblom
- - s.a. Dahl/Lindblom
- Little, J.M.D.: A critique of welfare economics. Oxford 1957
- Lively, J.: Pluralism and consensus. In: Birnbaum/Lively/Geraint (1978) 185-202.
- -s.a. Birnbaum/Lively/Geraint
- Locke, D.: The principle of equal interest. In: Philosophical review 90 (1981) 531-539.
- Loeffel s. Bühlmann/Loeffel/Nievergelt

- Lorenzen, P.: Normative logic and ethics. Mannheim,Zürich 1969.
- Lorenzen, P./ Schwemmer, O.: Konstruktive Logik, Ethik und Wissenschaftstheorie. Mannheim 1973.
- Luce, R.D./ Suppes, P.: Preference, utility and subjective probability. In: Luce/Bush/Galanter (1963)Bd 3, 249-410
- Luce, R.D./ Bush, R.R./ Galanter, E. (Hrsg.): Handbook of mathematical psychology , Bd 3. New York,London 1963.
- Luce, R.D./Raiffa, H.: Games and decisions. New York u.a.1957.
- Luce s. Tucker/Luce , s.a.Krantz/Luce/Suppes/Tversky
- Lumsden, M.: The Cyprus conflict as a prisoner's dilemma game. In: Journal of conflict resolution 17 (1973) 7-32.
- Lyons, D.: Forms and limits of utilitarianism. Oxford 1965.
- Rawls versus utilitarianism. In: Journal of philosophy 69 (1972) 535-545.
- Nature and soundness of the contract and coherence arguments. In: Daniels (o.J.) 141-167.
- Machan, T.R.: Rational choice and public affairs. In: Theory and decision 12 (1980) 229-258.
- Mack, E.: Natural and contractual rights. In: Ethics 87 (1976) 153-159.
- Mackenzie, J.C.: Prescriptivism and rational behavior. In: Philosophical quarterly 18 (1968) 310-319.
- Mackie, J.L.: Ethics. Inventing right and wrong (1977). Dt. : Auf der Suche nach dem Richtigen und Falschen. Stuttgart 1981.
- Maier, H./ Ritter, K./ Matz, U. (Hrsg.): Politik und Wissenschaft München 1971.
- Markovič,M.: Scientific and ethical rationality. In: Hilpinen (1980).
- Markowitz s. Goodman/Markowitz
- Marschak, J.: Rational behavior, uncertain prospects, and measurable utility. In: Econometrica 18 (1950) 111-141.
- Nutzenmessung und Wahrscheinlichkeit. In: Shubik (1965) 103-118.
- Maskin, E.: Decision-making under ignorance with implications for social choice. In: Theory and decision 11 (1979) 319-337.
- - s.a. Laffont/Maskin
- Matz s. Maier/Ritter/Matz
- Mayer-Tasch,P.C.: Autonomie und Autorität. Rousseau in den Spuren von Hobbes?. Neuried,Berlin 1968.
- McClellan, E.F.: The minimax theory and expected utility reasoning. In: Hooker/Leach/McClenren (1978) 337-367.
- Medina, M.: Ansätze einer normativen Spieltheorie. In: Kambartel, F., Mittelstraß, J. (Hrsg.): Zum normativen Fundament der Wissenschaft. Frankfurt/M. 1973, 200-215.
- Medlin, B.: Ultimate principles and ethical egoism. In: Australasian journal of philosophy 35 (1957) 111-118.

- Meeker s. Bruns/Meeker
- Meggle, G. (Hrsg.): Handlung, Kommunikation, Bedeutung. Frankfurt/M. 1979.
- s.a. Grewendorf/Meggle
- Menger, S.: Grundmodelle wirtschaftlicher Entscheidungen. Köln, Opladen 1969.
- Menges, G. (Hrsg.): Information, inference and decision. Dordrecht 1974.
- Meyer-Krahmer, F.: Politische Entscheidungsprozesse und ökonomische Theorie der Politik. Frankfurt 1979.
- Michalos, A.C.: Estimated utility and corroboration. In: British journal for the philosophy of science 16 (1966/67) 327-331-
- Postulates of rational preference. In: Philosophical science 34 (1967) 18-22.
- Mill, J.St.: Utilitarianism. London 1863. Dt.: Der Utilitarismus. Stuttgart 1976.
- Miller, R.W.: Rights and reality. In: The philosophical review 90 (1981) 383-407.
- Milnor, J.: Games against nature. In: Thrall/Coombs/Davis (1954) 49-59. Dt.: Spiele gegen die Natur. In: Shubik (1965) 129-139.
- s.a. Herstein/Milnor
- Minas, J.S./ Ackoff, R.L.: Individual and collective value judgments. In: SHelly/Bryan (1964) 351-359.
- Mironesco, C.: Individual and collective interests in game theory. In: Birnbaum/Lively/Geraint (1978) 217-334.
- Moeschlin, O./ Pallaschke, D. (Hrsg.): Game theory and related topics. Amsterdam, New York, Oxford 1979.
- Moore, G.E.: Principia ethica. Cambridge 1903. Dt.: Principia ethica 1970.
- Morgenstern, O.: Spieltheorie. Ein neues Paradigma der Sozialwissenschaft. In: Zeitschrift für Sozialwissenschaft 23 (1968) 145-164.
- s.a. v. Neumann/Morgenstern
- Morscher, E./ Stranzinger, R. (Hrsg.): Ethik. Grundlagen, Probleme, und Anwendungen. Akten des 5. Intern. Wittgenstein-Symposiums. Wien 1981.
- Moses s. Chernoff/Moses
- Mosteller s. Noguee
- Mueller, D.C.: Public choice. A survey. In: Journal of economic literature 14 (1976) 395-433.
- Muench, T./ Walker, M.: Identifying the free rider problem. In: Laffont (1979) 61-87.
- Murakami, Y: Logic and social choice. London 1968.
- Nakhtnikian s. Castaneda/Nakhtnikian
- Nash, J.F.: The bargaining problem. In: Econometrica 18 (1950) 155-162

- Nash, J.F.: Non cooperative games. In: Annals of mathematics 54 (1951) 286-295.
- Neumann, J. v./ Morgenstern, O.: Theory of games and economic behavior. Princeton 1947. Dt.: Die Theorie der Spiele und das wirtschaftliche Verhalten. Würzburg 1961.
- Spieltheorie und wirtschaftliches Verhalten. Würzburg 1973 (3. Auflage, s.O.)
- Ng, Y.K.: The possibility of a paretian liberal. Impossibility theorems and cardinal utility. In: Journal of political economy 79 (1971) 1397-1402.
- s.a. Kemp/Ng
- Nievergelt s. Bühlmann/Loeffel/Nievergelt
- Nogee, P./ Mosteller, F.: An experimental measure of utility. In: Journal of political economy 59 (1951) 371-404.
- Nowell-Smith, P.H.: A theory of justice. In: Philosophy of the social sciences 3 (1973). Dt. in: Höffe (1977) 77-107.
- Nozick, R.: Distributive justice. In: Philosophy and public affairs 3 (1973) 45-126.
- Anarchy, state and utopia. Oxford 1974. Dt.: Anarchie, Staat, Utopia. München (o.J.).
- Oksenberg, A. (Hrsg.): Essays on Aristotle's ethics. Berkeley 1980.
- Olson, M.: The logic of collective action. Cambridge 1965. Dt.: Die Logik kollektiven Handelns. Tübingen 1968.
- Ordeshook, P. (Hrsg.): Game theory and political science. New York 1978.
- s.a. Riker/Ordeshook
- Orwant s. Rapoport/Orwant
- Osborne, D.K.: On liberalism and the pareto-principle. In: Journal of political economy 83 (1975) 1283-1288.
- Owen, G.: Spieltheorie. Berlin, Heidelberg, New York 1971.
- Pallascke s. Moeschlin/Pallaschke
- Pareigis s. Kasch/Pareigis
- Patzig, G.: Die Begründung moralischer Normen. In: Patzig/Scheibe, E./ Wieland, W.: Logik, Ethik, Theorie der Geisteswissenschaften. (11. Kongreß der Philosophie. Göttingen 1975). Hamburg 1977.
- Pazner, E.: Recent thinking on economic justice. In: Journal of peace science 2 (1976) 143-153.
- Pitfalls in the theory of fairness. In: Journal of economic theory 14 (1977) 458-465.
- Equity, non feasible alternatives and social choice. A reconsideration of the concept of social welfare. In: Laffont (1979) 161-173.
- Pazner, E./ Schmeidler, D.: A difficulty in the concept of fairness. In: Review of economic studies 41 (1974).



