

**Duisburger Volkswirtschaftliche Schriften**

---

**Band 33**

# **Das vertragstheoretische Argument in der Ökonomik**

**Von**

**Christian Müller**



**Duncker & Humblot · Berlin**

DOI <https://doi.org/10.3790/978-3-428-50084-0>

Generated for Hochschule für angewandtes Management GmbH at 88.198.162.162 on 2025-12-20 04:34:49

FOR PRIVATE USE ONLY | AUSSCHLIESSLICH ZUM PRIVATEN GEBRAUCH

CHRISTIAN MÜLLER

Das vertragstheoretische Argument  
in der Ökonomik

# Duisburger Volkswirtschaftliche Schriften



Herausgeber:

Prof. Dr. Manfred Tietzel (geschäftsführend)  
Prof. Dr. Dieter Cassel · Prof. Dr. Helmut Cox  
Prof. Dr. Günter Heiduk · Prof. Dr. Ullrich Heilemann  
Prof. Dr. Carsten Herrmann-Pillath · Prof. Dr. Dietmar Kath  
Prof. Dr. Werner Pascha · Prof. Dr. Hans-Joachim Paffenholtz  
Prof. Dr. Josef Schira · Prof. Dr. Klaus Tiepelmann

Band 33

# Das vertragstheoretische Argument in der Ökonomik

Von  
Christian Müller



Duncker & Humblot · Berlin



Gedruckt mit Unterstützung der Deutschen Forschungsgemeinschaft

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

**Müller, Christian:**


Das vertragstheoretische Argument in der Ökonomik /  
von Christian Müller. – Berlin : Duncker und Humblot, 2000  
(Duisburger Volkswirtschaftliche Schriften ; Bd. 33)  
Zugl.: Duisburg, Univ., Diss., 1999  
ISBN 3-428-10084-0

D 464

Alle Rechte vorbehalten

© 2000 Duncker & Humblot GmbH, Berlin  
Fotoprint: Berliner Buchdruckerei Union GmbH, Berlin  
Printed in Germany

ISSN 0936-7020  
ISBN 3-428-10084-0

Gedruckt auf alterungsbeständigem (säurefreiem) Papier  
entsprechend ISO 9706 

## **Geleitwort der Herausgeber der Duisburger Volkswirtschaftlichen Schriften**

Der erste Band der Duisburger Volkswirtschaftlichen Schriften erschien im September 1987 im S+W Steuer- und Wirtschaftsverlag, Hamburg. In mehr als zehn Jahren vertrauensvoller und erfolgreicher Zusammenarbeit mit unserem damaligen Verleger und Kollegen, Herrn Professor Dr. Lothar Haberstock, folgten weitere 31 Veröffentlichungen in dieser Schriftenreihe. Nach dem plötzlichen Tod unseres geschätzten Kollegen im Januar 1996 sah sich seine Rechtsnachfolgerin veranlaßt, den S+W Steuer- und Wirtschaftsverlag im März 1998 auf den Erich Schmidt Verlag, Berlin, zu übertragen.

Obwohl das Programm des Erich Schmidt Verlages einen sehr deutlichen Schwerpunkt in der Veröffentlichung betriebswirtschaftlicher Schriften hat, trat er in den mit dem S+W Steuer- und Wirtschaftsverlag geschlossenen Herausgebervertrag ein. Da in Vertragsbeziehungen, ganz wie in Ehen, Entgegenkommen und Generosität hilfreiche, aber erst gemeinsame Ziele und Interessen hinreichende Bedingungen für ein beiderseitig förderliches Miteinander sind, wurde dieser Vertrag in gegenseitigem Einvernehmen gelöst, als sich die Möglichkeit bot, die Duisburger Volkswirtschaftlichen Schriften im Verlag Duncker & Humblot, Berlin, fortzuführen, in dessen Verlagsprogramm sich unsere Schriftenreihe nahtlos einfügt.

Wir sind zuversichtlich, daß der vorliegende 33. Band der Duisburger Volkswirtschaftlichen Schriften die gute Tradition der Reihe fortsetzt und den Anfang einer weniger bewegten und zugleich wissenschaftlich wie verlegerisch erfolgreichen Zukunft markiert.

Duisburg, im September 1999

Für die Herausgeber  
*Manfred Tietzel*



## Geleitwort

Versuche, die Entstehung und den Inhalt von Regeln zu erklären, nach denen die Mitglieder einer Gruppe ihr Zusammenleben und ihr gemeinsames Handeln gestalten, haben eine lange Geschichte; markante Höhepunkte dieser Geschichte stellen die Theorien des Gesellschaftsvertrages des 17. und 18. Jahrhunderts dar, wie sie etwa von Hobbes, Locke oder Rousseau entwickelt wurden. Während aber die klassischen Theorien des Gesellschaftsvertrages nicht nur erklärende Funktionen hatten, sondern, etwa in der amerikanischen Unabhängigkeitsbewegung, der Französischen Revolution von 1789 und den europäischen Revolutionen des Jahres 1848, auch zur Legitimierung bestimmter Verfassungsentwürfe in politische Dienste genommen wurden, schienen sie seither und bis in die jüngere Vergangenheit nur noch unter theoriegeschichtlichen Aspekten von Interesse zu sein.

Eine wohl ganz unerwartete Renaissance erfuhren vertragstheoretische Argumente erst in den letzten Jahrzehnten, in denen die „neuen Vertragstheoretiker“, wie sie Scott Gordon (1976) nannte, auf der Basis expliziter individualistischer Entscheidungstheorien genaue und allgemeine Theoreme über das Zustandekommen und die Legitimität solcher Regeln zu entwickeln suchten. „Ökonomisch“ kann man insbesondere zwei Vertragstheorien, die von John Rawls und jene von James Buchanan, nennen, insofern sie die üblichen *Homo-oeconomicus*-Prämissen, darunter die Rationalverhaltensannahme, verwenden. Gerade diese beiden Ansätze sind nicht nur auf ein breites Interesse gestoßen, das, weit über die Wirtschaftswissenschaften hinaus, in alle Sozialwissenschaften reichte, sondern sie haben auch eine umfangreiche Sekundärliteratur nach sich gezogen und sind bis zu der Dignität gereift, über ein eigenes wissenschaftliches Forum, die Zeitschrift „Constitutional Political Economy“, zu verfügen.

Der genauen Analyse der vertragstheoretischen Argumente von Rawls und Buchanan ist die vorliegende Untersuchung von Herrn Dr. Christian Müller, „Das vertragstheoretische Argument in der Ökonomik“, gewidmet. Zwei wesentliche Problemkomplexe stehen im Vordergrund dieser Untersuchung. Zum einen liefert sie eine methodologische Rekonstruktion des vertragstheoretischen Argumentes, durch die insbesondere die Art seines Geltungsanspruches und sein Geltungsbereich präzisiert und geklärt werden. Das zweite Anliegen der Arbeit besteht darin, u.a. mit Hilfe der Spieltheorie die Frage zu beantworten,



ob zutrifft, daß die - von Rawls und Buchanan verbal behaupteten und graphisch formulierten - Theoreme aus den jeweiligen Prämissen folgen und ob sie es mit der behaupteten Allgemeinheit tun.

Die Ergebnisse der Untersuchungen von Herrn Dr. Müller sind in vieler Hinsicht neuartig und überraschend; es kann als sicher gelten, daß sie auf die weitere Entwicklung der ökonomischen Vertragstheorie einigen Einfluß ausüben werden.

Nicht zuletzt aus diesem Grunde haben die Herausgeber der Duisburger Volkswirtschaftlichen Schriften diese Untersuchung, die im Sommersemester 1999 vom Fachbereich Wirtschaftswissenschaften der Gerhard-Mercator-Universität Duisburg als Dissertation angenommen wurde, gerne in die Schriftenreihe aufgenommen.

Duisburg, im September 1999

Für die Herausgeber  
*Manfred Tietzel*

## Vorwort

Die vorliegende Arbeit wurde im Sommersemester 1999 im Fachbereich Wirtschaftswissenschaften der Gerhard-Mercator-Universität Duisburg als Dissertation angenommen. Sie entstand zwischen November 1994 und Januar 1999 während meiner Tätigkeit als wissenschaftlicher Mitarbeiter an den Fachgebieten Wirtschaftswissenschaft, Methodologie und Finanzwissenschaft.

Während dieser Zeit genoß ich das Privileg einer in jeder Hinsicht idealen Promotionsbetreuung durch meinen Chef und Doktorvater, Herrn Prof. Dr. Manfred Tietzel. Herr Professor Tietzel hat mein Dissertationsprojekt stets mit größtem Interesse und Sachverstand begleitet und gefördert. Als Diskussionspartner und freundschaftlicher Ratgeber stand er mir jederzeit zur Verfügung. Für die schöne und vertrauensvolle Zeit der Zusammenarbeit mit ihm, aus der ich weit mehr mitnehme als eine solide theoretische Ausbildung, danke ich ihm herzlich.

Herrn Prof. Dr. Dieter Cassel bin ich sehr dankbar, daß er trotz vielfacher anderer Verpflichtungen ohne zu zögern bereit war, die Arbeit als Zweitgutachter zu übernehmen. Dank schulde ich auch den Herren Privatdozent Dr. Thomas Apolte und Prof. Dr. Volker Breithecker für die Mitwirkung in der Promotionskommission.

Dankbar bin ich für die Unterstützung, die ich aus dem Fachbereich Philosophie unserer Hochschule erfuhr. Wesentliche Grundlagen des ökonomischen Rationalverhaltensansatzes und der Spieltheorie lernte ich von den Herren Prof. Dr. Hartmut Kliemt und Dr. Bernd Lahno, die sich viel Zeit nahmen, die Probleme meiner Arbeit mit mir zu diskutieren. Auf diese Weise befreiten sie mich aus mancher theoretischen Sackgasse.

Ebenso danke ich Frau Dipl.-Math. Dipl.-Volksw. Birgit Hofmann für ihre Bereitschaft, wesentliche Teile meiner Arbeit kritisch mit mir zu diskutieren.

Für zahlreiche Diskussionen und freundschaftliche Zusammenarbeit danke ich meinen Mitarbeiterkollegen, insbesondere Privatdozent Dr. Gregor van der Beek, Dipl.-Volksw. Kornelia van der Beek, Dr. André Kuck, Dipl.-Volksw. Elsbeth Kuck, Dr. Siegfried Rauhut, Dipl.-Volksw. Raphaela Smarzczyk, Dr. Michael Terporten und Dipl.-Volksw. Thomas Wilke.

Mehrere „Generationen“ studentischer Hilfskräfte versorgten mich unermüdlich mit Lesestoff und waren mir technisch wie organisatorisch eine große Hilfe. Dafür danke ich Björn Bosch, Markus Brüggemann, Dipl.-Volksw. Sandra van Crooy, Dipl.-Volksw. Christoph Gwosc, Constanze Korte, Christian Krahe, Jochen Krölls, Dipl.-Kfm. Alexander Missal, Dipl.-Kfm. Carsten Neises, Dipl.-Kff. Petra Schiffmann, Matthias Schramm, Dipl.-Betriebsw. Marc Sturm, Sandra Thomas, Dipl.-Ök. Christoph Tillmanns, Dipl.-Ök. Lars Tutt und Kerstin Wolter. Ebenso danke ich unseren Sekretärinnen, Frau Heide Kessler und Frau Gisela Neugebauer, für langjährige angenehme Zusammenarbeit.

Der Deutschen Forschungsgemeinschaft, Bonn, danke ich für die freundliche Übernahme der Kosten der Drucklegung meiner Arbeit.

Ein unschätzbarer Halt waren und sind mir meine Eltern, Dr. Rolf und Renate Müller. Für ihre selbstverständliche, immerwährende und überaus großzügige Unterstützung in allen Lebenslagen bin ich sehr dankbar.

Nicht zuletzt danke ich meiner Frau Susanne, die stets vom Gelingen meiner Arbeit überzeugt war und diese - trotz größter eigener beruflicher Belastung als Lehrerin - tatkräftig gefördert hat. Ihr und unserer kleinen Tochter Sophie widme ich diese Arbeit.

Duisburg, im Februar 2000

*Christian Müller*

## Inhaltsverzeichnis

Einleitung .....	15
§ 1 Das vertragstheoretische Argument .....	25
A. Individualismus .....	25
B. Einstimmigkeit .....	26
I. Politik als Tausch .....	26
II. Einstimmigkeit und Verfahrensgerechtigkeit .....	31
C. Hypothetische Einstimmigkeit .....	32
I. Der Proberstein der Rechtmäßigkeit .....	32
II. Der methodologische Status des vertragstheoretischen Arguments .....	37
III. Fiktive Verträge und faktische Verpflichtungen .....	46
§ 2 Die starke Vertragstheorie .....	49
§ 3 Die schwache Vertragstheorie (I.): Die Situationsannahmen .....	58
A. Rawls' „Theorie der Gerechtigkeit“ und der „Schleier des Nichtwissens“ ....	60
B. Buchanan's Konstitutionenökonomik und der „Schleier der Unsicherheit“ ....	69
§ 4 Die schwache Vertragstheorie (II.): Die Konklusion .....	83
A. Die Gerechtigkeitskonzeption: Das Entscheidungskriterium unter Unsicherheit .....	86
I. Maximin-Gerechtigkeit .....	87
II. Laplace-Gerechtigkeit .....	96
B. Konsensinduzierung: Der Universalisierungsgrad .....	103
I. Partielle Universalisierung: Der Tausch von Positionen .....	104
II. Vollkommene Universalisierung: Der Tausch von Präferenzen .....	110
C. Ergebnis .....	118



§ 5 Diskriminierung und Dissens unter dem Schleier der Unkenntnis .....	120
A. Der Schleier der Unkenntnis und das Hobbessche Ordnungsproblem .....	120
B. Verfassungswahl unter dem Schleier der Unkenntnis .....	127
I. Das postkonstitutionelle Spiel .....	128
1. Das Zwei-Personen-Gefangenendilemma .....	128
2. Das $n$ -Personen-Gefangenendilemma .....	130
II. Das konstitutionelle Spiel .....	131
C. „Gerechtigkeit als Unfairneß“: Das konstitutionelle Spiel unter Rawls’ Schleier des Nichtwissens .....	136
D. Das konstitutionelle Spiel unter Buchanans Schleier der Unsicherheit .....	147
E. Bedingungen einer fairen Einigung .....	161
I. Fairneß und Dominanz .....	161
II. Fairneß und Pareto-Dominanz .....	163
III. Fairneß und Koordination .....	166
§ 6 Fairneß und Einstimmigkeit unter sicheren Erwartungen .....	171
A. Das konstitutionelle Zwei-Personen-Spiel unter perfekter Information .....	172
B. Das konstitutionelle $n$ -Personen-Spiel unter perfekter Information .....	176
C. Diskussion .....	179
I. Warum soviel Dissens? .....	180
II. Eine partielle Rehabilitierung des Schleiers .....	183
§ 7 Schluß .....	185
Literaturverzeichnis .....	192
Sachwortregister .....	211