- Fazner, E./ Schmeidler, D.: Egalitarian equivalent allocations.  
A new concept of economic equity. In: Quarterly journal of economics 92 (1978) 671-687.
- Pearce s. Leach/Butts/ Pearce
- Peleg, B.: Game theoretic analysis of voting schemes. In: Moeschlin/  
Pallaschke (1979) 83-89.
- Pennock, J.R./ Chapman, J.W. (Hrsg.): Coercion. Chicago, New York  
1972.
- Perner s. Rapoport/Perner
- Pflanzagl, J.: A general theory of measurement. Applications to  
utility. In: Naval research logistic quarterly 6 (1959) 283-294.
- Pieper, Annemarie: Analytische Ethik. Ein Überblick über die seit  
1900 in England und Amerika erschienene Ethik-Literatur. In:  
Philosophisches Jahrbuch 78 (1971) 144-176.
- Platon: Georgias. In: Platonis opera, ed. J. Burnet (Oxford 1900-  
1907; Nachdr. 1967), Bd 3, 447-527.
- Protagoras. A.a.O., S. 309-362.
- Politeia. In: Platonis opera, Bd 4, 327-621.
- Polsby s. Greenstine/Polsby
- Popper, K.: Logik der Forschung. Tübingen 1973.
- Potter, N.: Can all moral obligations be deduced from a principle  
of justice? In: Morscher/Stranzinger (1981) 255-257.
- Quine, W.v.: Roots of reference. La Salle/Ill. 1973. Dt.: Die  
Wurzeln der Referenz. Frankfurt/M. 1976.
- Raiffa, H.: Decision analysis. London 1978; s.a. Luce/Raiffa
- Ramsey, F.P.: The foundation of mathematics and other logical issues.  
London 1954.
- Rapoport, Amnon/ Kahan, I./ Funk, S./ Horowitz, A.: Coalition  
formation by sophisticated players. Berlin, Heidelberg, New York  
1979.
- Rapoport, Anatol: Fights, games and debates. Ann Arbor 1960.
- - Strategy and conscience. New York 1964.
- - (Hrsg.): Game theory as a theory of conflict resolution.  
Dordrecht, Boston 1974.
- Rapoport, An./ Chammah, A.M.: Prisoner's dilemma. A study in con-  
flict and cooperation. Ann Arbor 1965.
- Rapoport, An./ Guyer, M.J./ Gordon, D.G.: The 2 x 2 game. Ann Arbor  
1976.
- Rapoport, An./ Orwant, C.: Experimentelle Spiele. Eine Übersicht.  
In: Shubik (1965) 296-324.
- Rapoport, An./ Perner, J.: Testing Nash's solution of the coop-  
erative games. In: Rapoport (1974) 103-115.
- Raub, W./ Voss, T.: Individuelles Handeln und gesellschaftliche  
Folgen. Das individualistische Programm der Sozialwissenschaften.  
Darmstadt 1981.

- Rawls, J.: A theory of justice. New Haven 1971. Dt.: Eine Theorie der Gerechtigkeit. Frankfurt/M. 1975.
- Rescher, N.: The Logic of decision and action. Pittsburgh 1967.
- - Introduction to value theory. Englewood Cliffs 1969.
  - - The social issues in philosophical perspective. London 1972.
  - - Unselfishness. The role of the vicarious affects in moral philosophy and social theory. Pittsburgh 1975.
- Richter s. Chipman/Hurwicz/Richter
- Riedel, Manfred(Hrsg.) Rehabilitierung der praktischen Philosophie, Bd 1. 2. Freiburg 1972.
- - Norm und Werturteil. Stuttgart 1979.
- Riker, W.H./ Ordeshook, Peter C.: An introduction to positive political theory. Englewood Cliffs 1973.
- Ritter s. Maier/Ritter/Matz.
- Robbins, L.: An essay on the nature and significance of economic science. London 1932.
- Roberts, Kevin W.: Possibility theorems with interpersonally comparable welfare levels. In: Review of economic studies 47 (1980) 409-420. (1980a)
- - Interpersonal comparability and social choice theory. In: Review of economic studies 47 (1980) 421-439. (1980b)
- Rodriguez-Marin, J.: Norms, values and preference. In: Morscher/Stranzinger (1981) 181-184.
- Rorty, A.: Akrasia and pleasure. Nichomachean ethics, book 7. In: Oksenberg (1980).
- Rothkirch, C.v.: Rationales Handeln im öffentlichen Interesse. Frankfurt/M.,New York 1981.
- Rousseau, J.J.: Du contrat social(1762). Paris 1822. Dt.: Der Gesellschaftsvertrag. Stuttgart 1971.
- Runciman,W.G./ Sen, A.K.: Games, justice and general will. In: Mind 24 (1965) 554-562.
- Ryle, G. (Hrsg.) Contemporary aspects of philosophy. Stocksfields 1976.
- Samuelson, P.A.: Probalita, utility and the independence axiom.In: Econometrica 20 (1952) 670-678.
- Savage, L.J.: The theory of statistical decision. In: Journal of the American statistical association 46 (1951) 55-67.
- - The foundations of statistics. New York 1972.
- Savage s. Friedmann/Savage
- Savigny s. Austin
- Scheibe s. Patzig
- Schelling, T.C.: A strategy of conflict. Prospectus for a re-orientation of game theory. In: Conflict resolution 2 (1958) 203-264.
- - The strategy of conflict. Cambridge 1963.

- Schelling, T.C.: Game theory and the study of ethical systems.  
In: Journal of conflict resolution 12 (1968) 34-44.
- Schick, F.: Consistency and rationality. In: The journal of philosophy 60 (1963) 5-19.
- - Beyond utilitarianism. In: Journal of philosophy 68 (1971) 656-667.
- - Democracy and interdependent preferences. In: Theory and decision 3 (1972) 55-75.
- - Toward a theory of sociality. In: Hooker/Leach/McClennen ,Bd2. (1978) 121-130.
- Schlicht, E.: Die Theorie der kollektiven Entscheidung und der individualistische Ansatz. In: Leviathan, Bd 2 (1974) 263-280.
- Schmeidler s. Pazner/Schmeidler
- Schmitt, C.: Der Leviathan in der Staatslehre des Thomas Hobbes. Hamburg 1938.
- Schneeweiß, H.: Nutzenaxiomatik und Theorie des Messens. In: Statistische Hefte 4, 178 (1963).
- - Entscheidungskriterien bei Risiko. Berlin, Heidelberg, New York 1967.
- - Probability and utility. Dual concepts in decision theory. In: Menges, G. (Hrsg.): Information and decision. Dordrecht 1974, 113-144
- Schneider s. Frey/Schneider
- Schnur, R.: Individualismus und Absolutismus. Zur politischen Theorie des Thomas Hobbes. Berlin 1963.
- Schopenhauer, A.: Die beiden Grundprobleme in der Ethik (1841). Parerga und Paralipomena (1851). In: Gesamtausgabe, Bd 1-6. Stuttgart 1891.
- Schulz, W.: Wirklichkeit und Reflexion. Pfullingen 1973.
- Schwemmer s. Lorenzen/Schwemmer
- Sellars, W.: Imperatives, intentions and the logic of "ought". In: Castaneda/Nakhnikian (1963) 159-218.
- Selten, R.: Anwendungen der Spieltheorie auf die politische Wissenschaft. In: Maier/Ritter/Matz (1971) 287-320.
- Sen, A.K.: Collective Choice and social welfare. San Francisco 1970. (1970a)
- - Interpersonal aggregation and partial comparability. In: Econometrica 38 (1970). (1970b)
- - The impossibility of a paretian liberal. In: Journal of political economy 78 (1970). (1970 c)
- - Behavior and the concept of preference. In: Economica 40 (1973).
- - Choice, ordering and morality. In: Körner (1974) 54-67. (1974a)
- - Reply to comments. In: Körner (1974) 78-82.

- Sen, A.K.: Informational bases of alternative welfare approaches. In: Journal of public economics 3 (1974) 387-403. (1974b)
- - Rawls versus Bentham. An axiomatic examination of the pure distribution problem. In: Daniels (o.J.) 283-292.
  - - Welfare inequality and Rawlsian axiomatics. In: Theory and decision 7 (1976) 243-262. (1976a)
  - - Liberty, unanimity and rights. In: Econometrica (1976) 217-245. (1976b)
  - - On weights and measures. Informational constraints in social welfare analysis. In: Econometrica 45 (1977) 1539-1572. (1977a)
  - - Rationality and morality. A reply. In: Erkenntnis 11 (1977) 225-232. (1977b)
  - - Personal utilities and public judgements. Or what is wrong with welfare economics? In: The economic journal 89 (1979) 537-558. (1979a)
  - - Social choice theory. (Febr. 1979).(Ms)  
erscheint in Arrow/Intriligator, Handbook of math. ec., Bd.3
  - - A positive concept of negative freedom. In: Morscher/Stranzinger (1981) 43-56.
  - - s. a. Runciman/Sen
- Sermat s. Ells
- Shapley, L.S.: A value for n-person games. In: Kuhn/Tucker (1953) 307-317.
- Shelly, M.W./Bryan, G.L.(Hrsg.): Human judgements and optimality. New York, London, Sydney 1964.
- Shimony, A.: Coherence and the axiom of confirmation. In: Journal of symbolic logic 20 (1955) 1-28.
- Shubik, Martin (Hrsg.): Game theory and related approaches to social behavior. New York u.a. 1965. Dt.: Spieltheorie und Sozialwissenschaft. Frankfurt 1965.
- - Game theory, behavior, and the paradox of prisoner's dilemma. Three solutions. In: Journal of conflict resolution 13 (1969) 182-193.
  - - Game theory models and methods in political economy. In: Arrow/Intriligator, Bd 1 (1981) Kap.7.
- Sidgwick, H.: The method of ethics. London 1907.
- Siegel s. Davidson/Suppes/Siegel
- - s.a. Hurst
- Smart, J.J.C.: An outline of a system of utilitarian ethics. Melbourne 1961.
- Smart, J.J.C./ Williams, B.: Utilitarianism - For and against. Cambridge 1973.

- Smith s. Blocker/Smith  
- - s.a. Finnin/Smith
- Smoke, W./ Zajone, R.B.: On the reliability of group judgement and decisions. In: Criswell/Solomon/Suppes (1962) 322-333.
- Sneed, J.: Strategy and the logic of decision. In: Synthese 16 (1966) 270-283.
- - Entropy, information and decision. In: Synthese 17 (1967) 302-407.
- - John Rawls and the liberal theory of society (Vortrag Ludwig-Maximilians-Univ. München, Seminar f. Philosophie, Logik und Wissenschaftstheorie, August 1975).
- Snyder, G.H.: Deterrence and power. In: The journal of conflict resolution 4 (Juni 1960) 163-178.
- Sobel, J.H.: Interaction problems of utility maximizers. In: Canadian journal of philosophy 4 (1975) 677-688.
- Solomon s. Criswell/Solomon/Suppes
- Sonnenschein s. Chipman/Hurwicz/Richter/Sonnenschein (1971)
- Sosa s. Chisholm/Sosa
- Spinoza, B.: Tractatus theologico-politicus (=Werke, Bd 1, hrsg. von Gawlik u. Niewöhner. Darmstadt 1979). Dt.: Theologisch-politischer Tractat hrsg. v. C. Gebhardt. Hamburg 1965.
- Spohn, W.: Grundlagen der Entscheidungstheorie. Kronberg /Ts. 1978.
- - How to make sense of game theory. In: Studies in contemporary economics, Bd 2: Philosophy of economics, hrsg. v. W. Stegmüller, W. Balzer, W. Spohn. Berlin, Heidelberg, New York 1982.
- Stegmüller, W.: Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Bd 1: Wissenschaftliche Erklärung und Begründung. Berlin, Heidelberg, New York 1969.
- - Bd 2: Theorie und Erfahrung. 1970. Bd 4: Personelle und statistische Wahrscheinlichkeit. 1973.
- - On the interrelations between ethics and other fields of philosophy and science. In: Erkenntnis 11 (1977) 55-80.
- - Hauptströmungen der Gegenwartsphilosophie, Bd 2. Stuttgart 1979.
- - s. a. Carnap/Stegmüller
- Stranzinger s. Morscher/Stranzinger
- Strasnick, S.: The problem of social choice: Arrow to Rawls. In: Philosophy and public affairs 5 (1975) 241-273.
- - Social choice and the derivation of Rawl's difference principle. In: Journal of philosophy 73 (1976) 85-99.
- - Moral structures and axiomatic theory. In: Theory and decision 11 (1979) 195-206.
- Strauss, L.: Die Unterscheidung zwischen Tatsachen und Werten. In: Albert (1971) 73-91.
- Stroh, G.W.: American ethical thought. Chicago 1979.
- Suppes, P.: The role of subjective probability and utility in decision-making. In: The proceeding of the third Berkeley Symposium on Mathematical Statistics and Probability, 1954, 55, Bd 5 (1956) 61-73. Abgedr. in: Suppes (1969).

- Suppes, P.: Behavioristic foundations of utility. In: *Econometrica* 29 (1961) 186-202. Abgedr. in: Suppes (1965) 130-147.
- - The philosophical relevance of decision theory. In: *The journal philosophy* 58 (1961) 605-614. Abgedr. in: Suppes (1969) 105-114. (1961a)
  - - Some formal models of grading principles. In: *Synthese* 16 (1966) 284-306. Abgedr. in: Suppes (1969) 148-169.
  - - Studies in the methodology and foundations of science. Dordrecht 1969.
  - - Axiomatic set theory. New York u.a. 1969. (1969a)
  - - The concept of obligation in the context of decision theory. In: Leach /Butts/Pearce (1973) 1-14.
  - - s.a.Arrow/Karlin/Suppes; Criswell/Solomon/Suppes; Davidson/Suppes/Siegel; Krantz/Luce/Suppes/ Tversky; Luce/Suppes
- Suppes, P./ Winet, M.: An axiomatization of utility based on the notion of utility differences. In: *Management science* 1 (1955) 259-270, abgedruckt in: Suppes (1969) 115-129.
- Suppes, P./ Zinnes, J.L.: Basic measurement theory. In: Luce/Bush/Galanter (1963) 1-76.
- Svensson, L.G.: Social justice and fair distributions. Lund 1977.
- Taylor, M.: The theory of collective choice. In: Greenstein/ Polsby, Bd 3, 413-481.
- Teitelman, M.: The limits of individualism. In: *The journal of philosophy* 69 (1972) 545-556.
- Thibaut, J.W./ Kelly, H.H.: The social psychology of groups. New York, London/Sidney 1959.
- Thompson, G.: Game theory and social value states. In: *Ethics* 75 (1964/65) 36-39.
- Thompson, W.: Maximin stragedies and elicitation of preference. In: Laffont (1979) 245-267.
- Thorndike, R.L.: The effect of discussion upon the correctness decisions, when the facto- of majority influence is allowed for. In: *Journal of social psychology* 9 (1938) 343-363.
- Thrall, R.M./Coombs, C.H./ Davis, R.L. (Hrsg.): Decision processes. New York, London 1954.
- Thurstone, L.L.: The measurement of values. In: *Psychological review* 61 (1954) 47-58.
- - The measurement of values. Chicago 1959.
- Topitsch s. Albert/Topitsch
- Trany, K.E.: Norms of inquiry. Rationality consistency requirements and normative conflict. In: Hilpinen (1980).