## Abbildungsverzeichnis

Abbildung 1.1	Die logische Struktur der vertragstheoretischen Als-ob-Erklärung/ Prognose .....	38
Abbildung 2.1:	Nutzendiagramm im Zwei-Personen-Fall .....	54
Abbildung 5.1:	Zwei-Personen-Gefangenendilemma $G_2$ .....	128
Abbildung 5.2:	Die Struktur des konstitutionellen Spiels $K_2$ unter dem Schleier der Unkenntnis (Zwei-Personen-Fall) .....	132
Abbildung 5.3:	$G_2$ -Szenarien 1 und 2 .....	138
Abbildung 5.4:	Konstitutionelles Spiel $K_2^R$ unter Rawls' Schleier des Nicht- wissens .....	139
Abbildung 5.5:	Einstimmige Wahl der diskriminierenden Regel $\bar{r}_{d1}$ im konstitu- tionellen Spiel $K_2^R$ unter Rawls' Schleier des Nichtwissens .....	141
Abbildung 5.6:	Einstimmige Wahl der fairen Regel $\bar{r}_f$ im konstitutionellen Spiel $K_2^R$ unter Rawls' Schleier des Nichtwissens .....	145
Abbildung 5.7:	$G_2$ -Szenarien 1 und 2 .....	150
Abbildung 5.8:	Konstitutionelles Spiel $K_2^B$ unter Buchanans Schleier der Unsicherheit .....	151
Abbildung 5.9:	Einstimmige Wahl der diskriminierenden Regel $\bar{r}_{d2}$ im konstitu- tionellen Spiel $K_2^B$ unter Buchanans Schleier der Unsicherheit .....	152
Abbildung 5.10:	GD- und konstitutionelle Präferenzen, die die Wahl einer dis- kriminierenden Regel $\bar{r}_d$ im konstitutionellen Spiel $K_{1000}^B$ unter dem Schleier der Unsicherheit implizieren .....	154
Abbildung 5.11:	Gefangenendilemma- und konstitutionelle Präferenzen für Grup- pen a und b, die Dissens über die Wahl irgendeiner Regel unter dem Schleier der Unsicherheit implizieren (Zahlenbeispiel für $n = 10$ ) .....	156

Abbildung 5.12: Nichteinigung zwischen Mitgliedern von Gruppe a und Gruppe b im konstitutionellen Spiel unter dem Schleier der Unsicherheit (Zahlenbeispiel für $n = 10$ ) .....	157
Abbildung 5.13: Einstimmige Wahl der fairen Regel $\bar{r}_f$ im konstitutionellen Spiel $K_2^B$ unter Buchanans Schleier der Unsicherheit .....	159
Abbildung 5.14: GD- und konstitutionelle Präferenzen, die die Wahl der fairen Regel $\bar{r}_f$ im konstitutionellen Spiel $K_{200}^B$ unter dem Schleier der Unsicherheit implizieren .....	160
Abbildung 5.15: GD-Szenarien und konstitutionelles Spiel unter Buchanans Schleier der Unsicherheit als reines Koordinationsspiel (Fall 3) für $n = 2$ .....	167
Abbildung 6.1: Konstitutionelles Zwei-Personen-Spiel $K_2$ unter perfekter Information .....	173
Abbildung 6.2: Zwei-Personen-Gefangenendilemma mit Rawls- und Buchanan-Indifferenzkurven (graphisch) .....	175

## Einleitung

Die Frage, welche Handlung eines Herrschers gerecht und legitim sei, beschäftigt die Staatsphilosophie bereits seit Jahrtausenden. Schon bei einigen Philosophen der griechischen Antike finden sich Ansätze, die Legitimität staatlichen Handelns auf die Zustimmung der von ihm Betroffenen zurückzuführen. Gerechtigkeit ist nach dieser Auffassung aus dem individuellen Interesse abgeleitet und gründet sich, wie andere später explizit formulierten, auf einen Vertrag aller Betroffenen. „Gerechtigkeit an sich“, so wandte Epikur gegen die Platonische Ideenlehre ein, „hat es nie gegeben. Alles Recht beruhte vielmehr stets nur auf einer Übereinkunft zwischen Menschen, die sich in jeweils verschiedenen großen Räumen zusammenschlossen und sich dahin einigten, daß keiner dem anderen Schaden zufügen oder von ihm erleiden soll.“<sup>1</sup>

Dieser Begriff konsensualer Gerechtigkeit bildet den Kern der *Theorien des Gesellschaftsvertrags*, wie sie seit Jahrhunderten - von Thomas Hobbes, Jean-Jacques Rousseau oder John Locke<sup>2</sup> - vertreten werden. Er ist auch kennzeichnend für die ökonomisch argumentierenden Rechtfertigungsargumente der sog. „neuen Vertragstheoretiker“<sup>3</sup> John Rawls und James Buchanan, deren „Theorie der Gerechtigkeit“ bzw. ökonomische Theorie der Verfassung im Vordergrund der vorliegenden Arbeit stehen.<sup>4</sup> Alle diese Ansätze sind Spielarten des gleichen kontrakttheoretischen Grundgedankens.

Mit ihrer Betonung des Konsenses zwischen allen Gesellschaftsmitgliedern lenken die Vertragstheorien ihr Augenmerk von der postkonstitutionellen Per-

---

<sup>1</sup> Epikur (1949), Hauptlehrsatz 33, S. 61. Vgl. auch Hauptlehrsatz 31, S. 60: „Das der Natur gemäße Recht ist eine den Nutzen betreffende Übereinkunft, einander nicht zu schädigen noch voneinander Schaden zu leiden.“

<sup>2</sup> Vgl. Hobbes (1976 [1651]); Locke (1992 [1689]); Rousseau (1977 [1762]). Die Literatur zur Vertragstheorie ist unüberschaubar. Vergleichende Gesamtdarstellungen geben etwa Gough (1957); Lessnoff (1986); Kersting (1994). Gute Einblicke in die Grundideen der Vertragstheorien geben Ballestrem (1983); Vanberg (1994).

<sup>3</sup> Gordon (1976).

<sup>4</sup> Vgl. grundlegend Rawls (1994 [1971]) sowie Buchanan/Tullock (1962) und Buchanan (1984 [1975]). Gesamtdarstellungen der Rawlsschen Theorie finden sich z.B. bei Wolff (1977); Kley (1989); Kukathas/Pettit (1990); Kersting (1993); Pogge (1994). Einen Überblick über die von Buchanan begründete „Konstitutionenökonomik“ geben Buchanan (1987), ders. (1990); Leschke (1993); Pies/Leschke (Hrsg.; 1996); Mueller (1996).



spektive, wie sie für die ökonomische Theorie ansonsten üblich ist, auf die konstitutionelle Analyse der Politik.<sup>5</sup> Die *postkonstitutionelle* Sichtweise betrachtet Wahlhandlungen *innerhalb* exogen vorgegebener Restriktionen, die natürlicher oder historischer Art, von anderen Personen, durch Sitte oder Gewohnheit gesetzt sein können. Die *konstitutionelle* Sichtweise berücksichtigt demgegenüber die Tatsache, daß Individuen, soweit sie über *Selbstbindungsfähigkeit* verfügen, auch Wahlentscheidungen *zwischen* alternativen Restriktionen treffen können, unter denen sie später selbst handeln. Gegenstand von Wahlhandlungen sind dann gegenseitige Selbstbeschränkungsversprechen in Form von *Regeln*, die sich die Entscheidungsträger auf der konstitutionellen Ebene - dem „*Urzustand*“<sup>6</sup> - selbst geben.<sup>7</sup> Der Blickwinkel verlagert sich insoweit auf jene Institutionen, die in der postkonstitutionellen Gesellschaft den „Datenkranz“ der individuellen Entscheidungen darstellen. Bewertungsmaßstab sind nicht mehr nur die postkonstitutionell relevanten Handlungsinteressen der Beteiligten, sondern ihre „konstitutionellen Interessen“ an einer Selbstbindung durch Regeln.<sup>8</sup>

Eine beliebte Metapher zur Verdeutlichung beider Perspektiven ist die eines Spiels oder Wettkampfs<sup>9</sup>: Während des Spiels stellen die Spielregeln die Grenzen legitimer Spielzüge dar und dienen als Daten, die das Verhalten der Spieler in mehr oder weniger vorhersagbarer Weise steuern; weil sie „quasi-permanent“ gelten, liefern die Regeln den Spielern eine Information, an die sie ihr Verhalten anpassen können.<sup>10</sup> Wechselt man jedoch auf die konstitutionelle Entscheidungsebene, so läßt sich analysieren, wie sich der Spielablauf und die Spielergebnisse ändern, wenn die Regeln in bestimmter Hinsicht geändert werden. (Spiel-)Regeln gelten mithin nicht uneingeschränkt und sind, in einer prägnanten, auf Frank Knight zurückgehenden Formulierung Buchanans, lediglich „relativ absolute Absolutheiten“<sup>11</sup>.

<sup>5</sup> Diese Terminologie wird v.a. in der *Buchananschen* Konstitutionenökonomik verwendet; vgl. *Buchanan/Tullock* (1962), S. 110 f.; *Buchanan* (1972), S. 126; *Buchanan/Bush* (1977 [1974]), S. 186; *Brennan/Buchanan* (1993 [1985]), S. 7 ff. Die Unterscheidung einer konstitutionellen und einer postkonstitutionellen Ebene findet sich aber (wie auch *Buchanan*, 1972, S. 126, anmerkt) ebenso bei *Rawls* in seiner Beschränkung auf den sozialen Regelrahmen, den er als die „Grundstruktur“ einer Gesellschaft bezeichnet; vgl. *Rawls* (1994 [1971]), v.a. Abschnitt 2, S. 23-17; *ders.* (1992 [1977]).

<sup>6</sup> Den *Buchananschen* Begriff der „konstitutionellen Ebene“ und den *Rawls'schen* Begriff des „Urzustandes“ verwende ich hier und im folgenden synonym.

<sup>7</sup> Vgl. *Buchanan* (1987), S. 585, *ders.* (1990), S. 2 f.; *Vanberg* (1990), S. 16.

<sup>8</sup> Vgl. grundlegend zur Unterscheidung von Handlungs- und konstitutionellen Interessen *Vanberg/Buchanan* (1988). Vgl. auch *Petersen* (1996), Zweiter Teil.

<sup>9</sup> Vgl. z.B. *Brennan/Buchanan* (1988 [1980]), S. 3; *dies.* (1993 [1985]), S. 7-9.

<sup>10</sup> Vgl. *Brennan/Buchanan* (1983), S. 46.

<sup>11</sup> *Buchanan* (1989).

Eine einstimmige Wahl von Regeln stößt in modernen Millionengesellschaften jedoch auf zwei grundlegende Probleme: *Erstens* setzt die Verwendung der Konsensregel als Legitimationsnorm die *Nichtexistenz von Konsensfindungs- und Informationskosten* voraus. Das Einstimmigkeitsprinzip ist zum einen nur dann eine kosteneffiziente Kollektiventscheidungsregel, wenn die Kosten der Anbahnung und des Abschlusses des Gesellschaftsvertrags vernachlässigbar sind<sup>12</sup>; andernfalls wäre die effiziente Lösung, auf eine einstimmige Legitimation der Kollektivhandlungen zu verzichten. Zum anderen erfordert die einstimmige Beschlußfassung über alle Belange der Politik, daß die Bürger hinreichend über den Abstimmungsgegenstand informiert sind; diese Kosten der Information dürfen für keine Vertragspartei prohibitiv sein.

In realen Gesellschaften vieler Millionen Individuen ist diese Voraussetzung des Abschlusses von Gesellschaftsverträgen in der Regel aber nicht erfüllt. Aus der Geschichte ist kein Beispiel bekannt, daß es je eine politische Handlung gegeben hätte, welche die einmütige Zustimmung aller Bürger erhalten hätte. Der Gesellschaftsvertrag, über den die großen Staatstheoretiker so lange nachdachten, ist eine Fiktion. Moderne Massengesellschaften sind geprägt durch das „Faktum des Pluralismus“<sup>13</sup>; die Hoffnung auf reale Einigkeit ist utopisch. Schon früh haben sich die meisten Vertragstheoretiker daher auf die Untersuchung der Frage zurückgezogen, auf welche Regeln ihres Zusammenlebens sich denn wenigstens fiktive Rationalentscheider einigen würden. An die Stelle einer tatsächlichen Rechtfertigung durch einen realen Konsens tritt hier die (Quasi-) Legitimation durch einen *hypothetischen Gesellschaftsvertrag*; statt Zustimmung aller verlangen diese Theorien lediglich die *Zustimmungsfähigkeit* durch alle Betroffenen.

Ein *zweites* Problem einer einstimmigen Legitimation praktischer Politik ist die Frage, ob und inwieweit einstimmige Entscheidungen unabhängig von den Bedingungen, unter denen sie getroffen wurden, zur Rechtfertigung von Staatshandeln herangezogen werden können. Als Voraussetzung für die Akzeptanz einstimmiger Regelentscheidungen als Legitimationsquelle verlangen die meisten - gleichwohl nicht alle - Theorien des Gesellschaftsvertrags, daß die *Entscheidungssituation selbst gerecht* zu sein habe. Im allgemeinen werden Entscheidungen, die unter Zwang getroffen wurden, nicht als legitimationsstiftend anerkannt. Andere Theoretiker gehen noch einen wesentlichen Schritt weiter und wollen nur solche Verträge als gerecht akzeptieren, die unter der Bedingung formaler rechtlicher oder sogar sozioökonomischer Gleichheit zustande gekommen sind.

---

<sup>12</sup> Vgl. *Buchanan/Tullock* (1962) sowie unten § 3 B.

<sup>13</sup> *Rawls* (1992 [1987]), S. 294.