- Tucker s. Kuhn/Tucker
- Tucker, A.W./ Luce, R.D. (Hrsg.): Contributions to the theory of games, Bd 4: Annals of mathematics studies, 40. Princeton 1959.
- Tugendhat, E.: Bemerkungen zu einigen methodischen Aspekten von Rawls "Theorie der Gerechtigkeit". In: G. Patzig: 'Rawls Theorie der Gerechtigkeit'. Ein Symposium (1977).
- Tullock, G.: The general irrelevance of the general impossibility theorem. In: Quarterly journal of economics 81 (1967) 256-270.
- - s.a.Buchanan/Tullock
- Tversky s. Krantz/Luce/Suppes/Tversky
- Ullman-Margalit, E.: Coordination norms and social choice. In: Erkenntnis 11 (1977) 143-155.
- Vanacke, W.E./ Arkoff, A.: An experimental study of coalitions in the triad. In: American social review 22 (1957) 406-414.
- Varian, H.R.: Equity, envy and efficiency. In: Journal of economic theory 9 (1974) 63-91.
- - Distributive justice, welfare economics and the theory of fairness. In: Philosophy of public affairs 4 (1974/75) 223-247.
- Vogelsang, R.: Die mathematische Theorie der Spiele. Bonn, München 1963.
- Vorobjeff, N.N.: Grundlagen der Spieltheorie. Würzburg, Wien 1972.
- Voss s. Raub/Voss
- Wagner, C.: Consensus through respect. A model of rational group decision-making. In: Philosophical studies 34 (1978) 335-349.
- Wald, A.: Statistical decision functions. New York 1950.
- Wald, A.: The theory of games. In: Shubik (1954) 33-42.
- Waldner, J.: Good desires. In: Black (1975) 453-481.
- Walker s. Muench/Walker
- Walsh, D.: "Akrasia" reconsidered. In: Ethics 85 (1974/75) 151-158.
- Watkins, J.W.N.: Methodological individualism. A reply. In: Philosophy of science 22 (1955) 58-62.
- - Comment 'Self-interest and morality'. In: Körner (1974) 67-77.
- Weisser, G.: Politik als System aus normativen Urteilen. Göttingen 1951.
- Weymark, J.A.: Unselfishness and prisoner's dilemma. In: Philosophical studies 34 (1978) 417-425.
- Wheeler, J.R.: Ethical egoism in hellenic thought. Ann Arbor 1976.
- Whiteley, P.: Public choice. A dissenting view. In: Buchanan (1978) 150-155.
- Wieland s. Patzig
- Williams s. Smart/Williams
- Wilson, K.V./ Bixenstine, V.E.: Forms of social control in two-person, two-choice-games. In: Behavioral science 7(1962) 91-102.

Winet s. Suppes

Wittman, D.: A diagrammatic exposition of justice. In: Theory and decision 11 (1979). in: Brock, Teil 1 (1979) 207-237.

Wolff, R.P.: Reflections on game theory and the nature of value. In: Ethics 72 (1962).

- - Understanding Rawls. Princeton 1977.

Wright, G.H.V.: The logic of preference. Edinburgh 1963.

Yaari, M.E.: Endogenous changes in tastes. A philosophical discussion. In: Erkenntnis 11 (1977) 157-196.

Zajone s. Smoke/Zajone

Zelger, J.: Zum Begriff der vernünftigen Entscheidung. In: Morscher, stranzinger (1981) 215-218.

Zellner, A.: War and peace. A fantasy in game theory? In: Journal of conflict resolution 6 (1962) 39-41.

Zinnes s. Suppes/Zinnes



PERSONENREGISTER

- Albert 151  
Aldrup 151  
Allais 10, 31, 58  
Allen 262  
Alt 74  
Armbruster 191  
Armstrong 78f  
Arnaszus 31  
Arrow 8, 34, 49, 52, 58,  
65, 73, 82f, 117  
d'asprenment 58, 65  
Austin 152  
Baier 9, 132, 152, 170, 203,  
218, 230, 247, 254 f  
Bar-Hillel 34  
Barry 125, 193  
Becker 58  
Bentham 215  
Bergson 51, 56  
Bernoulli , D. 8, 27, 34, 57  
Bernoulli , N. 27  
Bertrand 34  
Bixenstine 212  
Bjerring 191  
Black 83, 197  
Blau 116  
Bock 74  
Böge 191  
Bohnen 99  
Boulding 197  
Braithwaite 34, 157f, 176,  
215, 235  
Brandt 9, 31  
Braungart 163  
Braybrooke 99, 197  
Brehmer 125  
Brew 198  
Brock 28, 170f, 198, 215  
Buchanan 125  
Bühlmann 16, 47, 84  
Burger 47  
Burns 197  
Butler 168  
Camacho 58  
Carnap 34, 74  
Chammah 47, 212  
Chernoff 16, 47, 57  
Chipman 58, 74  
Condorcet 8  
Daniel 235  
Daniels 65, 160  
Davidson 58, 98  
Debreu 28, 57, 58  
De Groot 58  
Dechamps 60  
Diamond 107  
Downs 193  
Dreyfus, H.L. 31  
Dreyfus, S.E. 31  
Dworkin 125  
Egle 16  
Ellsberg 58  
Ellsworth 125  
Essler 74, 125  
Fauvre 125  
Feinberg 98  
Fletscher 171  
de Finetti 34  
Finnin 197  
Fishburn 16, 85, 130  
Foley 75, 235  
Frankana 127, 151, 153, 155  
Frey 193  
Friedmann 57  
Funk 47, 212  
Gäfgen 16, 57  
Gärdenfors 31, 39  
Gaertner 65, 116, 160  
Gardano 34  
Gauthier 125, 170, 183, 187,  
189f, 203  
Georgescu-Roegen 78 f  
Gevers 58, 60, 65  
Gibbard 116  
Gnedenko 34  
Goodmann 78 f  
Gordon 213  
Gottinger 9, 16, 57, 145, 170,  
191 f  
Gorman 58,  
de Graaf 105  
Gray 125  
Grewendorf 151  
Guyer 213  
Hagen 31, 58,  
Hammond 60, 73, 160  
Hardin 197  
Hare 125  
Harsanyi 28, 97, 99, 100, 106ff,  
122, 125, 154, 159, 176 f,  
191, 195f

- Hayek 159  
Hempel 74, 130  
Herstein 57  
Hirsch 31  
Hobbes 162, 171  
Höffe 9, 31, 99  
Hoerster 163, 203, 215  
Hooker 191 f, 197, 198  
Horowitz 47, 212  
Houthakker 58  
Howard 176, 182, 211f  
Hume 169  
Hurst 58  
Hurwicz 58  
Inciarte 151  
Isbell 66  
Jeffrey 16, 17ff, 61, 66f, 197  
Jehle 18  
Jevons 27  
Jones 74  
Kahan 47, 212  
Kalin 190  
Kant 111, 153, 155, 169, 215  
Karlin 58  
Kasch 229  
Katzner 125  
Kelly 47, 212  
Kern 60, 117  
Keynes 34  
Keyt 125  
Kilgour 235  
Kirman 130  
Kirsch 193  
Klaus 47  
Körner 224, 230  
Kolmogoroff 30  
Krelle 16  
Krüger 116  
Kuhn 176  
Laffont 58, 75  
Leach 191 f, 197, 198  
Lehning 125  
Leinfellner 9, 161, 191f  
Lewis 188, 204  
Loeffel 16, 47, 84  
Luce 16, 37, 47, 58, 66, 130,  
172, 183, 195  
Lyons 99, 125  
Mackie 106, 152, 215  
Markowitz 78f  
Marschak 57, 58  
Mayer-Tasch 171  
McClennen 191f, 197, 198  
Medlin 190  
Meeker 197  
Meggle 151, 204  
Menger 27  
Meyer-Krahmer 193  
Michalos 58  
Mill 99, 215  
Miller 125  
Milnor 39, 57  
Moeschlin 191  
Moore 99f, 190, 215  
Morgenstern 31, 57, 145  
Morscher 163  
Moses 16, 47, 57  
Mosteller 58  
Nash 176, 183  
v. Neumann 31, 57, 145  
Newing 83  
Nievergelt 16, 47, 84  
Nigel 213  
Nogee 58  
Nowell-Smith 9, 125  
Nozick 116  
Oksenberg 98  
Olson 193  
Ordenshook 171, 193, 215  
Orwant 47, 212  
Osborne 116  
Owen 47, 67  
Pallaschke 191  
Pareigis 229  
Parfit 211  
Pazner 75, 215, 235  
Perner 212, 213  
Pfanzagl 130  
Pieper 151  
Platon 167, 169  
Popper 151  
Potter 125  
Pringsheim 27  
Quine 110  
Raiffa 16, 31, 37, 47, 172,  
183, 195  
Ramsey 57  
Rapoport, Am. 47, 212  
Rapoport, An. 47, 176f, 197, 209ff  
Raub 159  
Rawls 9, 125, 160, 170, 214  
Rescher 111, 161, 198f, 259, 261  
Richter 58  
Riedel 9, 151  
Riker 193  
Robbins 74  
Roberts 60, 73, 160  
Rorty 98  
Rothkirch 50, 64, 125  
Rousseau 171  
Runciman 262  
Samuelson 57  
Savage 16, 17, 34, 36, 57  
Schelling 47, 176, 188  
Schick 91 f, 94, 132, 198  
Schlicht 78  
Schmeidler 75, 215, 235  
Schmitt 170  
Schneeweiß 16, 31, 57, 85, 145  
Schnur 170

Schopenhauer 111  
Sen 48ff, 60, 62, 73, 75,  
105, 116ff, 129E, 160,  
224 ff, 262  
Shapley 176  
Shimony 30  
Shubik 47, 57  
Sidgwick 190, 215  
Siegel 58  
Singer 215  
Smart 99, 215,  
Smith 111, 197  
Sobel 180  
Sonnenschein 58  
Spinoza 111  
Spohn 16, 21, 22, 74, 191  
Stegmüller 16, 31, 74, 97,  
127, 130  
Stranzinger 163  
Strasnick 65  
Suppes 16, 57, 58, 61, 75,  
79, 130, 137f, 147, 215  
Svensson 75, 215, 235  
Teitelman 160  
Teuffel 167  
Thibaut 47, 212  
Thomson 58  
Thurstone 74  
Tucker 66, 176  
Tugendhat 125  
Tulloch 83  
Vanacke 47, 212  
Varian 75, 215, 235  
Vogelsang 47, 212  
Vorobjeff 47  
Voss 159  
Wald 35  
Waldner 197  
Walras 27  
Walsh 98  
Watkins 159, 230  
Weizsäcker 210  
Weymark 261  
Wheeler 168  
Williams 215  
Wilson 47, 212  
Winet 61  
Wittman 64, 197, 235  
Wolff 124  
Zinnes 130

SACHREGISTER

- Abneigung 82
- äquivalent 174
- Aggregation: 48
  - von Präferenzen 91, 101, 126
- Aggregationspostulate 49ff, 61, 105, 119, 214
- , normative 122
- Aggregationstheorien, normative 80f, 93ff, 103
- der Sozialethik 116
- Als-Ob-Spiele 225 ff
- Altruismus 246ff
- Altruissummspiel 226ff
- Analytisch 198
- Anaxzustand 85ff, 128
- Anonymitätsbedingung 73, 79
- Anti-Egalitarismus 246ff
- Argument des unzureichenden Grundes 126
- Argument der Verallgemeinerung 215
- außermoralisch 146, 154ff
- austauschbar 174
- Auswahlfunktion 49
- Auswahlmenge 49
- Auszahlung s. "Bewertung, subjektive"
- Axiomatik 143
- Axiomatisierung sozialethischer Kriterien 65f
- axiologisch 151ff
- axiologische Theorie 151ff
- Basis, außermoralische 100
- Bayesianisierung 178, 191
- Bayesianismus, ethischer 97ff, 105 ff
- Bayessche Entscheidungstheorie 15, 29
- Bayessches Kriterium 41, 219
- Begründbarkeitsproblem 151
- Begründung 65
  - , intersubjektive 164
- Begründungsmethode 104
- Behaviorismus 60
- Bernoulli-Prinzip 16
- Bernoulli-Nutzenfunktion 111
- Bewertung 15, 26ff, 59ff
  - , individuelle 59f
  - , Konsequenzenbewertung 21, 30, 211ff
  - , subjektive 25, 45f, 59f, 194f
  - Wahrscheinlichkeitsbewertung 21, 24, 32
  - Zustandsbewertung 21, 199
- Bewertungsfunktion 29, 45, 134
  - , bedingte 84ff
  - Beschränktheit subjektive B. 115
  - , individuelle 59ff

- , kardinale 59ff, 111f
- , ordinale 59ff
- , resultierende moralische 113ff
- , subjektive 42, 61, 85ff, 195
- , subjektive moralische 113ff, 216ff
- collective-choice-Theorie 56, 157, 164
- Differenzenprinzip 65, 214
- Dilemma, praktisches 208 ff
- , theoretisches 208ff
- Dominanzkriterium 37, 173
- Egoist 183
- Einzelfallbeurteilung 65
- Entscheidung, dominante 29, 226
- , irrationale 30
- , kollektive 172ff, 220ff, 225ff
- , kollektive kooperative 226
- , partielle 139
- Entscheidungsbereich, persönlicher 118ff
- Entscheidungskriterium 19ff
- Entscheidungslogik 15ff
- Entscheidungsregel, lexikographische 141
- Entscheidungssituation 42f, 173f, 206f
- Definition 19
- Kontinuum der E. 41f
- ohne Interaktion 144
- Entscheidungstheoretische Modelle 16, 88, 127
- Entscheidungsträger 19, 145
- epistemisch 24
- Ereignis 15, 17f, 72, 139
- (subjektive) Wahrscheinlichkeit von Ereignissen 42, 144
- Erkenntnis, moralische 163
- Erwartungswert, subjektiver 15
- erweiterte individuelle Ordnung 75
- Ethifizierung 176, 191ff
- Ethik, egoistische 167f
- erster Ordnung 158ff
- , nicht-axiologische 157f
- , normative 110
- , utilitaristische 167f
- zweiter Ordnung 157ff, 241
- Evidenz 65
- Existente Form 44f, 233
- Externe Effekte 214
- Fairempfehlung 165
- Faireßprinzipien 215
- Faireßtheorie 158ff
- Gefangenendilemma 172ff, 206f, 208ff, 216ff, 225ff, 239ff
- Gefühl, moralisches 111
- Gegenstand der moralischen Beurteilung 234ff, 258f
- Gemeinwohl 102, 161
- Generalisierbarkeitskriterium 194ff
- Gerechtigkeit, distributive 102
- Gerechtigkeitsempfinden 245ff