Die „Theorie der Gerechtigkeit“ von Rawls und Teile des Buchanan'schen Werkes unterstellen statt dessen, daß die betrachteten Individuen unter einem „Schleier des Nichtwissens“ (Rawls) bzw. einem „Schleier der Unsicherheit“ (Buchanan) - für beide Annahmen werde ich den Oberbegriff „*Schleier der Unkenntnis*“ verwenden - entscheiden, der sie in Unkenntnis darüber versetzt, in welcher sozialen Position sie selbst sich in der postkonstitutionellen Gesellschaft befinden und welche daraus resultierenden Partialinteressen sie haben werden. Der Grad des Nichtwissens von Entscheidungsträgern über ihre individuellen Eigeninteressen - so die Grundidee - ist positiv korreliert mit der *Unparteilichkeit* und der *Konsensfähigkeit* der erzielten Abstimmungsergebnisse. Je unsicherer die Beteiligten im Zeitpunkt der konstitutionellen Abstimmung über ihre eigene Stellung in der Gesellschaft und die daraus resultierenden Partialinteressen sind, desto wahrscheinlicher werden faire und zugleich einstimmige Entscheidungen sein. Ein Beispiel mag dies verdeutlichen.

Wollten alle Mitglieder einer Gesellschaft einstimmig darüber entscheiden, ob und nach welcher Regel das in ihrer Volkswirtschaft erwirtschaftete Sozialprodukt unter ihnen umverteilt werden solle, so würden sie sich vermutlich niemals einigen. Denn die Bedürftigen unter ihnen würden für, die Wohlhabenden - als potentielle Geber - gegen eine solche Redistribution votieren. Nach dem zentralen ökonomischen Wohlfahrtskriterium ist ein Verteilungskampf wie dieser sogar als paretooptimal anzusehen, weil die Wohlfahrt der Armen sich durch Umverteilung nur dann erhöhen ließe, wenn den Reichen Teile ihrer Einkommen und Vermögen entzogen, diese also schlechter gestellt würden. Allein dann, wenn alle potentiellen Geber Altruisten wären, in deren Nutzenfunktionen das Wohl der Armen als Argument eingeht, oder die Reichen aus anderen Gründen ein Eigeninteresse hätten, einer Umverteilung zuzustimmen, wäre eine „paretooptimale Umverteilung“<sup>14</sup> denkbar.

Ein ganz anderes Ergebnis wäre aber zu erwarten, wenn dieselben Individuen zusammensäßen, bevor sie wissen, ob sie einmal reich oder arm sein werden. Kann am Start des „Spiels des Lebens“ noch niemand ahnen, ob er sein Glück machen wird oder nicht, wäre ein *einstimmiges* Ergebnis nicht unwahrscheinlich. Da niemand weiß, ob er selbst einmal zu den Wohlhabenden oder den Hungerleidern gehören wird, wird er mit seiner Stimme weder die einen noch die anderen bevorzugen: Jeder wird, um sich selbst gegen den Fall der Armut zu „versichern“, aus eigenem Interesse für eine Umverteilung votieren, die den Armen ein auskömmliches Dasein garantiert und die gleichwohl die Reichen - in deren Rolle sich ja jeder selbst einmal wiederfinden könnte - nicht zu sehr schröpft. Da das fehlende Wissen der Entscheidungsträger über ihre eigenen

---

<sup>14</sup> Vgl. Hochman/Rodgers (1969) für den altruistischen, Brennan (1973) für den nicht-altruistischen Fall paretooptimaler Umverteilung.

künftigen Partialinteressen jede Art von positiver oder negativer Diskriminierung unmöglich macht, wird man diese Entscheidung im allgemeinen auch als „fair“ bezeichnen können. Da die Umverteilungsregel im *eigenen* Interesse *aller* Gesellschaftsmitglieder liegt, stellt die Vereinbarung eines solchen rudimentär umverteilenden Sozialstaats unter unsicheren Erwartungen eine (konstitutionelle) Pareto-Verbesserung gegenüber der marktlichen Primärverteilung dar.

Um den aufgezeigten Konflikt zwischen (konstitutionellen) Wohlfahrtsverbesserungen nach dem üblichen Pareto-Kriterium und Pareto-Verbesserungen unter einem Schleier der Unkenntnis, wie sie sich in den Theorien von Rawls und Buchanan darstellen, kreist die vorliegende Arbeit. Zwar verwenden auch andere Theorien wie jene von William Vickrey, John C. Harsanyi und Paul A. Samuelson die Annahme eines Schleiers der Unkenntnis<sup>15</sup>; da diese Ansätze, die sich allesamt dem Utilitarismus zurechnen, keine Vertragstheorien darstellen und - mit Ausnahme von Harsanyis Ethik - auch nicht weiter ausformuliert sind, werde ich mich im folgenden allein mit den Theorien des hypothetischen Gesellschaftsvertrags von Rawls und Buchanan auseinandersetzen.

Zwei Einschränkungen sind jedoch bereits an dieser Stelle erforderlich: Es wird im folgenden nicht der Anspruch erhoben, die Rawlsche oder die Buchanansche Theorie auch nur ansatzweise erschöpfend zu behandeln. Die Rawlsche Theorie wäre für ein solches Unternehmen viel zu komplex. Ich werde mich daher auf die für Ökonomen besonders interessanten sozioökonomischen Aspekte der „Theorie der Gerechtigkeit“ und damit, wie in ökonomischen Analysen dieses Ansatzes üblich, auf den zweiten Teil des zweiten Rawlschen Gerechtigkeitsprinzips beschränken, das berühmte Differenzprinzip, nach dem die Verteilungspolitik am Wohl des jeweils am schlechtesten gestellten Gesellschaftsmitglieds auszurichten ist.<sup>16</sup> Auch werden allein die vertragstheoretischen Argumentationsstränge verfolgt; eine Analyse der kohärenztheoretischen Normbegründungsstrategie von Rawls mit seiner Methode des sog. „Überlegungsgleichgewichts“ wird dagegen unterbleiben.<sup>17</sup>

Buchanans Vertragstheorie ist für eine umfassende Analyse zu widersprüchlich. So erscheint es kaum möglich, *die* Buchanansche Position - so es sie überhaupt gibt - zu ermitteln. Auf der einen Seite lehnt er den Utilitarismus und die moderne Wohlfahrtsökonomik wegen ihrer Orientierung an subjektiven ergeb-

<sup>15</sup> Vgl. Vickrey (1945), ders. (1960); Harsanyi (1953), ders. (1955); Samuelson (1964). Daß ihre Modelle in Unkenntnis der Arbeiten von Vickrey - dem wohl eigentlichen Erfinder des Schleiers der Unkenntnis - entstanden, betonen Harsanyi (1982 [1977]), S. 47, und Samuelson (1981), S. 256.

<sup>16</sup> Für eine Darstellung dieses Prinzips vgl. unten § 4 A.

<sup>17</sup> Vgl. dazu z.B. Hoerster (1979).



nisbezogenen Gerechtigkeitsnormen ab; auf der anderen Seite aber führt er mit dem Schleier der Unsicherheit eine Prämisse ein, die ganz ähnliche Implikationen hat wie die kritisierten Theoriezweige. Das eine Mal verwirft er das Rawlsche Differenzprinzip; ein anderes Mal zeigt er, wie man es anders als Rawls herleiten kann.<sup>18</sup> Es erscheint nicht übertrieben festzustellen, daß Buchanan (mindestens) zwei verschiedene Vertragstheorien verfaßt hat: einen „*hobbesianischen*“, auf einer „realistischen“ Ausgangssituation beruhenden Ansatz einerseits, den er vor allem in „Die Grenzen der Freiheit“ präsentiert; und eine „*kantianische*“, auf die Legitimation moralisch richtigen Handelns zielende Vertragstheorie andererseits, die ansatzweise bereits in „The Calculus of Consent“ durchscheint, besonders deutlich aber in dem neuesten Buch „Politics by Principle, Not Interest“ zum Vorschein kommt.<sup>19</sup>

In § 1 werde ich zunächst ganz allgemein versuchen, den methodologischen Status des vertragstheoretischen Arguments zu klären und das konstitutionelle Wohlfahrtskriterium, das es impliziert, als kollektive Auswahlregel (*collective choice rule*) im Sinne der ökonomischen *Social-choice*-Theorie zu formulieren. Dabei werde ich das kontraktualistische Argument als einen ökonomischen Rationalwahlansatz rekonstruieren, der zumindest im Grundsatz alle logischen Adäquatheitsbedingungen einer wissenschaftlichen Erklärung bzw. Prognose erfüllt, nicht aber die empirische Bedingung, nach der die Aussagen im Explanans faktisch wahr zu sein haben. Ich werde argumentieren, daß eine hypothetische Rechtfertigung staatlichen Handelns die Falschheit zumindest einiger Prämissen im Explanans konstitutiv voraussetzt.

In den übrigen Kapiteln der Arbeit werde ich darüber hinaus zu zeigen versuchen, daß sowohl die „Theorie der Gerechtigkeit“ von Rawls als auch die verschiedenen Varianten von Buchanans konstitutionenökonomischer Theorie - neben der von allen Vertragstheorien verletzten empirischen Adäquatheitsbedingung eines wahren Explanans - in der einen oder anderen Weise ebenfalls die logischen Adäquatheitsbedingungen deduktiv-nomologischer Argumente verletzt.

Im Mittelpunkt von § 2 steht die sog. starke Vertragstheorie, wie sie James Buchanan in „Die Grenzen der Freiheit“ vorstellt. Stark werden Vertragstheorien genannt, welche versuchen, allein mit der kontraktualistischen Konsensnorm auszukommen. Legitim ist danach, was die betrachteten Individuen in ihrer hypothetischen Ausgangssituation einstimmig entscheiden; weitergehende Anforderungen an die Gerechtigkeit der Entscheidungsbedingungen werden nicht gestellt, da allein die Werturteile der betrachteten Individuen in die Analyse

---

<sup>18</sup> Vgl. Buchanan/Faith (1980).

<sup>19</sup> Die Unterscheidung zwischen einem „*hobbesianischen*“ und einem „*kantianischen*“ Buchanan verdanke ich Hartmut Kliemt.

einfließen sollen, auch nicht die „externen Normen“ des Vertragstheoretikers selbst.

Buchanan versucht zu zeigen, daß Individuen in einer völlig offenen, d.h. un-spezifizierten und Ungleichheiten zulassenden Entscheidungssituation einen Rechtsschutzstaat wählen werden. Ich werde aber darlegen, daß eine starke Vertragstheorie nur aufzeigen kann, daß die betrachteten Entscheidungsträger irgendeine paretoverbessernde Vertragsalternative wählen werden, aber die Buchanansche Vorhersage, die Individuen wählten einen Rechtsstaat, nur dann eine zulässige Folgerung ist, wenn zusätzliche inhaltliche Gerechtigkeitsnormen eingeführt werden, die Buchanan erklärtermaßen aus der Vertragstheorie heraushalten will. Dem starken Kontraktualismus geht es insofern wie der strikt paretianischen Wohlfahrtsökonomik: Da ihr Legitimationsergebnis völlig unbestimmt bleibt, muß sie jedes Anliegen verfehlen, praktische Gestaltungsempfehlungen zu geben, wie eine gerechte Politik auszusehen habe.

Die sog. schwache Vertragstheorie, welche im Mittelpunkt von § 3 steht, versucht dieses Problem dadurch zu lösen, daß sie mit dem Schleier der Unkenntnis eine Informationsprämisse einführt, welche die betrachteten konstitutionellen Entscheidungsträger zwingt, so zu handeln, *als ob* sie die ethische Universalisierungsnorm teilen, und zwar unabhängig davon, ob sie dies tatsächlich tun. Die Rawlssche Annahme eines Schleiers des Nichtwissens beruht auf einer Norm. Es kommt hier nicht darauf an, daß die Entscheidungsbedingungen in irgendeiner Weise „realistisch“ sind; der Urzustand ist vielmehr allein so gestaltet, daß er als faire Ausgangssituation akzeptiert werden kann. Die Folgen dieser Annahme sind weitreichend. In Abschnitt A. von § 3 werde ich aufzeigen, daß die grundlegende vertragstheoretische Konsensnorm - das Pareto-Prinzip - damit faktisch aufgegeben wird: die Entscheidung im Urzustand fällt nicht nach Maßgabe der Vorstellungen der betrachteten Individuen, sondern gemäß einer ihnen oktroyierten Fairnessnorm.

Buchanan hingegen behauptet die Existenz des Schleiers der Unsicherheit als ein empirisches Phänomen, das bei Verfassungsentscheidungen aller Art auftritt und die Individuen dazu veranlaßt, verallgemeinerbare Entscheidungsergebnisse zu erzielen. In Abschnitt B. von § 3 werde ich jedoch darlegen, daß die Begründung, die Buchanan für den behaupteten empirischen Charakter seiner Annahme anführt, nicht nur widersprüchlich ist, sondern darüber hinaus auch auf einem zirkulären Argument beruht. Hieraus wird der Schluß gezogen werden, daß die Behauptung, der Schleier der Unsicherheit sei ein empirisches Phänomen, eine *ad hoc* eingeführte Interpretation ist, um einerseits die ethische Universalisierungsnorm aus der Vertragstheorie herauszuhalten, andererseits aber Ergebnisse herzuleiten, die mit diesem Werturteil kompatibel sind. Offensichtlich tritt mit der Annahme konstitutioneller Unsicherheit der „kantianische“ Buchanan auf den Plan. Wie die Rawlssche werde ich daher auch die mit dem

Schleier der Unsicherheit operierenden Teile der Buchanan'schen Theorie den schwachen, auf ergebnisbezogenen Werturteilen beruhenden Kontraktansätzen zuordnen.

Die Ausführungen in § 4 sind dem eigentlichen „Begründungs“-Schritt der beiden Vertragstheorien gewidmet. Wer den durch den Schleier der Unkenntnis konkretisierten Urzustand als legitimitätsstiftend akzeptiert, so die Argumentation, der muß auch eine konkrete Gerechtigkeitsauffassung akzeptieren. Nach der Rawlsschen Theorie einigen sich die Beteiligten im Urzustand auf die berühmten beiden Gerechtigkeitsgrundsätze<sup>20</sup>; der Buchanan'sche Ansatz impliziert bei Anwendung der üblichen ökonomischen Entscheidungslogik eine der utilitaristischen ähnliche soziale Wohlfahrtsfunktion.

Ich werde jedoch - erstens - sowohl für die Rawlssche als auch für die Buchanan'sche Theorie die Ansicht zurückweisen, daß der behauptete Deduktionsschritt tatsächlich eine zwingende Folgerung aus den Prämissen der jeweiligen vertragstheoretischen Entscheidungssituation ist. Vielmehr werde ich aufzuzeigen versuchen, daß - bei entsprechender Annahmenspezifikation über die Wahrscheinlichkeitserwartungen - aus den vertragstheoretischen Urzustandsprämissen buchstäblich *jede* denkbare soziale Wohlfahrtsfunktion gefolgt werden kann. Zweitens werde ich darlegen, daß selbst dann, wenn alle Individuen im Urzustand durch ihr Rationalverhalten die gleiche soziale Wohlfahrtsfunktion maximieren, ein Konsens unter dem Schleier der Unkenntnis in Buchanan's Ansatz überhaupt nicht und in der Rawlsschen Theorie nur formal - durch das Wegdefinieren des Problems intersubjektiver individueller Wohlfahrtsvergleiche - garantiert werden kann.