Gesellschaft, wohlgeordnete 170  
Gesetz, moralisches 169  
Glaubensaxiom 141  
Gleichbehandlung 73  
Gleichgewichtspunkte 226ff, 245ff  
Gleichgewichtsstrategie 174ff  
-, dominante 188ff  
  
Gleichgewichtstheorie 180  
Gleichheitsprinzipien 72  
Gleichverteilung 71  
Glückseligkeit 198  
Glücksstatus 198  
Güter 28, 69ff  
Güterbündel 75  
Güterräume 133  
Handlungsanleitung 137  
Hurwicz-Kriterium 37ff  
hypothetische Interpretation 166  
Immunisierung 176  
Individualismus, methodischer 154, 159ff  
-, sozialetischer 161  
individualistischer Ansatz 158  
Information 56  
-, unvollständige 47  
-, vollständige 47  
Informationsbasis 58ff  
Informationsbeschränkung 126  
Informationsmenge 45f  
Inkonsistenz 15  
Institutionen 102  
-, staatliche 170  
Intention 111, 200  
Interaktion 16  
-, einfach reduzierte 184  
-, n-fach reduzierte 184  
-, reduzierte 184  
Interaktionssituation 42ff, 172ff, 222f, 224ff  
-, kooperative 47, 196, 211ff  
-, nicht-kooperative 47, 210ff  
Interdependenzprobleme 93ff  
Interdependenzstrukturen 199  
Interdependenztheorie 112  
Interesse 110, 126  
- Eigeninteresse 170  
- Gemeininteresse 170  
-, gemeinsames 103, 204f  
-, persönliches 111ff, 203f, 209ff, 217  
Interessenaggregation 17, 150 ff  
Interessenausgleich, fairer 103f, 234f  
Interessenkonflikt 73, 158ff  
Interessenkonstellation 103, 235  
Interpersonelle Vergleichbarkeit 55ff  
Intrapersonelle Vergleichbarkeit 59ff

Irrationalität, kollektive 221ff, 246ff  
iterativ 199  
Kardinalität 78  
Kardinalisierbarkeit 86  
kategorisch 165  
Kollektive Präferenzordnung 50ff  
kollektives Urteil 105  
Kolmogoroff-Axiome 30  
Konfliktregelung 163  
Konfliktsituationen 47  
Konnexität 28, 48, 113f  
Konsens 77  
-, praktischer 103  
-, theoretischer 103  
Konsequentialismus 263f  
konsequentialistische Interpretation der Entscheidungstheorie 263f  
Konsequenzen 19, 25f, 73, 139, 144, 211ff, 225 ff  
Konstantsummen-Spiele 47, 177  
konstativ 152  
Kontinuität 28  
Kooperation, soziale 218  
Kooperationsbereitschaft 227  
Koordination, soziale 218  
Korrelation 89  
Laplace-Kriterium 34f, 38f  
Liberales Paradoxon 116ff  
Liberalismus, klassischer 169  
Liberalitätsbedingung 73  
Lösung im strengen Sinne 174  
Lotterie 33, 145  
Marschakbedingungen 84  
Maximin-Kriterium 35ff  
- -Maximin-Kriterium 41  
Maximax-Kriterium 37ff  
Mehr-Personen-Spiele 47  
Meßbarkeitsproblem 74  
Metaethik 164  
Metakriterium 65, 72  
Metapräferenzen 197  
Metapräferenzrealtion, moralische 232, 254  
Metapräferenztheorie 224ff, 239ff  
Metarationalität 197  
Metrisierung von Präferenzen 199  
Minimalmoral 162 ff, 239ff  
Minimax-Kriterium 36ff  
mögliche Welt 17, 21, 48, 84ff  
Modell, behavioristisches 38, 57  
- behavioristisches Präferenz-M. 197  
-, entscheidungstheoretisches 88  
- Jeffreys M. 22  
moralisches System 200 ff  
Moralphilosophie 169f  
-, analytische 155

- Nash-Lösung 174
- nicht-axiologische Ethik 157
  - Theorie 154ff
- Nicht-Konstantsummen-Spiel 47
  - , nicht-kooperative 174ff
- Normalform 45f
- normative Ethik 110
- Nutzen 29, 137ff
  - , absoluter 56, 66ff
  - , abstrakter 147
  - , bedinger 85ff
  - Durchschnittsnutzen 107
  - Gesamtnutzen 68ff
  - Grenznutzen 133
- Nutzenbewertung 233f
  - Einschätzung der N. 88
  - , intrapersonell divergierende individuelle 92
- Nutzendifferenzen 56ff
- Nutzerwartungswert (NEW) 40, 171, 178ff
- Nutzenfunktion 30, 69f, 133, 171, 193f
  - Bernoulli-N. 110
  - Beschränktheit von N. 66f
  - , kardinale 28, 57f, 98
  - , konkave 71
  - , konvexe 70
  - Normierung von Nr. 67, 72
  - , ordinale 28
- Nutzeninterdependenz 84
  - , wechselseitige 89ff
- Nutzenmaximierer, rationaler 181ff
- Nutzenniveau 56ff
- Nutzenniveauveränderung 73
- Nutzentheorie, ökonomische 122
- Nutzenvergleich, intrapersoneller kardinaler 192f
- Nutzenbergleichbarkeit, vollständige interpersonelle 67f
- objektivistische Interpretation 21
- Optimierungsverhalten 177
- Ordnung, erweiterte individuelle 75
  - , soziale 154
- Ordnungsrelation 137
  - , konnexe 48, 145
- Pareto-Regel 64, 79, 95ff, 118ff, 162f
- pareto-inklusiv 172, 214, 216, 244ff
- pareto-optimal 95, 173, 174ff, 208ff, 219
- performativ 152
- Pflicht 169
- PIRPIW-Rationalität 244ff
- Position 75ff
- Präferenzen
  - A-Präferenzen 227ff
  - , abgeleitete 257f
  - , bedingte 127ff
  - Bedingtheit moralischer P. 131



- , fiktive 232
- , höhere Ordnung 135f
- , individuelle 38, 61
- , individuelle resultierende 104f, 217, 235ff
- , intrinsische 131ff, 255ff
- , kollektive 51ff
- , kontradiktorisch bestimmte 82
- Metapräferenzen 197
- , n-ter Ordnung 136
- overall-P. 157, 198
- , persönliche 107
- , propositionale 25, 128ff
- , resultierende 112f, 123ff, 211ff, 237
- , subjektive 55
- , subjektive moralische 216ff, 235ff
- , unbedingte 127ff
- V-Präferenzen 227ff
- Wohlfahrtspräferenzen 55, 120ff, 202ff, 208ff, 219f, 239ff
- , zirkuläre bestimmte 81, 199
- Zustandspräferenzen 25
- zweiter Ordnung 135
- Präferenzenaggregation s. "Aggregation von P."
- Präferenzeninterdependenz 76, 80ff, 97
- Präferenzordnung 16, 38, 139
  - , azyklische kollektive 120
  - , azyklische konnexe 121
  - , individuelle 51f, 240f
  - , kollektive 51f
  - , vollständige 75ff
  - , vollständige subjektive 75ff
  - , vollständige erweiterte subjektive 75ff
  - , zyklische 121
- Präferenzrelation 75ff, 141, 231
  - , antisymmetrische 49f
  - , azyklische 52
  - , kollektive 59f
  - , konnexe 49ff
  - , quasi-transitive 50
  - , relexive 49ff
  - , transitive 49f
  - , zyklische kollektive 119f
- Präferenzstruktur 49f, 172ff, 211, 244ff
- Präferenzzuordnung, kollektive 50ff
- präskriptiv 152ff
- probabilistische Äquivalenz 29
- Propositionen 22ff, 33, 84, 128ff, 263f
- Reflexivität 48
- Regelutilitarismus 195
- Rationalität, kollektive 17, 48ff, 163, 206, 219, 226, 239ff
  - 246ff
  - , kollektive i.e.S. 220f
  - , maximale kollektive 220f
- Rationalitätsbedingungen 49ff, 111

Rationalitätskriterium 16, 243f  
Rationalitätspostulate 29f  
revealed-preference-concept 113, 124, 217  
Risiko, Entscheidung unter 125  
Risikosituation 16, 32, 43, 145  
Sachverhalte 24, 205  
Sätze, normativ-ethische 154  
Selbsteinschätzung 146  
Sicherheitssituation 32  
Sozialethik 17  
-, subjektivistische 78  
-, vertragstheoretische 162  
Sozialwahl-situation 48ff  
Sozialwahltheorie 104f, 129  
Spiele  
- Als-Ob-S. 225ff  
- Altruismus-spiel 226ff  
-, kooperatives 194  
- Konstantsummenspiel 47, 177  
- Mehr-Personen-Spiel 47  
- Nicht-Konstantsummenspiel 47, 174ff  
- Vertrauensspiel 225ff  
- Zwei-Personen-Spiel 47  
Spielbaum 44f, 233  
Spieler 44f  
Spieltheorie 42f, 176  
Sprachphilosophie 204f  
Staatstheorie 170  
Staatsverfassung 155f  
Strategie 43ff, 178ff, 208ff  
Strategienkombination 45f, 233f  
-, erreichbare 184f  
Symmetrieforderung 79  
System, moralisches 200ff  
teleologisch 102, 216, 263f  
Testreihe 212f  
Theorie, axiologische 151ff  
-, nicht-axiologische 154ff  
-, objektivistische 156  
-, subjektivistische 112  
Transferfaktoren 216  
Transformation, lineare 58ff  
-, positiv-monotone 59ff, 78  
Transitivität 48  
Unmöglichkeitstheorem 82  
Unsicherheit, Entscheidung unter 125  
Unsicherheitssituation, 32ff, 43  
Unvergleichbarkeit 65  
Utilitarismus 64ff, 97ff, 105ff, 156, 215  
-, eudämonistischer 100  
-, idealer 100  
Verfassungswahl 154  
Vergleichbarkeit, interpersonelle 55ff

- , intrapersonelle 59ff
- , kardinale 56, 66
- , subjektive (ordinale) interpersonelle 77ff
- Verhalten, irrationales 15, 29, 211
- , inkonsistentes 15
- , konsistentes 28f
- , kooperatives 211, 217, 226ff
- , moralisches 202ff
- NEW-maximierendes 29, 171, 178ff
- , nicht-kooperatives 210ff, 217
- , rationales 28ff, 145, 175, 211, 227
- Verhaltensempfehlung, inkonsistente 183
- Verhaltensregel 200 ff
- , vollständige 184ff
- vernunftgemäßes Handeln 169
- Verpflichtung 137ff
- Prima-facie-Verpflichtung 141
- Verpflichtungskonsistenz 141
- Verpflichtungsurteil 137, 154ff
- Verteilungsgerechtigkeit 126
- Vertrauen 204ff, 226
- , soziales 218
- Vertrauensintervalle 41
- Vertrauensregel 204f
- Vertrauensspiel 225ff
- Voters Paradox 192ff
- Wahrhaftigkeit 204ff
- Wahrscheinlichkeit 15, 26ff, 39ff, 125, 138
- , objektive 33, 41
- , subjektive 34, 133
- , (subjektive) W. von Ereignissen 42, 144
- zweiter Ordnung 41
- Wahrscheinlichkeitsaxiome 137
- Wahrscheinlichkeitsfunktion 29
- Wert, subjektiver 25
- Wertaussage 154
- Werterkenntnis, objektive 164
- Werturteile 129
- Wohlergehen 123
- Wohlfahrtsfunktion 115, 124ff
- , individuelle 61ff, 236
- , soziale 52
- , subjektive 113f
- Wohlfahrtspräferenzen 55, 120ff, 202ff, 208ff, 219f, 235, 239ff
- Wohlfahrts-theorie, ökonomische 157
- Wünschbarkeit 16, 24, 144
- Zufall 44f
- Zustand 17ff, 28, 84ff, 119, 129ff, 219ff, 233f
- Weltzustand 33, 132
- Zustandsmenge 25, 50
- Zwei-Personen-Spiele 47

SYMBOLVERZEICHNIS

$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\Xi$	Menge der Zustände
$Z$	Menge der Anax-Zustände
$K$	Menge der Personen. $i \in K, {}_i \in K$
$E$	Menge der Ereignisse. $\varepsilon \in E$
$U$	Menge der Umstände. $e \in U$
$C$	Menge der Konsequenzen. $c \in C$
$\Sigma^1$	Menge der der Person $i$ offenstehenden Handlungen. $\sigma \in \Sigma^1$ .
$\rho_i$	subjektive Wahrscheinlichkeitsfunktion. $\rho_i(A B)$ bedingte Wahrscheinlichkeit von A bei B.
$\rho(A B)$	bedingte Wahrscheinlichkeit von A bei B $\rho(A B) = \frac{\rho(A \wedge B)}{\rho(B)}$
$u_i$	subjektive Bewertungsfunktion über $\Xi$ (ab §9 resultierende subjektive Bewertungsfunktion)
$v_i$	subjektive Bewertungsfunktion über $Z$
$0_i$	subjektive Bewertungsfunktion über $C$
$u_i$	subjektive Bewertungsfunktion von Handlungen (und Lotterien)
$\pi_i$	subjektive Wohlfahrtsfunktion
$\mu_i$	subjektive moralische Bewertungsfunktion
$R_i$	individuelle schwache Präferenzrelation, auch $>$
$Rw_i$	Wohlfahrtspräferenzrelation $i$
$Rm_i$	subjektive moralische Präferenzrelation
$Rres_i$	resultierende Präferenzrelation
$\tilde{R}_i$	individuelle erweiterte Präferenzordnung
$R$	kollektive schwache Präferenzrelation, auch $f(U)$
$g$	Präferenzstruktur
$f$	kollektive Präferenzzuordnung (bzw. kollektive Wahlregel)
$p$	Position, $P$ Menge der Positionen
$s$	gesellschaftliche Situation, $S$ Menge der Situationen

# **tuduv-Studie**

## **Reihe Sozialwissenschaften**

Wilhelms, Simone

### **Ghetto und Staat**

Der politische Faktor sozialwissenschaftlicher Perspektiven und praktisch-administrativer Lösungsstrategien in der Behandlung gesellschaftlicher Probleme, dargestellt am Beispiel der schwarzen Ghettos in den U.S.A.