§ 5 untermauert diesen Befund mit Hilfe einer einfachen spieltheoretischen Modellierung der konstitutionellen Konsensfindung unter dem Schleier der Unkenntnis. In der Literatur wurden bislang erstaunlich wenig Versuche unternommen, die Regelwahl unter den Bedingungen konstitutioneller Unkenntnis spieltheoretisch zu modellieren. Soweit man überhaupt zu formalen Argumenten griff, modellierte man die Entscheidungsfindung unter dem Schleier der Unkenntnis in der Tradition der Entscheidungstheorie als Spiel gegen die Natur<sup>21</sup> oder verzichtete - unter Hinweis auf ihre fehlende Realitätsnähe - ganz auf

---

<sup>20</sup> Die beiden Rawlsschen Gerechtigkeitsgrundsätze bestehen zum einen in einem Prinzip gleicher Grundrechte und -freiheiten und zum anderen in der Forderung, sozio-ökonomische Ungleichheiten nur insoweit zuzulassen, wie allen Individuen Ämter und Positionen in gleicher Weise offenstehen (sog. Chancenprinzip) und wie diese Ungleichheiten den am wenigsten begünstigten Gesellschaftsmitgliedern zum Vorteil reichen (sog. Differenzprinzip). Vgl. im einzelnen § 4 A. I.

<sup>21</sup> Vgl. Ellsworth (1978); Pies (1994).

die Verwendung der Prämisse der Positionsunsicherheit.<sup>22</sup> Erst seit einigen Jahren finden sich einige Ansätze, die konstitutionelle Wahl von Regeln unter dem Schleier der Unkenntnis mit dem Instrumentarium der Spieltheorie zu analysieren.<sup>23</sup> Eine solche formalisierte Argumentation vermag einerseits die Bedingungen zu präzisieren helfen, unter denen die von den Vertragstheoretikern verbal behaupteten Theoreme gelten; andererseits lassen sich durch sie auch eventuelle Irrtümer, die sich in einer rein qualitativen Argumentation verbergen können, leichter aufdecken.

Die Modellanalyse in § 5 wird die in § 4 aufgestellte Behauptung stützen, daß man, wenn man das Modell des Urzustands als Legitimationsquelle akzeptiert, nicht auch gezwungen ist, die jeweilige konkrete Gerechtigkeitskonzeption zu teilen. Es wird gezeigt werden, daß die Rawlssche und die Buchanansche soziale Wohlfahrtsfunktion schon deshalb nicht aus der Universalisierungsnorm und den vertragstheoretischen Prämissen des Urzustands als deren metaphorischer Ausgestaltung folgen können, weil sie Implikationen haben, die mit der ethischen Verallgemeinerungsnorm nicht in Einklang zu bringen sind. Diese Konflikte können bereits dann auftreten, wenn die in der Vertragstheorie übliche Annahme getroffen wird, daß Gesellschaftsverträge von einem Gefangenendilemma (GD) betroffen sind. Für die konstitutionelle Regelwahl in einem einfachen Zwei- und  $n$ -Personen-Gefangenendilemma wird dargelegt, daß ein Schleier der Unkenntnis à la Rawls oder Buchanan einerseits nicht garantieren kann, daß das Einigungsergebnis der Beteiligten fair im Sinne der ethischen Verallgemeinerungsnorm ist; andererseits vermag ein Buchananscher Schleier der Unsicherheit auch nicht sicherzustellen, daß überhaupt eine Einigung auf der konstitutionellen Entscheidungsstufe erzielt wird.

In § 6 wird darüber hinaus dargelegt, daß alle diese Fairneß- und Konsensprobleme sich gegebenenfalls dadurch lösen lassen, daß die Annahme eines Schleiers der Unkenntnis einfach aufgegeben wird: Beruht die Vertragstheorie allein auf dem Pareto-Prinzip als einziger normativer Grundlage, so ist im Spiel, das die Beteiligten auf der konstitutionellen Ebene spielen, die faire Lösung des Zwei-Personen-Gefangenendilemmas immer und die des  $n$ -Personen-Gefangenendilemmas unter einer bestimmten Bedingung das einzige Nash-Gleichgewicht in sukzessiver Dominanz. Entgegen einer weitverbreiteten Ansicht in der ökonomischen Vertragstheorie hat der Schleier der Unkenntnis in

---

<sup>22</sup> Vgl. *Faber* (1973).

<sup>23</sup> Zu nennen sind v.a. die Arbeiten von *Howe/Roemer* (1981); *Gaa* (1984); *Binmore* (1989), *ders.* (1994); *Schweizer* (1988), *ders.* (1990); *Eichberger/Pethig* (1993), *ders.* (1994); *Wessels* (1993), *ders.* (1994); *Mueller* (1998a); *Müller* (1998a).

diesen Fällen also keine positive - d.h. fairneß- und konsensfördernde - Funktion; er kann das Erreichen von Fairneß und Einstimmigkeit auf der konstitutionellen Ebene allerdings verhindern.

§ 7 beschließt die Arbeit mit einer Zusammenfassung und einer kurzen Gesamtwürdigung der Ergebnisse.

## § 1 Das vertragstheoretische Argument

§ 1 der vorliegenden Arbeit zeichnet den Weg der Theorien des hypothetischen Gesellschaftsvertrags nach, beginnend mit dem Wunsch nach individualistischer (A.) und konsensualer (B.) Legitimation allen Staatshandelns durch die von ihm Betroffenen bis hin zum Konstrukt der hypothetischen Einstimmigkeit als „Probierstein der Rechtmäßigkeit eines jeden öffentlichen Gesetzes“ (Kant). Kapitel C. rekonstruiert zunächst (in Abschnitt I.) das konstitutionelle Wohlfahrtskriterium der Theorien des hypothetischen Vertrags in allgemeiner Form als kollektive Auswahlregel im Sinne der ökonomischen *Social-choice*-Theorie. Abschnitt II. versucht daraufhin, den methodologischen Status des kontrakttheoretischen Gedankenexperiments zu klären. Das vertragstheoretische Argument, so werde ich argumentieren, weist die gleiche logische Struktur wie eine deduktiv-nomologische Erklärung oder Prognose nach dem Hempel-Oppenheim-Schema auf mit dem wesentlichen Unterschied, daß die faktische Falschheit der (meisten) Annahmen im Explanans des vertragstheoretischen Arguments die konstitutive Voraussetzung für die Erfüllung seiner theoretischen Aufgabe - der hypothetischen Legitimation politischer Handlungen - ist. Abschließend werde ich - in Abschnitt III. - die These vertreten, daß die hypothetisch-kontraktualistische Theorie auf ein Induktionsproblem stößt, aufgrund dessen sie keine reale Verpflichtung realer Individuen begründen kann.

### A. Individualismus

Das höchstrangige Werturteil der Theorien des Gesellschaftsvertrags ist der *normative Individualismus (NI)*, der die „meta-normative“ Leitlinie für die Ableitung „richtiger“ Werturteile über soziale Situationen darstellt.<sup>1</sup> In der Formulierung Vanbergs verlangt diese Wertprämisse:

*Normativer Individualismus (NI)*. „(S)ocial states are to be judged ‘good’ to the extent that the individuals concerned judge them to be good. And social states are to be judged ‘better’ than others if they are judged that way by the individuals concerned, i.e. if they are preferred by these individuals over potential alternatives.“<sup>2</sup>

---

<sup>1</sup> Vgl. Vanberg (1986), S. 114; Kersting (1994), S. 14.

<sup>2</sup> Vanberg (1986), S. 115.

Bewertungen sollen nach diesem Postulat, das in der ökonomischen Theorie auch in Gestalt des Prinzips der Konsumenten-souveränität auftritt, allein jenen Individuen überlassen sein, die von den Folgen der bewerteten Zustände „betroffen“ sind. Die einzige Instanz, welche die Legitimität einer Regel oder einer ganzen Gesellschaftsordnung zu bewerten in der Lage angesehen wird, sind die Menschen, die unter diesen Institutionen leben müssen; sie allein sollen „die Quelle aller Werte“<sup>3</sup> bzw. „selbstschaffende Quellen berechtigter Ansprüche“<sup>4</sup> sein. „Meta-normativ“ ist die Forderung (NI) insofern, als es sich hierbei um ein Werturteil „zweiter Ordnung“ handelt, das eine Leitlinie zur Einstufung jeglicher Werturteile „erster Ordnung“, d.h. jeder Art individueller Nutzenbewertungen, darstellt.<sup>5</sup>

Die individualistische Grundnorm verlangt, die Menschen und ihre Entscheidungen so zu akzeptieren, wie sie sind, unabhängig davon, ob ihre Präferenzen rein egoistisch, altruistisch oder sogar malevolent sind.<sup>6</sup> Die extremste Form des Individualismus stellt nicht einmal normative Anforderungen an die Bedingungen, unter denen die legitimitätsstiftenden Entscheidungen zustande gekommen sind. Buchanan etwa, der hier wohl die am weitesten reichende Position vertritt, will sogar unter Zwang getroffene Entscheidungen als Rechtfertigungsquelle akzeptieren.<sup>7</sup>

## B. Einstimmigkeit

### I. Politik als Tausch

Die Theorien des Gesellschaftsvertrags betrachten die Einstimmigkeit als die natürliche „Verlängerung“ des meta-normativen Postulats (NI) in die Sphäre der Kollektiventscheidung.<sup>8</sup> Die konstitutionenökonomische Variante der Vertrags-

<sup>3</sup> Brennan/Buchanan (1993 [1985]), S. 28.

<sup>4</sup> Rawls (1992 [1980]), S. 119.

<sup>5</sup> Vgl. zur Unterscheidung von Werturteilen erster und zweiter Ordnung Mackie (1981 [1977]), S. 12 f., mit Bezug auf den normativen Individualismus S. 119.

<sup>6</sup> Vgl. z. B. Mueller (1989b), S. 3.

<sup>7</sup> Vgl. etwa die Ausführungen zum sog. „Sklavereivertrag“ bei Buchanan (1984 [1975]), S. 85 f. Vgl. auch unten § 2.

<sup>8</sup> Vgl. Buchanan/Congleton (1998), S. 16. - Gleichwohl impliziert das Postulat des normativen Individualismus nicht das Einstimmigkeitsprinzip, wie man gelegentlich liest. Auch andere Normen - etwa das „Recht des Stärkeren“, andere Individuen mit Gewalt zu unterdrücken - können durch das Postulat des normativen Individualismus gedeckt sein. Vorliegend genügt jedoch die Feststellung, daß auch einstimmige Entscheidungen jedenfalls Entscheidungen im Sinne des individualistischen Postulats sind.

theorie begründet dies mit einer Übertragung des auf Konkurrenzmärkten geltenden Tauschprinzips auf die Interaktion von Individuen im politischen Prozeß.<sup>9</sup> Auf Märkten vollzieht sich danach - ohne Rückgriff auf irgendwelche „externen“, außerhalb der Vertragsparteien stehende Kriterien<sup>10</sup> - die Ermittlung des Werts der ausgetauschten Sache oder Leistung im Rahmen von Austauschverträgen zwischen den Transaktionspartnern. Durch den Tauschakt offenbaren die Vertragsparteien, daß beide bei gegebener Ausstattung und gegebenen Spielregeln die Ergebnisse der vorgenommenen Transaktion höher schätzen als die Güterallokation ohne Tausch; der Tausch im Konsens ist insoweit aus der Sicht der Beteiligten wertmaximierend.<sup>11</sup>

Wie den Wert eines Tauschguts in Markttransaktionen läßt sich nach vertragstheoretischer Auffassung auch der Wert einzelner gesellschaftlicher Zustände oder einer ganzen Gesellschaftsordnung aus einem multilateralen, konsensualen „Verhaltens Austausch“<sup>12</sup> zwischen den sie tragenden Bürgern ableiten. Bereits Knut Wicksell<sup>13</sup> hatte vor mehr als einem Jahrhundert die Einstimmigkeit aller von einer politischen Wahlhandlung Betroffenen als „institutionelles Analogon zum Zwei-Personen-Tausch“<sup>14</sup> herausgestellt. In Analogie zum Marktparadigma interpretiert die Konstitutionenökonomik politische Prozesse daher idealiter als multilaterale Verträge zwischen Kollektivmitgliedern, im Laufe derer ständig Kompromisse vereinbart und unterschiedliche Individualinteressen miteinander versöhnt werden. Politik ist nach dieser Auffassung „die Struktur eines komplexen Austauschvorganges zwischen Individuen - eine Struktur, in deren Rahmen Personen versuchen, ihren eigenen, persönlich definierten Zielen gemeinschaftlich nachzugehen, weil sie diese Ziele durch einfachen marktwirtschaftlichen Austausch nicht effizient erreichen können.“<sup>15</sup>

Das Ergebnis des Verhaltenstausches ist die Vereinbarung bestimmter Regeln. Diese Regeln induzieren bestimmte gesellschaftliche Zustände (Situationen). Zwischen Regeln zu wählen, bedeutet daher - gegeben die vollkommene Selbstbindungsfähigkeit der Individuen<sup>16</sup> - stets zugleich, sich zwischen gesell-

<sup>9</sup> Vgl. *Buchanan* (1989 [1986]), S. 938 f.; *ders.* (1984 [1975]), S. 54 ff.; *Brennan/Buchanan* (1993 [1985]), S. 30 ff.; *Buchanan/Congleton* (1998), S. 4.

<sup>10</sup> So ausdrücklich *Brennan/Buchanan* (1993 [1985]), S. 31.

<sup>11</sup> Vgl. *Brennan/Buchanan* (1993 [1985]), S. 32.

<sup>12</sup> *Buchanan* (1984 [1975]), S. 85.

<sup>13</sup> Vgl. *Wicksell* (1969 [1896]), S. 113 ff. Zu Wicksells Bedeutung für die Public-Choice-Theorie vgl. *Wagner* (1988).

<sup>14</sup> *Buchanan* (1984 [1975]), S. 55; vgl. auch bereits *Buchanan* (1960 [1954]).

<sup>15</sup> *Buchanan* (1989 [1986]), S. 938. Für eine fast wortgleiche Formulierung vgl. *Buchanan/Congleton* (1998), S. 16.

<sup>16</sup> Die vollkommene Fähigkeit der Vertragsparteien zur Selbstbindung ist eine in der Kontrakttheorie übliche (meist implizite) Annahme (vgl. z. B. *Binmore* 1989, S. 93; *Schweizer* 1990, S. 33). Die Annahme vollkommener Selbstbindungsfähigkeit bedeutet,



schaftlichen Zuständen (Situationen) zu entscheiden. Im folgenden werden daher die Begriffe der Regel und des gesellschaftlichen Zustandes synonym gebraucht. Eine Regel (bzw. ein Gesellschaftszustand)  $r \in R$  (mit  $R$  = Menge aller Regeln/Zustände) wird durch den Vektor der Auszahlungen  $u_i$  aller betroffenen Individuen  $i$  (mit  $i = 1, \dots, n$ ) beschrieben:  $r = (u_1, u_2, \dots, u_n)$ .

Sofern aus Gründen prohibitiv hoher Konsensfindungskosten<sup>17</sup> und rationaler Uninformiertheit der Bürger ein direkt-demokratischer, einstimmiger Vertragsschluß über jede einzelne politische Maßnahme realiter unerreichbar sein dürfte, tritt an die Stelle der fallweisen, konsensualen Entscheidungsfindung aller Individuen eine indirekte, nicht-einstimmige Demokratie, deren Entscheidungsinhalte und Grenzen durch einen einzigen einstimmigen multilateralen Gesellschaftsvertrag - die *Verfassung* - gezogen werden.

Trotz ihrer offensichtlichen Strukturähnlichkeit unterscheiden sich individueller Markttausch und einstimmiger politischer Verhaltenstausch in einer modernen Massengesellschaft in einem wesentlichen Punkt: im multilateralen Politiktausch spielt der „Repräsentationsaspekt“ des Tausches, der im Zweipersonen-Tausch sehr ausgeprägt ist, praktisch keine Rolle; eine Analogie zwischen Markt und Politik besteht nur in bezug auf den „Schutzaspekt“ der Einstimmigkeit.

a) Unter dem *Repräsentationsaspekt* der Einstimmigkeit kann man die Möglichkeit verstehen, mittels einer einstimmigen Entscheidung ein beliebiges Tauschvorhaben nach eigenen Vorstellungen durchzusetzen. Die schwache Ausprägung des Repräsentationsaspekts einstimmiger Entscheidungen in multilateralen Tauschhandlungen läßt sich anhand einer einfachen Wahrscheinlichkeitsüberlegung demonstrieren. Nehmen wir an, in einer Menge von  $n$  Individuen wolle ein Individuum  $i$  den übrigen Beteiligten ein Tauschangebot  $x$  - über den Austausch eines Konsumgutes oder ein Steuerreformgesetz - machen. Jedes

---

daß eine einmal eingesetzte Institution durch irgendeine außerhalb der Vertragsparteien stehende Instanz kostenlos und vollkommen durchgesetzt wird; Neuverhandlungen über eine einstimmig angenommene Regel finden nicht statt. Auch vom Problem, daß die Durchsetzung einer gesellschaftsvertraglich festgelegten Regel einen Vertrag mit der Durchsetzungsinstanz voraussetzt, der seinerseits durchgesetzt werden muß, abstrahieren die Vertragstheorien regelmäßig (*Smyth* 1972). Materiell kann man sich vollkommene Selbstbindung so vorstellen, daß eine dritte Instanz - etwa der Staat - durch die Verhängung von Sanktionen, die Einholung eines Pfands oder durch „Umerziehung“ die Entscheidungssituation so variiert, daß die auf der konstitutionellen Ebene gewählte Regel *selbstdurchsetzend* wird, da es im Eigeninteresse der Spieler liegt, das von der Regel verlangte Verhalten zu wählen. Zur Diskussion der Selbstbindungsprämisse allgemein vgl. z. B. *Kliemt* (1993); *Lahno* (1995); *Beckmann* (1998), 401-511.

<sup>17</sup> Vgl. zu diesem Begriff *Buchanan/Tullock* (1962); im Zusammenhang mit dem Gesellschaftsvertrag z. B. *Buchanan/Congleton* (1998), S. 17.

Individuum stimme mit gleicher Wahrscheinlichkeit mit „Ja“ oder mit „Nein“ (vom Fall der Indifferenz zwischen Annahme und Ablehnung sei vereinfachend abstrahiert). Dann beträgt die Wahrscheinlichkeit, daß  $i$  für seine Idee ein einstimmiges Gesamtvotum erzielt,  $1/(2^{n-1})$ . In einem einstimmigen Zwei-Personen-Tausch ( $n=2$ ) setzt  $i$  seinen Vorschlag also mit einer Wahrscheinlichkeit von  $\frac{1}{2}$  durch; für große Kollektive ( $n \rightarrow \infty$ ) hingegen geht  $i$ 's Wahrscheinlichkeit, genau seinen eigenen Vorschlag unverändert durchzubringen, jedoch gegen Null. Diese Asymmetrie verschärft sich abermals, wenn man realistischerweise davon ausgeht, daß Individuum  $i$  sich auf dem Markt die Partner eines Zwei-Personen-Austausches selbst aussuchen kann, während die Tauschpartner in einem gegebenen politischen Entscheidungskollektiv bei Einstimmigkeit nicht austauschbar sind. Entscheidungsträger  $i$  wird beim Zwei-Personen-Tausch sein Tauschangebot solchen potentiellen Transaktionspartnern machen, die er kennt und mit deren Einwilligung er mit einer höheren Wahrscheinlichkeit rechnen kann als  $\frac{1}{2}$ . Während also die Repräsentation der eigenen Tauschvorstellung im Tauschergebnis in bilateralen Transaktionen einen hohen Stellenwert hat, spielt diese im Kontext einer Massengesellschaft praktisch keine Rolle.

b) Doch trotz der geringen Repäsentation des positiv formulierten Willens des Einzelnen im Ergebnis einstimmiger Kollektiventscheidungen, ist das Konsensprinzip hier keineswegs wertlos. Denn Entscheidungsregeln haben nicht allein die Funktion, die Durchsetzung individueller Vorstellungen zu erleichtern. Wenn dies so wäre, dann gäbe es auch in marktlichen Zwei-Parteien-Transaktionen eine Entscheidungsregel, die den individuellen Willen besser als die Einstimmigkeit zu fördern in der Lage wäre: die „Jedermann-Regel“, nach der jeder der Beteiligten das Recht hätte, sich das Tauschobjekt auch ohne die Zustimmung des „Transaktionspartners“ anzueignen. Unter den obigen Annahmen (und unter Abstraktion von Verteidigungsanstrengungen des „Anbieters“) betrüge dann die Wahrscheinlichkeit, den eigenen Willen durchzusetzen, eins. Offensichtlich hat die Einstimmigkeitsregel hier jedoch einen *Schutzaspekt*, der es dem Nachfrager nach einem Gut verbietet, dessen Anbieter dadurch „externe (Entscheidungs-)Kosten“<sup>18</sup> aufzuzwingen, daß er sich das Gut einfach aneignet.

Dieser Schutzaspekt der Einstimmigkeit ist es auch, der die Einstimmigkeit aus vertragstheoretischer Perspektive so attraktiv macht. Bei Geltung der Einstimmigkeitsregel kann niemand von einem Teil des Kollektivs gezwungen werden, eine Kollektiventscheidung, die er nicht wünscht, zu dulden und - worauf es wohl besonders ankommt - mitzufinanzieren. Durch das *Vetorecht*, das die Einstimmigkeitsregel jedem Individuum einräumt<sup>19</sup>, kann jedes Gesell-

<sup>18</sup> Vgl. zu diesem Begriff *Buchanan/Tullock* (1962).

<sup>19</sup> Vgl. bereits *Wicksell* (1969 [1896]), S. 143.

schaftsmitglied solche „externen Kosten“ der Kollektiventscheidung abwehren. Jede einstimmige Entscheidung bedeutet demnach, daß gegenüber der Ausgangssituation eine Wohlfahrtsverbesserung nach dem *Pareto-Kriterium* erreicht wird.

In ihrer starken Formulierung besagt die Pareto-Relation  $\bar{P}$  (mit  $R_i$  = „aus Sicht von Individuum  $i$  besser als oder gleichwertig zu“,  $P_i$  = „aus Sicht von Individuum  $i$  besser als“):

*Starke Pareto-Relation* ( $\bar{P}$ ). Für zwei beliebige gesellschaftliche Regeln (Situationen)  $r$  und  $q$  gilt:  $r$  ist gegenüber  $q$  genau dann eine starke Pareto-Verbesserung, wenn mindestens ein Kollektivmitglied  $r$  gegenüber  $q$  strikt vorzieht und alle übrigen Kollektivmitglieder nicht  $q$  gegenüber  $r$  schwach präferieren:

$$r \bar{P} q \leftrightarrow [\forall i: r R_i q \wedge \exists i: r P_i q].^{20}$$

Verlangt man  $r P q \leftrightarrow r \bar{P} q$  (mit  $P$  = „aus Sicht des gesamten Kollektivs besser als“), so determiniert die starke Pareto-Relation  $\bar{P}$  eine kollektive Auswahlregel (*collective choice rule*; CCR). Darunter versteht man in der ökonomischen *Social-Choice-Theorie* eine funktionale Beziehung  $f$ , die jedem  $n$ -Tupel individueller Präferenzordnungen  $R_1, \dots, R_n$  (eine Präferenzordnung je Individuum) genau eine gesellschaftliche Präferenzrelation  $R$  zuordnet,  $R = f(R_1, \dots, R_n)$ .<sup>21</sup>

Jede einstimmig getroffene Entscheidung von zwei oder mehr rationalen Individuen ist *paretoeffizient* in dem Sinne, daß eine Situation erreicht wird, von der ausgehend niemand mehr bessergestellt werden kann, ohne ein anderes Kollektivmitglied schlechterzustellen.<sup>22</sup> Andere, nicht-einstimmige Entscheidungsregeln für Kollektive - wie etwa die einfache Mehrheitsregel - haben diesen Vorzug nicht, so daß einzelne Individuen zu Zwangskonsumenten eines bestimmten Kollektivguts oder gar zu Zwangsfinanziers einer klientelorientierten Subventionierung bestimmter Gruppen der Gesellschaft werden können. Unter nicht-einstimmigen Kollektiventscheidungsregeln ist es auch

<sup>20</sup> Vgl. Sen (1970), S. 28.

<sup>21</sup> Zum Begriff der kollektiven Auswahlregel (CCR) vgl. grundlegend Sen (1970), S. 28 ff. Vgl. auch z. B. Schmidt (1991), Kap. II.

<sup>22</sup> Wie in der Konstitutionenökonomik, aber auch in der *Social-choice-Theorie* (vgl. Sen 1970, Kapitel 2) üblich, werden in der vorliegenden Arbeit die Begriffe der Einstimmigkeit und der Pareto-Verbesserung synonym verwendet. Rawls (1994 [1971]), S. 317, hingegen sieht zwischen dem Einstimmigkeitskriterium und dem Pareto-Prinzip den Unterschied, daß der erste im Unterschied zum zweiten Grundsatz eine gerechte Ausgangsverteilung von Einkommen und Vermögen voraussetze. Die Forderung, die Verwendung des Einstimmigkeitsprinzips als Gerechtigkeitsgrundsatz sei nur akzeptabel, wenn die Ausgangsverteilung selbst gerecht sei, findet sich bereits bei Wicksell (1969 [1896]), S. 143, und Lindahl (1919), S. 100 f.

möglich, daß der einzelne zum Opfer des „Wohllollens“ der jeweils herrschenden Gruppe wird, die sich - möglicherweise mit den besten Absichten - daran macht, die ihrer Ansicht nach „verzerrten“ Präferenzen der überstimmten Individuen zu „korrigieren“.<sup>23</sup> Nur das durch die Einstimmigkeitsregel verbürgte Vetorecht gewährt dem Individuum einen Schutz vor einer möglichen „Tyrannei der Mehrheit“<sup>24</sup>.

## II. Einstimmigkeit und Verfahrensgerechtigkeit

Hat das Individuum sein Vetorecht nicht ausgeübt, so muß es eine Kollektiventscheidung auch dann gegen sich gelten lassen, wenn sie ihm später mißfällt, getreu dem alten Rechtssatz „*Volenti non fit iniuria*“ - der Zustimmungende kann nicht aus dem von ihm Akzeptierten ungerecht behandelt werden.<sup>25</sup> „Starke“ Vertragstheorien, die mit dem Pareto-Prinzip als alleiniger Wertbasis auszukommen versuchen<sup>26</sup>, verwenden den Einstimmigkeitsbegriff daher auch synonym mit den Begriffen der *Gerechtigkeit*, *Fairneß* oder *Legitimität*<sup>27</sup>; „schwa-

---

<sup>23</sup> Zu einer modelltheoretischen Analyse und konstitutionenökonomischen Einschätzung der Wohlfahrtsverluste, die bei einer solchen „Meritorisierung“ des Angebots privater Güter entstehen können, vgl. Tietzel/Müller (1998).

<sup>24</sup> *de Tocqueville* (1976 [1848]), S. 289-292. Für eine moderne ökonomische Analyse dieser Tyrannei der Mehrheit vgl. *Buchanan/Congleton* (1998). - Ein zweiter wichtiger Unterschied zwischen Markttasch und Politik sei, da er im Kontext der vorliegenden Arbeit keine Rolle spielt, nur am Rande erwähnt: Während die Transakteure im Zwei-Personen-Tasch stets für das Zustandekommen des Vertrags *entscheidend* sind, hat ein beliebiger Kollektiventscheidungssträger *i* bei Geltung der Einstimmigkeitsregel nur dann mit seinem Veto kausalen Einfluß auf das kollektive Entscheidungsergebnis, wenn *alle* übrigen Kollektivmitglieder der vorgeschlagenen Maßnahme zustimmen. Macht aber außer *i* noch mindestens ein anderes Individuum von seinem Ablehnungsrecht Gebrauch, so wird der Vorschlag verworfen, *völlig unabhängig* davon, wie *i* sich entscheidet. Ohne jeden Einfluß auf das Ergebnis hat *i* nun aber möglicherweise einen Anreiz, „expressiv“ - nach Gesichtspunkten, die möglicherweise gegen seine eigenen Präferenzen verstoßen - zu wählen. Auf diese Weise kann es zu einstimmigen Beschlüssen kommen, die kein einzelnes Individuum befürwortet hätte, wäre es jeweils entscheidend gewesen. Zwischen Einstimmigkeit und Pareto-Effizienz besteht mithin keine *logische* Beziehung. Vgl. hierzu *Brennan/Lomasky* (1984) und (1993), v.a. Kap. 8. Von diesem Unterschied wird in dem in §§ 5 und 6 entwickelten Modell durch die Annahme, daß das Wählen gemäß den eigenen Präferenzen eine dominante Strategie ist, abstrahiert.