1976, Band 1, 210 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 031 8

Tuchhändler, Klaus

### **De Gaulle und das Charisma**

Elemente charismatischer Führung im Gaullismus der V. Republik

1977, Band 4, 384 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 045 8

Johler, Johannes

### **Erziehung zur Autonomie**

Das psychoanalytische Theorem der Abhängigkeitsmotivation (sears) und seine Konsequenzen für eine Erziehung zur Autonomie

1977, Band 5, 209 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 048 2

Hömberg, Erentraud

### **Tourismus**

Funktionen, Strukturen, Kommunikationskanäle

1977, Band 6, 240 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 055 5

Leßner-Abdin, Dietlinde

### **Zur sozialen Lage der Frau in Entwicklungsländern**

Eine Fallstudie: Algerien

1982, Band 7, 2. neuüberarb. Aufl., 210 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 117 9

Kühner, Hans

### **Abenteuer der Dialektik in China**

Die Wandlungen des Basis-Überbau-Begriffs auf dem Weg zur „Kulturrevolution“

1977, Band 8, 345 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 063 6

Höhn, Jan

### **Außenpolitik der EG-Staaten**

Im Fall der KSZE - Geschichte - Struktur - Entscheidungsprozeß - Aktion - Möglichkeiten und Grenzen

1978, Band 9, 391 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 064 4

Nadolny, Sten

### **Abrüstungsdiplomatie 1932/33**

Deutschland auf der Genfer Konferenz im Übergang von Weimar zu Hitler

1978, Band 10, 512 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 066 0

# **tuduv**

Paik, Kyung Nam

**Korea und Japan im Kräftefeld des Nord-West-Pazifik**

Zur Entstehung und Problematik des koreanisch-japanischen Normalisierungsvertrages von 1965

1978, Band 11, 347 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 070 9

Bellers, Jürgen

**Reformpolitik und EWG-Strategie der SPD**

Die innen- und außenpolitischen Faktoren der europapolitischen Integrationswilligkeit einer Oppositionspartei (1957 - 1963)

1979, Band 12, 570 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 077 6

Bailey, Paul J.

**Möglichkeiten der Kontrolle multinationaler Konzerne**

Die Rolle internationaler Organisationen

1981, Band 13, 263 Seiten, 2. Aufl., Manuskriptdruck, Paperback, ISBN 3 88073 104 7

Mempel, Hans Christian

**Die Vermögenssäkularisation 1803/10**

Verlauf und Folgen der Kirchengutenteignung in verschiedenen Territorien, 2 Bände  
1979, Band 15, Teil I Materialien, 437 Seiten; Teil II Text, 269 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 085 7

Hefty, Georg P.

**Schwerpunkte der Außenpolitik Ungarns 1945 - 1973**

Vorgeschichte, Infrastruktur, Orientierung, Interaktionsprozesse

1980, Band 16, 662 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 088 1

Sommerfeld, Reinhard

**Evolution, Kommunikation und Sprache**

Versuch einer Synthese ethologischer und linguistischer Semiotik

1980, Band 17, 480 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 094 6

Hofstetter, Helmut

**Die Leiden der Leitenden**

Zur Pathologie intrapersonaler und interpersoneller Störungen von Führungskräften in Organisationen

1980, Band 18, 292 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 097 0

Probst, Ulrich

**Das Regierungssystem der Republik Malta**

Die Zeit von 1976 bis zur Gegenwart

1980, Band 19, 220 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 100 4

Erlwein, Willy

**Transnationale Kooperation im Alpenraum**

Dargestellt am Beispiel der Arbeitsgemeinschaft der Alpenländer (ARGEAlp)

1981, Band 20, 270 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 112 8

tuduv

Graf, Dietrich

**Bürokratische Erziehung, Strukturen, Funktionen und Folgen**

1982, Band 21, 225 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 122 5

Kmuche, Wolfgang

**Wirtschaftswachstum und Energieverbrauch**

Vergangenheitsentwicklung und Zukunftsperspektiven am Beispiel der Bundesrepublik Deutschland

1982, Band 22, 208 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 123 3

Kompa, Ain

**Gesundheitspolitik und Wohlstandskrankheiten**

Tabak, Alkohol, Ernährung u. Bewegung – Problemfelder ungesunder Lebensformen

1982, Band 23, 682 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 127 6

Yang, Dae-Hyun

**Die Waffenstillstandsverhandlungen in Korea 1951 – 1953**

Eine multiperspektivische Konstellationsanalyse

1982, Band 24, 576 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 126 8

Ng, Hong Chiok

**Staatstheorie von Marx und Ihre Aktualität in unterentwickelten Ländern**

1983, Band 25, 515 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 129 2

Cap, Peter

**Analysen und Modelle auswärtiger Politik**

Die wissenschaftsgeschichtliche Entwicklung vergleichender Studien auf dem Gebiet der Außenpolitik

1983, Band 26, 464 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 133 0

Schmidt, Peter

**Wirkungen der militärischen Entspannung auf die Verteidigungspolitik der NATO-Staaten**

Der Fall MBFR

1983, Band 27, 208 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 136 5

Orland, Nachum

**Die Cherut-Analyse einer rechtsorientierten israelischen Partei**

1983, Band 28, 303 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 137 3

Bader, Monika

**Der sino-amerikanische Normalisierungsprozeß: Vom Dogmatismus zum Pragmatismus**

1983, Band 29, 449 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 141 1

tuduv

Yeh, Young-ming

**Sun Yat-sens Regierung in Südchina von 1917–1922 und die "Bewegung zum Schutz der Verfassung"**

1983, Band 30, 704 Seiten, Manuskriptdruck, Paperback, ISBN 3 88073 140 3

Nida-Rümelin, Julian

**Rationalität und Moralität**

Eine Untersuchung der Bedeutung rationaler Entscheidung für die Sozialethik

1985, Band 31, ca. 360 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 154 3

Schönweiß, Reinhard

**Die Agrarpolitik der Bundesrepublik Deutschland**

Ziele, Maßnahmen und Kritik staatlicher Interventionen auf dem Agrarsektor unter besonderer Berücksichtigung der struktur- und markt-/preispolitischen Beschlüsse

1984, Band 32, 75 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 155 1

Alms, Wolfgang

**Der Einfluß von Zeitdruck auf das Problemlösen in Abhängigkeit von Selbstkonzept, Attribuierungstendenz, Geschlecht und Schultyp**

1984, Band 33, 475 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 158 6

von Törne, Monika

**Grenzen staatlicher Schulreform**

Fallstudie: Die Reform der Hauptschule in Bayern

1985, Band 34, ca. 455 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 161 6

Imle, Walter

**Zwischen Vorbehalt und Erfordernis**

Eine historische Studie zur Entstehung des nachrichtendienstlichen Verfassungsschutzes nach 1945.

1984, Band 35, 325 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 166 7

Seidl, Otmar

**Die Territorialität des Menschen**

Untersuchungen über einige biologische Grundlagen sozialen Verhaltens

1985, Band 36, 465 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 178 0

Tworek-Müller, Elisabeth

**Kleinbürgertum und Literatur**

Zum Bild des Kleinbürgers im bayerischen Roman der Weimarer Republik

1985, Band 37, 357 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 179 9

Hennies, Werner

**Die politische Theorie August Ludwig von Schlözers zwischen Aufklärung und Liberalismus**

1985, Band 38, 308 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 180 2

tuduv



Mörth, Ingo

**Lebenswelt und religiöse Sinnstiftung.**

Ein Beitrag zur Theorie des Alltagslebens

1986, Band 39, 228 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 192 6

Phocas, Andreas

**Biologische Aspekte politischen Verhaltens**

1986, Band 40, 228 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 202 7

Vogel, Beatrix

**Der primärtherapeutische Begriff der Erfahrung**

Versuch einer erkenntnistheoretisch-forschungslogischen Begründung  
der Primärtherapie Arthur Janovs

1987, Bd. 41, 558 Seiten, Manuskriptdruck, Pb., ISBN 3 88073 227 2

**tuduv**