<sup>25</sup> Vgl. *Kant* (1983 [1797]), § 46, S. 432.

<sup>26</sup> Vgl. zum Begriff der starken Vertragstheorie unten § 2.

<sup>27</sup> Vgl. v.a. *Buchanan* (1977a), S. 124 und S. 128 ff.; *Brennan/Buchanan* (1993 [1985]), S. 25 und S. 29.

che“ Kontraktansätze machen demgegenüber diese Gleichsetzung von der Erfüllung zusätzlicher Bedingungen abhängig. Die vertragstheoretische Gerechtigkeitsvorstellung ist damit - wenigstens im Grundsatz - nicht inhaltlich, im Hinblick auf gesellschaftliche Endzustände definiert, sondern ein Begriff der *Tausch- oder Verfahrensgerechtigkeit (iustitia commutativa)*<sup>28</sup>. Gerecht ist eine kollektiv vereinbarte Regel, wenn sie - Informiertheit und Rationalität der Beteiligten vorausgesetzt - das Ergebnis eines bestimmten Tausch- oder Abstimmungsverfahrens, einer einstimmigen Beschlußfassung, ist.

## C. Hypothetische Einstimmigkeit

### I. Der Probierstein der Rechtmäßigkeit

So wünschenswert aus individualistischer Sicht einstimmige Entscheidungen aller politischen Belange auch sind, sie haben einen entscheidenden Nachteil: im Konsens getroffene Entscheidungen sind besonders teuer. Eine erste Kostenrestriktion einstimmiger Entscheidungen stellen die *intrapersonellen Konsensfindungskosten* dar. Jedes Individuum, das an einer Entscheidung teilnimmt, muß sich über die zur Abstimmung stehende Sache informieren, um sich eine Meinung zu bilden, wie es in der bevorstehenden Wahl stimmen will. Informationen sind jedoch nicht kostenlos; die Aufnahme, der Erwerb und die Speicherung von Informationen verursachen monetäre Kosten sowie Opportunitätskosten der Zeit. Jedes Individuum wird rationalerweise nur so lange Informationen suchen, wie der Grenznutzen der zusätzlichen Informationseinheit deren Grenzkosten übersteigt. Es wird daher, je nach Nutzenfunktion, nicht im Interesse eines jeden einzelnen Individuums liegen, sich vollkommen über jeden Sachverhalt zu informieren. Vielmehr werden die meisten Individuen zu einem gewissen Grad „rational ignorant“ sein.<sup>29</sup>

Neben den intrapersonellen Kosten der Entscheidungsfindung verhindert eine zweite Kostenart, daß alle Entscheidungen der laufenden Politik einstimmig getroffen werden können: die *(interpersonellen) Konsensfindungskosten*.<sup>30</sup> Wenn zwei oder mehr Personen eine einstimmige Entscheidung über einen bestimmten Gegenstand treffen sollen, müssen sie Zeit und Mühe, gegebenenfalls auch Geld, aufwenden, um zu einem Ergebnis zu kommen. Diese Ent-

<sup>28</sup> Vgl. zum Begriff der Verfahrensgerechtigkeit z. B. *Ballestrem* (1983), S. 3; *Barry* (1984), S. 584.

<sup>29</sup> Vgl. *Stigler* (1961), S. 216. Zu den Problemen eines solchen „Informationsoptimums“ vgl. jedoch *Tietzel* (1985), S. 14 ff.

<sup>30</sup> Vgl. *Buchanan/Tullock* (1962), S. 68 ff.

scheidungskosten sind um so höher, je mehr Personen an der zu treffenden Entscheidung beteiligt sind. Einstimmige Entscheidungen über jeden Gegenstand der Tagespolitik werden daher zumindest in modernen Massengesellschaften regelmäßig prohibitive Konsensfindungskosten verursachen.

Aufgrund dieser Kostenrestriktionen ist eine direkt-demokratische Entscheidung aller tagespolitischen Entscheidungen regelmäßig nicht effizient. Nach allem, was man heute weiß, ist es in einer großen Gesellschaft bislang nicht einmal gelungen, wenigstens eine einstimmige Entscheidung aller Bürger über die Verfassung ihres Zusammenlebens - einen multilateralen *Gesellschaftsvertrag* - herbeizuführen. Die Historizität einer einstimmigen Legitimation einer Gesellschaftsordnung durch einen Gesellschaftsvertrag wird in der heutigen vertragstheoretischen Diskussion, angesichts praktisch völlig fehlender historischer Beispiele<sup>31</sup>, kaum mehr bejaht.<sup>32</sup> Aber selbst dann, wenn es irgendwann einmal einen solchen Vertrag gegeben haben sollte, von dem wir heute nichts mehr wissen, fragt es sich, warum heute lebende Generationen noch an Verträge gebunden sein sollten, die ihre Vorfahren einmal geschlossen haben.<sup>33</sup> Alle realen Gesellschaften, auch die demokratischen, weisen deshalb - mangels eines einstimmigen Gründungsaktes - aus vertragstheoretischer Sicht ein gewisses und wohl unvermeidliches „Legitimationsdefizit“<sup>34</sup> auf.

Doch handeln, ob legitimiert oder nicht, müssen Kollektive in jedem Fall, stellt doch selbst der Verzicht auf ein Tun letztlich eine „Handlung“ dar. Ziel der Theorien des hypothetischen Gesellschaftsvertrags ist es daher, Gestaltungsempfehlungen für die praktische Politik zu formulieren, die wenigstens eine gewisse *Quasi-Legitimität* - eine erkennbare „Nähe“ zum normativen Ideal der konsensualen Legitimation - aufweisen können. Ausgangspunkt ist dabei die Überlegung, wie - wenn schon nicht die wirklich von einer Kollektiventscheidung Betroffenen - wenigstens rationale Durchschnittsindividuen in einer gedachten Situation der Verfassungswahl das individualistische Postulat

<sup>31</sup> Ein berühmter Ausnahmefall ist allerdings jener der puritanischen „Mayflower“-Pilger, die zum Zweck der Gründung einer eigenen, auf biblischen Grundsätzen beruhenden Kolonie 1620 den „Mayflower Compact“ abschlossen. Der Vertragsschluß der Pilger war jedoch seinerseits von den Theorien des Gesellschaftsvertrags inspiriert. Vgl. *Benn/Peters* (1959), S. 319; *Lutz* (1988), S. 26 ff.

<sup>32</sup> So konstatiert etwa *Buchanan* (1977 [1976]), S. 82: „We know, factually and historically, that the ‘social contract’ is mythological, at least in many of its particulars. Individuals did not come together in some original position and mutually agreed on the rules of social intercourse.“ Auch *Rawls* (1992 [1985]), S. 271, betont den ahistorischen Charakter seiner Gesellschaftsvertragslehre. Noch *Locke* (1992 [1689]), §§ 100 f., S. 262 f., hielt den Gesellschaftsvertrag für eine historische Tatsache.

<sup>33</sup> Dies ist ein altes Argument, das bereits bei *David Hume* (1993), S. 167 f., und *John Locke* (1992 [1689]), § 116, S. 274, anzutreffen ist.

<sup>34</sup> *Eschenburg* (1980), S. 37.

verwirklichen würden. Solange das Ideal des *realen* Konsenses unerfüllbar ist, bleibt als (erheblich schwächeres) Substitut nur die *hypothetische Einstimmigkeit*, um gesellschaftliche Zustände positiv auszuzeichnen, die herbeizuführen die praktische Politik anstreben sollte. Unter den klassischen Sozialvertragstheoretikern war Immanuel Kant der exponierteste Vertreter dieser Lehre vom hypothetischen Gesellschaftsvertrag:

„Allein dieser Vertrag (contractus originarius oder pactum sociale genannt) ... ist keineswegs als *Faktum* vorauszusetzen nötig (ja als solches gar nicht möglich). ... Sondern es ist eine *bloße Idee* der Vernunft, die aber ihre unbezweifelte (praktische) Realität hat: nämlich jeden Gesetzgeber zu verbinden, daß er seine Gesetze so gebe, als sie aus dem vereinigten Willen eines ganzen Volks haben entspringen können, um jeden Untertan, sofern er Bürger sein will, so anzusehen, als ob er zu einem solchen Willen mit zusammen gestimmt habe. Denn das ist der Probestein der Rechtmäßigkeit eines jeden öffentlichen Gesetzes.“<sup>35</sup>

Den gleichen „Probestein der Rechtmäßigkeit“ scheint auch Buchanan zu verwenden, sofern er anmerkt:

„We may evaluate any element of the existing legal structure in terms of its possible consistency with ‘that which *might* emerge’ from a genuine ‘social contract’ among all persons who are involved in the interaction. This test applies equally to those elements of legal structure that may have evolved without conscious design or intent and to those elements which may have been quite explicitly ‘laid on’ for the achievement of a particular purpose at some time in the past. The evaluative analyst must test all ‘law’ on such ‘*as if*’ contractarian criteria.“<sup>36</sup>

Auch Rawls versteht seine „Theorie der Gerechtigkeit“ ganz ausdrücklich in diesem Sinne „kantisch“:

„Eine kantische Gerechtigkeitskonzeption versucht nun diesen Konflikt zwischen den verschiedenen Auffassungen von Freiheit und Gleichheit dadurch zu lösen, daß sie fragt, auf welche traditionell anerkannten Prinzipien der Freiheit und Gleichheit (oder eine ihrer natürlichen Varianten) sich freie und gleiche moralische Personen einigen würden, wenn sie sich in fairer Weise allein als solche Personen darstellten und als Bürger betrachteten, die ihr ganzes Leben in einer bestehenden Gesellschaft verbringen. Durch ihre Übereinkunft (vorausgesetzt, es käme zu einer Übereinkunft), würden, so die Annahme, die angemessensten Grundsätze von Freiheit und Gleichheit ausgewählt und damit zugleich die Grundsätze der Gerechtigkeit festgelegt.“<sup>37</sup>

<sup>35</sup> Kant (1983 [1793]), Band 9, S. 153 (Hervorhebungen im Original).

<sup>36</sup> Buchanan (1977 [1977]), S. 33 (Hervorhebungen von mir; C. M.). Weitere Stellen der Annahme eines hypothetischen Konsenses finden sich z. B. bei Buchanan (1977a), S. 127; Brennan/Buchanan (1988 [1980]), S. 5; dies. (1993 [1985]), S. 35; Buchanan/Congleton (1998), S. 5. - Zu Unterschieden zwischen Kants und Buchanans Lehre des Gesellschaftsvertrags vgl. jedoch Petersen (1996), S. 216 ff.

<sup>37</sup> Rawls (1992 [1980]), S. 83.



Der gedankenexperimentelle Nachweis einer *allgemeinen Zustimmungsfähigkeit* wird damit zum zentralen normativen Legitimitätstest institutioneller Arrangements - als hilfswieser Ersatz für die in der Regel unaufhebbare Tatsache, daß keine (Verfassungs-)Regel in Sicht ist, deren einstimmige Verabschiedung in einem großen Kollektiv realiter nicht prohibitiv hohe Entscheidungskosten verursachen würde. Auf den tatsächlichen Abschluß eines Vertrags kommt es dabei gar nicht an; die Frage nach der *Möglichkeit* eines Vertrags ist nur das analytische Instrument, entgegenstehende Interessen und damit die eventuelle Illegitimität einer kollektiven Entscheidung herauszufinden.<sup>38</sup> Theorien des hypothetischen Gesellschaftsvertrags haben also immer dort ihre Rechtfertigungsaufgabe zu erfüllen, wo eine tatsächliche Legitimation durch Einstimmigkeit *nicht* erreicht ist. Sie dienen der Legitimation des Herbeiführens von Situationen (Regeln), die dem Status quo nach dem Pareto-Kriterium unvergleichbar sind und daher nicht Gegenstand eines realen Vertrags der wirklich Betroffenen sein könnten.<sup>39</sup> In diesem Sinne sind die Ansätze des modernen Kontraktualismus „Vertragstheorien ohne Vertrag“<sup>40</sup>.

Das legitimationstheoretische Argument der Theorien des hypothetischen Gesellschaftsvertrags schlägt dem Theorieadressaten ein Gedankenexperiment der folgenden Art vor:

„Wir wollen eine fiktive Entscheidungssituation fiktiver Entscheidungsträger definieren, die so gerecht ist, daß alle Entscheidungen, die in ihr einstimmig getroffen werden, ebenfalls als gerecht angesehen werden können. Wie immer das unter diesen Umständen zustande gekommene Entscheidungsergebnis auch lautet: seine Realisierung durch die praktische Politik wäre aus individualistischer Sicht legitimiert.“<sup>41</sup>

Indem für jedes konkret zu lösende Problem der Vertragstheorie nachzuweisen ist, daß die jeweils zur Diskussion stehende Regel - aus Sicht der konstitutionellen Entscheidungssituation - eine Pareto-Verbesserung darstellt, läßt sich die *konstitutionelle soziale Auswahlregel (CCR)* der Theorien des hypotheti-

<sup>38</sup> Daher trifft das von *Brennan/Lomasky* (1993), S. 135-138, vorgetragene Argument, daß die Wähler im hypothetischen Urzustand der *Rawlsschen* Vertragstheorie eine bestimmte Gerechtigkeitskonzeption zwar präferieren können, aber möglicherweise „expressiv“ - d.h. eine andere als die präferierte Lösung - wählen könnten (vgl. S. 31, Fußnote 24), nicht den Kern des vertragstheoretischen Legitimationsarguments: Auf den *Akt* der konstitutionellen Wahl kommt es nach der Theorie des hypothetischen Gesellschaftsvertrags überhaupt nicht an.

<sup>39</sup> Vgl. etwa *Rawls* (1992 [1980]), S. 82 f.: „Wann immer ... eine ausreichende Basis der Übereinstimmung zwischen Bürgern *unbekannt* oder *nicht anerkannt* ist, stellt sich die Aufgabe der Rechtfertigung einer Gerechtigkeitskonzeption ...“ (Hervorhebungen von mir; *C. M.*)

<sup>40</sup> *Kliemt* (1998), S. 633.

<sup>41</sup> Eine ähnliche Formulierung des vertragstheoretischen Gedankenexperiments findet sich bei *Zintl* (1983), S. 30 f.



schen Gesellschaftsvertrags mit Hilfe der starken konstitutionellen Relation wie folgt formulieren (mit  $K =$  „aus konstitutioneller Sicht kollektiv besser als“):

*Starke konstitutionelle Relation (K).* Für zwei beliebige pareto-unvergleichbare Regeln (Situationen)  $r$  und  $q$  gilt:  $r$  wird  $q$  kollektiv vorgezogen, wenn  $r$ , nicht aber  $q$ , im Vergleich zu einer fiktiven und als (verfahrens-)gerecht empfundenen Verfassungswahlsituation  $z$  eine starke Pareto-Verbesserung ist:

$$r K q \leftrightarrow r \bar{P} z \wedge \neg (q \bar{P} z).$$

Die starke konstitutionelle Relation  $K$  ist eine kollektive Auswahlregel (CCR), falls gilt:  $r P q \leftrightarrow r K q$  (mit  $P =$  „aus Sicht des gesamten Kollektivs besser als“).

Die Feststellung einer konstitutionellen Wohlfahrtserhöhung ist somit ein zweistufiger Prozeß. In einem ersten Schritt ist festzustellen, ob die vorgeschlagene Regel  $r$  im Vergleich zur gedachten Situation der Fairneß  $z$  eine Pareto-Verbesserung darstellt (d.h.  $r \bar{P} z$ ). Dieser erste Schritt indes ist, was häufig übersehen wird, in den Theorien des hypothetischen Gesellschaftsvertrags nur eine notwendige Bedingung, um  $r$  zu realisieren. Ist  $q$  der bereits realisierte Gesellschaftszustand, so könnte es nämlich sein, daß auch  $q$  eine Pareto-Verbesserung gegenüber  $z$  darstellt. In diesem Fall wären sowohl  $r$  als auch  $q$  legitime Institutionen. Als moralisch geboten wird man die Herbeiführung von  $r$  aber nur dann betrachten, wenn der zu verlassende Zustand  $q$  aus konstitutioneller Perspektive illegitim ist. Eine hinreichende Bedingung für die Realisierung von  $r$  ist daher der in einem zweiten Schritt geführte Nachweis, daß  $q$  nicht selbst eine Pareto-Verbesserung gegenüber  $z$  darstellt (d.h.  $\neg (q \bar{P} z)$ ), was impliziert, daß  $q$  und  $z$  - wegen Nutzenänderungen in verschiedene Richtungen beim Übergang von  $z$  nach  $q$  - nach dem Pareto-Kriterium nicht miteinander vergleichbar sind.<sup>42</sup>

---

<sup>42</sup> Da  $r \bar{P} z$ , ist zugleich ausgeschlossen, daß  $q$  im Vergleich zu  $z$  paretoinferior ist, also  $z \bar{P} q$ . Wegen der Transitivität der Pareto-Relation müßte dann auch gelten:  $(r \bar{P} z \wedge z \bar{P} q) \rightarrow r \bar{P} q$ , was der Annahme widerspräche, daß  $r$  und  $q$  pareto-unvergleichbar sind.

## II. Der methodologische Status des vertragstheoretischen Arguments

Wie aber läßt sich zeigen, daß  $r$  im Vergleich zur gedachten Entscheidungssituation  $z$  eine Pareto-Verbesserung (und daß  $q$  im Vergleich zu  $z$  keine Pareto-Verbesserung<sup>43</sup>) darstellt? Als Mittel für diesen Nachweis verwenden die Theorien des hypothetischen Gesellschaftsvertrags im allgemeinen einen in der ökonomischen Theorie üblichen Rationalwahlansatz. Methodologisch gesehen, hat das vertragstheoretische Argument die gleiche logische Struktur wie eine wissenschaftliche Erklärung bzw. Prognose zielintendierten, rationalen Individualverhaltens.

Allgemein besteht eine wissenschaftliche (deduktiv-nomologische) Erklärung<sup>44</sup> in der logischen Folgerung einer singulären, die Existenz eines Ereignisses  $e$  behauptenden Aussage  $E$  - dem Explanandum - aus einer Menge von Prämissen, dem Explanans. Dieses Explanans besteht aus mindestens einer deterministischen Gesetzesaussage  $G$  mit strikter Universalität und Konditionalität und zusätzlichen Antezedensbedingungen  $A$  in Form von singulären Existenzaussagen, die bestimmte,  $e$  betreffende Ereignisse beschreiben. Einen singulären Sachverhalt  $e$  zu erklären, heißt mithin zu folgern:  $G_1, \dots, G_n$  zusammen mit  $A_1, \dots, A_m$  implizieren  $E(e)$ . Diese Ableitungsbeziehung ist eine kausale Erklärung in dem Sinne, daß die Antezedensbedingungen  $A$  die Ursache der vom Explanandum-Satz beschriebenen Wirkung angeben.<sup>45</sup>

Eine wissenschaftliche Prognose - eine Behauptung über ein zukünftiges Ereignis - hat die gleiche logische Struktur wie eine deduktiv-nomologische Erklärung.<sup>46</sup> Gesetze und Anwendungsbedingungen werden hier als Projectans, der daraus gefolgerte singuläre Existenzsatz als Projectandum bezeichnet. Erklärung und Prognose unterscheiden sich nur bezüglich der sog. „pragmatischen Gegebenheitsrelation“: Bei einer Erklärung ist zunächst allein der Explanandum-Satz  $E$  gegeben, und gesucht wird eine Begründung - in Form von Gesetzen und erfüllten Antezedensbedingungen - für das von  $E$  behauptete Ereignis  $e$ . Bei der Prognose hingegen sind bestimmte singuläre Ereignisse gegeben, die als Anwendungsfälle bekannter Gesetzeshypothesen interpretiert werden; hieraus

---

<sup>43</sup> Der Nachweis des Vorliegens der hinreichenden Bedingung für die Realisierung von  $r$  hat die gleiche logische Struktur wie der Nachweis ihrer notwendigen Bedingung mit dem einzigen Unterschied, daß beide zu einem entgegengesetzten Ergebnis führen (sollen). Die folgenden Ausführungen beschränken sich daher auf die notwendige Bedingung.

<sup>44</sup> Vgl. grundlegend *Hempel/Oppenheim* (1948) und *Hempel* (1977) sowie *Stegmüller* (1969).

<sup>45</sup> Vgl. *Popper* (1987), S. 97; *Hempel/Oppenheim* (1948), S. 139.

<sup>46</sup> Vgl. *Hempel/Oppenheim* (1948), S. 138 f.; *Hempel* (1977), S. 43 ff. Vgl. jedoch kritisch zu dieser Identitätsbehauptung *Stegmüller* (1969), S. 153 ff.

wird dann das Auftreten des Ereignisses *e* als logische Folge des Vorliegens des Projectans vorausgesagt.

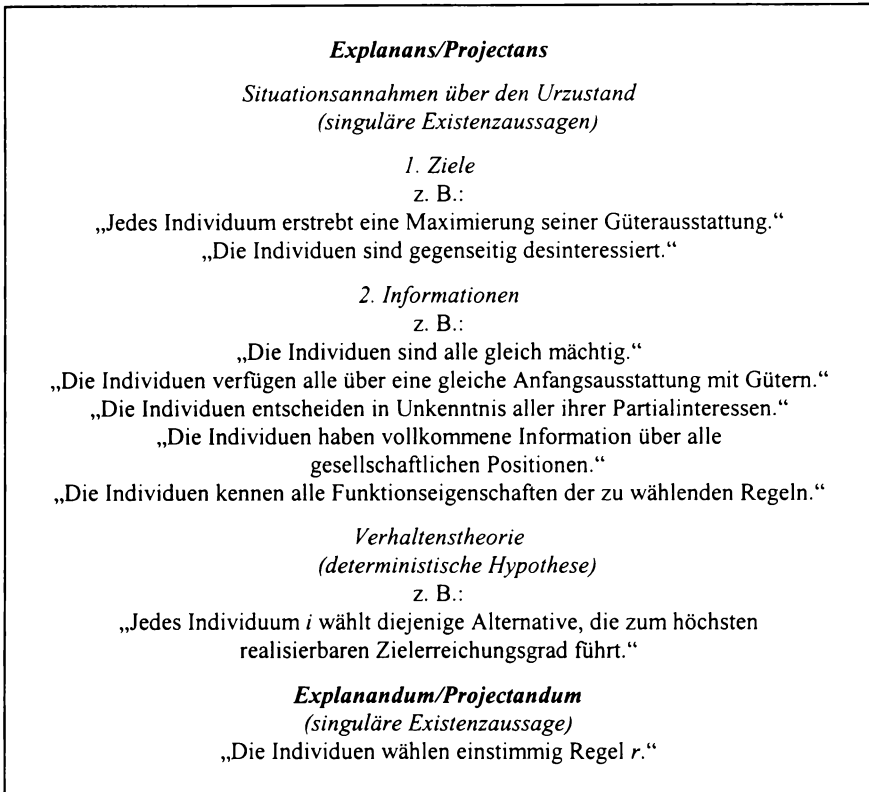


Abbildung 1.1: Die logische Struktur der vertragstheoretischen  
Als-ob-Erklärung/-Prognose

In einer sozialwissenschaftlichen *Rational-choice*-Erklärung bzw. -Prognose finden sich im Explanans als Gesetz die ökonomische Maximierungshypothese (die *Rationalverhaltenstheorie*) und als Antezedensbedingungen eine Beschreibung der Entscheidungssituation (die *Situationsannahmen*), bestehend aus den Zielen des Entscheidungsträgers und seinen technologischen Informationen über die Entscheidungsalternativen sowie aus singulären Informationen über Machtverhältnisse, Kosten oder andere entscheidungsrelevante Einflüsse. Das hieraus abgeleitete Explanandum enthält eine Information über die durch den Entscheidungsträger gewählte Handlungsalternative.<sup>47</sup>

<sup>47</sup> Vgl. Tietzel (1981), S. 132 f.; ders. (1985), S. 21-25.

Das vertragstheoretische Argument hat die gleiche logische Struktur wie eine „normale“ ökonomische Rationalwählerklärung (vgl. Abbildung 1.1). Jede der Theorien des hypothetischen Gesellschaftsvertrags trifft - von der bei allen Theorien gleichen Annahme der konstitutionellen Fähigkeit zur Selbstbindung abgesehen - ihre eigenen (zum Teil impliziten) *Situationsannahmen* über einen „gerechten“ Urzustand, etwa über die Machtverteilung, die Informiertheit oder die „zulässigen“ Präferenzen der fiktiven Entscheider auf der konstitutionellen Ebene<sup>48</sup>. Und jede dieser Annahmen, welche allesamt die Eigenschaften kontrafaktischer singulärer Existenzaussagen haben, verändert das Vertragsergebnis selbst und damit die Politikempfehlung, welche die Vertragstheorie als hypothetisch legitimiert erweisen soll.

Man könnte  $z$  etwa in der Weise explizieren, daß man annimmt, alle in dieser Situation getroffenen einstimmigen Entscheidungen fielen *freiwillig*<sup>49</sup>. Gesellschaftliche Institutionen sind folglich nur dann als „legitim“ anzusehen, wenn sich schlüssig argumentieren ließe, daß sich alle Betroffenen freiwillig auf sie einigten. Mit dieser Festlegung erklärt man zugleich alle politischen Handlungen und Institutionen für illegitim, von denen angenommen werden kann, daß ihnen nicht alle Beteiligten freiwillig zustimmen würden: Andere Länder dürften nicht annektiert, Minderheiten nicht diskriminiert und „Frauen-Quoten“ im öffentlichen Dienst nicht erlassen werden, da nicht davon auszugehen ist, daß alle Bürger im annektierten Land, alle Repräsentanten einer Minderheit und alle Männer den gegen sie gerichteten Handlungen aus freien Stücken zustimmen würden.

Auch eine Umverteilung von Einkommen und Vermögen von Reichen hin zu Armen wäre unter der Entscheidungsbedingung der Freiwilligkeit nicht akzeptabel. Sei etwa  $q$  die in einer Gesellschaft realisierte Situation ohne eine Redistributionsregel;  $r$  hingegen stehe für den Zustand derselben Gesellschaft nach erfolgter Umverteilung. Ist die vertragstheoretische Vergleichssituation  $z$  allein durch die Bedingung der Freiwilligkeit (Index „ $f$ “) gekennzeichnet (kurz:  $z_f$ ), so könnte nicht ohne weiteres davon ausgegangen werden, daß alle Reichen einer solchen Politik zustimmen würden; einige unter ihnen würden vermutlich ihr Veto einlegen, so daß ein Gesellschaftsvertrag mit einer solchen Redistributionsklausel nicht allgemein zustimmungsfähig wäre;  $r$  ist somit keine Pareto-Verbesserung gegenüber  $z_f$  und ist damit auch nach dem konstitutionellen Wohlfahrtskriterium  $K$  nicht der Situation  $q$  vorzuziehen:

<sup>48</sup> Ausgeschlossen sein könnten etwa malevolente Präferenzen von Akteuren, die nur darauf gerichtet sind, andere Individuen zu schädigen.

<sup>49</sup> Zu einer solchen Konzeption der Vertragstheorie, welche die Bedingung der *Freiwilligkeit* als einzige Wertprämisse über die Konsensnorm hinaus enthält vgl. *Vanberg* (1986).

$$\neg (r \bar{P} z_f) \rightarrow \neg (r K q).$$

Dies bedeutet aber nicht, daß Umverteilung aus vertragstheoretischer Sicht generell als illegitim anzusehen wäre. Nach Jean-Jacques Rousseau beispielsweise wäre ein Vertragsschluß zwischen Reichen und Armen - ein „Gesellschaftsvertrag der beiden Stände“<sup>50</sup> - notwendig ein unsittlicher Täuschungs- und Betrugsvertrag, der lediglich zum Ziel hätte, die Privilegien, welche die Reichen im Naturzustand genießen, in den gesellschaftlichen Zustand hinüberzuretten.<sup>51</sup> Ein solcher Vertrag verletzte nach Rousseau die grundlegende Gerechtigkeitsbedingung der *Gleichheit* der Kontraktparteien. Formale menschenrechtliche Gleichheit ohne materielle Gleichheit sei nur der Versuch, das Unterwerfungsverhältnis der Armen unter die Reichen zu verschleiern. Folglich hält Rousseau nur solche Gesellschaftsverträge für legitimitätsstiftend, die in einer Situation unter der Bedingung einer weitgehenden Gleichheit des Besitzes und dem Fehlen einer Schicht der Besitzlosen abgeschlossen worden sind<sup>52</sup>; wir wollen diese Situation  $z_g$  nennen.

Wären die gleichen Entscheider, die sich im Urzustand  $z_f$  nicht auf eine Umverteilungsregel  $r$  einigen konnten, in der Rousseauschen Vertragssituation  $z_g$ , so würden sie sich sehr wohl einstimmig darauf verständigen, im Gesellschaftsvertrag festzuschreiben, daß die konstitutionelle materielle Gleichheit aller Menschen auch in der postkonstitutionellen Gesellschaft gelten soll. Die Situation mit Umverteilung,  $r$ , wäre gegenüber  $z_g$  eine Pareto-Verbesserung, d.h.  $r \bar{P} z_g$ . Zugleich gilt dies nicht für die tatsächlich realisierte Situation  $q$  ohne Umverteilungsregel, da - ausgehend von  $z_g$  - beim Übergang zu  $q$  einige Abstimmungsteilnehmer - die in  $q$  Armen - schlechter gestellt würden. Es gilt also:  $\neg (q \bar{P} z_g)$ . Daher ist nach der Rousseauschen Vertragstheorie eine Gesellschaft mit Umverteilung (= Situation  $r$ ) einer solchen ohne Umverteilung (= Situation  $q$ ) vorzuziehen:

$$r \bar{P} z_g \wedge \neg (q \bar{P} z_g) \rightarrow r K q.$$

Die Individuen in  $z_g$  werden sich, in den Worten Rousseaus, auf einen Gesellschaftsvertrag verständigen, der „eine sittliche und rechtliche Gleichheit an die Stelle dessen setzt, was die Natur an physischer Ungleichheit unter den Menschen hervorbringen kann, und daß die Menschen, die möglicherweise nach Stärke und Begabung ungleich sind, durch Vertrag und Recht alle gleich werden.“<sup>53</sup> Eine egalisierende Redistribution von den Besitzenden zu den Habe-

<sup>50</sup> Rousseau (1977 [1755]), S. 50.

<sup>51</sup> Vgl. Rousseau (1977 [1755]), S. 50 f.; ders. (1977 [1762]), Buch I, Kap. 9, S. 26. Vgl. auch Fetscher (1968), S. 33-46; Kersting (1994), S. 144 ff.

<sup>52</sup> Vgl. Fetscher (1968), S. 34.

<sup>53</sup> Rousseau (1977 [1762]), Buch I, Kap. 9, S. 26.

nichtsen wäre insofern nicht nur ein Recht, sondern eine moralische Pflicht der praktischen Politik.

Welche Annahmen man als konstitutiv ansieht für die Gerechtigkeit der vertragstheoretischen Ausgangssituation  $z$  und welche nicht, ist eine Frage der persönlichen Bewertung, die keine intersubjektive Gültigkeit beanspruchen kann. In der Regel ist sie abhängig von den persönlichen - wenngleich zumeist unausgesprochenen - Gerechtigkeitsnormen des jeweiligen Vertragstheoretikers selbst und, wie man aus den angeführten Beispielen sieht, nicht zuletzt von dem Rechtfertigungsergebnis, das seine Theorie „legitimieren“ soll. John Rawls, dessen „Theorie der Gerechtigkeit“, neben der Vertragstheorie Buchanans, im Mittelpunkt dieser Arbeit stehen wird, gibt dies unumwunden zu, indem er schreibt: „Wir möchten den Urzustand so bestimmen, daß die gewünschte Lösung herauskommt.“<sup>54</sup> Ganz so frei, wie es hier anscheint, ist der Kontrakttheoretiker in der Ausgestaltung der vertragstheoretischen Wahlsituation gleichwohl nicht, zumindest dann nicht, wenn sich möglichst viele Menschen den Prämissen seines Ansatzes und den aus ihnen gefolgerten Politikempfehlungen anschließen können sollen. In der Regel wird ein Autor die konstitutionellen Entscheidungsbedingungen daher nach Maßgabe solcher Werturteile explizieren, die im jeweiligen Adressatenkreis der Theorie zum Bestand der weitverbreiteten und tief verwurzelten moralischen Normen gehören. Vertragstheorien werden daher im allgemeinen keine universellen, interkulturell akzeptablen Rechtfertigungsargumente sein, sondern ihre Rechtfertigungsfunktionen bestenfalls relativ zu einer bestimmten Wertegemeinschaft entfalten können.<sup>55</sup>

Ein zweiter wesentlicher Bestandteil des vertragstheoretischen Rationalwahlarguments ist - wie bei allen *Rational-choice*-Argumenten - die jeweilige *Verhaltenshypothese*, welche die Ansätze des hypothetischen Vertrags verwenden. Eine solche Verhaltenstheorie muß darlegen, daß - gegeben die speziellen Informationsbedingungen, die konkrete Machtverteilung und die übrigen Annahmen, welche die konstitutionelle Vertragssituation definieren - die betrachteten Individuen einstimmig eine bestimmte Kollektiventscheidung treffen würden. Die beiden im Mittelpunkt dieser Arbeit stehenden Vertragstheorien von Rawls und Buchanan greifen hier, wie einst Hobbes, auf ein Konzept egoistischer Zweckrationalität zurück; ihre Verhaltenstheorie ist die Rationalitätshypothese der neoklassischen Wirtschaftstheorie, die besagt, daß rationale Individuen aus der Menge der verfügbaren Alternativen jene auswählen, die zum höchsten realisierbaren Zielerreichungsgrad führen.

---

<sup>54</sup> Rawls (1994 [1971]), S. 165.

<sup>55</sup> Daß sie die spezifische Sozialisation von Menschen in den westlichen Demokratien des 20. Jahrhunderts widerspiegele, kritisiert Pettit (1974) an der Rawlsschen „Theorie der Gerechtigkeit“.

Aus Situationsannahmen und Rationalitätshypothese folgen die Theorien des hypothetischen Sozialvertrags die Explanandum-Aussage *E*, die besagt, auf welche Regel *r* sich die Individuen im Urzustand einigen werden. Dieses singuläre Ereignis *e* - die Einigung auf Regel *r* - wird mittels des kontraktualistischen Arguments, je nachdem, wo man sich auf dem fiktiven Zeitpfad einordnet, entweder hypothetisch erklärt oder prognostiziert. Buchanan interpretiert seine in „Die Grenzen der Freiheit“ vorgetragene Vertragstheorie vor allem als Erklärung: Er versucht zu zeigen, daß es Situations- und Verhaltensannahmen gibt, die logisch hinreichend sind, um ein gewünschtes Rechtfertigungsergebnis - hier: die Existenz eines liberalen Rechtsstaats - als Ergebnis eines Vertragsschlusses aller Betroffenen in einem gedachten Zustand der Anarchie zu behaupten.<sup>56</sup> Meist - etwa auch im Falle von Rawls' „Theorie der Gerechtigkeit“ - erscheint die Vertragstheorie jedoch als ein Prognoseargument. Der Theorieadressat wird zunächst - ohne ihm das Legitimationsziel des jeweiligen Ansatzes zu nennen - davon überzeugt, daß die Situationsannahmen ein auch nach seiner Meinung gerechter Ausgangspunkt für das Treffen legitimer Entscheidungen sind; daran anschließend wird „prognostiziert“, welche Regel sich unter diesen Umständen als Ergebnis einer einstimmigen Entscheidung ergeben wird.

Bestehen die von der Vertragstheorie explizit oder implizit behaupteten Deduktionsbeziehungen tatsächlich, so erfüllt das konstitutionelle Legitimationsargument die *logischen* Adäquatheitsbedingungen wissenschaftlicher Erklärungen und Prognosen.<sup>57</sup> Diese postulieren erstens, daß der Explanandum-Satz *E* aus dem Explanans logisch ableitbar sein muß. Daß dies jeweils der Fall ist, legen letztlich alle Vertragstheorien nahe. Nicht alle jedoch behaupten dies zu recht, wie wir im Verlauf dieser Arbeit sehen werden; in diesem Fall bietet das kontraktualistische Argument nicht nur eine fiktive, sondern darüber hinaus auch

---

<sup>56</sup> „Das Problem besteht ... in dem Versuch, die Beziehungen zwischen den Individuen und zwischen Individuen und der Regierung zu *erklären* und zu *verstehen*. Für diesen Zweck könnte es notwendig sein, ungeachtet der historisch aufgezeichneten Fakten verschiedene Als-ob-Modelle für die Entstehungsgeschichte von Begriffen zu erstellen. Das *Schwergewicht* sollte dabei auf '*erklären*' und '*verstehen*' liegen ... Dieses Buch ist ein Versuch zu *erklären*, wie das '*Recht*', '*die Eigentumsrechte*' und '*die Verhaltensregeln*' aus nichtidealisiertem, vom Eigeninteresse bestimmten Verhalten abgeleitet werden können, ohne vorauszusetzen, daß in einer Art Naturzustand Gleichheit bestünde ...“ Buchanan (1984 [1975]), S. 77 f. (Hervorhebungen von mir; C. M.).

<sup>57</sup> Zu den logischen und empirischen Adäquatheitsbedingungen deduktiv-nomologischer Erklärungen vgl. Hempell/Oppenheim (1948), S. 137 f.; Stegmüller (1969), S. 86-90. Auf die Probleme, die sich im Zusammenhang mit diesen Adäquatheitsbedingungen ergeben können, soll hier nicht weiter eingegangen werden; vgl. im einzelnen Küttner (1976).



eine lediglich „partielle Erklärung“<sup>58</sup>. Zweitens muß das Explanans mindestens eine Gesetzesaussage enthalten, die eine Antwort auf die Frage gibt, warum das Eintreten des durch den Explanandum-Satz behaupteten Ereignisses *e* zu erwarten war. Wie in anderen Rationalwahlansätzen auch, übernimmt diese Aufgabe in der Vertragstheorie die ökonomische Maximierungshypothese. Drittens müssen die Gesetzesaussagen im Explanans *Informationsgehalt* besitzen, d.h. falsifizierbar sein.<sup>59</sup> Die ökonomische Maximierungshypothese besagt, daß rationale Individuen aus der Menge der verfügbaren Alternativen jene auswählen, die zum höchsten Zielerreichungsgrad führen. Da sie damit alle Handlungen bis auf eine ausschließt, ist der Informationsgehalt dieser Verhaltenstheorie extrem hoch.

Das hypothetische Vertragsargument verletzt aber stets die *empirische* Adäquatheitsbedingung, die verlangt, daß alle Aussagen im Explanans faktisch wahr sein müssen; für empirische Gesetzesaussagen, die aufgrund ihrer strikten Universalität nicht verifizierbar sind, begnügt man sich im allgemeinen mit einem hohen Maß an empirischer Bewährung. Wahre Annahmen im Explanans des vertragstheoretischen Arguments dürften eher die Ausnahme denn die Regel sein. Die Entscheidungssituation ist meist fiktiv. Die vertragstheoretischen Entscheidungsbedingungen - etwa die Situation der geordneten Anarchie bei Locke (und Nozick)<sup>60</sup> oder die Bedingungen völliger sozio-ökonomischer Gleichheit bei Rousseau - haben, soweit bekannt, in dieser Form zu keinem Zeitpunkt und an keinem Ort im Universum jemals tatsächlich vorgelegen, noch ist zu erwarten, daß diese Annahmen irgendwann einmal irgendwo erfüllt sein werden.

Nicht einmal das von den modernen Kontraktansätzen verwendete ökonomische Verhaltensgesetz erfüllt die Adäquatheitsbedingung hoher empirischer Bewährung. Daß die Rationalverhaltenshypothese nicht nur hochgradig falsifizierbar, sondern sogar falsifiziert - empirisch falsch - ist<sup>61</sup>, wird unter Ökono-

---

<sup>58</sup> „Partielle Erklärungen“ sind nach *Stegmüller* (1969), S. 108-110, solche, in denen das vorgeschlagene Explanans nicht ausreicht, um die Explanandum-Aussage in all den Hinsichten zu erklären, in denen es beschrieben wird. Eine partielle Erklärung liefert statt dessen nur eine Erklärung für einige dieser Aspekte.

<sup>59</sup> Der empirische oder Informationsgehalt einer Gesetzesaussage ist ihre Eigenschaft, informativ in bezug auf die Realität zu sein; er bezeichnet die Menge aller Zustände in ihrem Anwendungsbereich, die von der Aussage ausgeschlossen werden. Der Informationsgehalt steigt mit der Allgemeinheit (Universalität) der Wenn-Komponente und der Bestimmtheit (Präzision) der Dann-Komponente der generellen Aussage. Vgl. grundlegend *Popper* (1994) sowie *Chmielewicz* (1979), S. 123-129.

<sup>60</sup> Vgl. *Locke* (1992 [1689]), §§ 4 ff., S. 201 ff.; *Nozick* (1976 [1974]), S. 25 ff.

<sup>61</sup> Vgl. zur Informationshaltigkeit und empirischen Falschheit dieser Annahme im einzelnen *Tietzel* (1985), S. 83-97.



men heute kaum mehr bestritten.<sup>62</sup> Da die Rationalitätsannahme trotzdem nach wie vor weitgehend verwendet wird, können ökonomische *Rational-choice*-Ansätze, in der Diktion von Hayeks, niemals mehr sein als „Erklärungen im Prinzip“ - im Unterschied zu „Erklärungen im Detail“<sup>63</sup>; statt adäquater sind ökonomische Ansätze lediglich „idealisierte Erklärungen“ realer Phänomene.

Idealisiert werden Erklärungen genannt, welche aufgrund fehlender oder nicht feststellbarer Wahrheit der Gesetzesaussage im Explanans die empirische Adäquatheitsbedingung von Erklärungen nicht erfüllen, weil die Faktoren, von denen ein zu erklärender Sachverhalt abhängt, zu komplex oder nicht bekannt sind.<sup>64</sup> Idealisierte *Als-ob*-Erklärungen, welche die logischen, aber nicht die empirische Adäquatheitsbedingung erfüllen, sind immer nur *hinreichende* Bedingungen, um die Aussage *E* zu behaupten. Da aber die Adäquatheitsbedingung der faktischen Wahrheit der Explanans-Aussagen verletzt ist, ist das erklärende Argument niemals *notwendig* zur Ableitung des Explanandum-Satzes; der gleiche Explanandum-Satz kann möglicherweise auch aus einer Vielzahl ganz anderer, ebenso denkbarer Explanantia deduziert werden. Im Zusammenhang mit der Erklärung des Auftretens bestimmter singulärer empirischer Ereignisse ist daher die bewußte Verwendung eines faktisch falschen Explanans nur insofern akzeptabel, als es erstens in der überwiegenden Anzahl von Anwendungsfällen zu Erklärungserfolgen führt und es zweitens als ein vorläufiges heuristisches Verfahren betrachtet wird.<sup>65</sup>

Im Kontext eines hypothetischen Vertragsarguments hingegen spielt die empirische Wahrheit der Rationalitätshypothese ebensowenig eine Rolle wie die faktische Falschheit der singulären Existenzaussagen, welche die Entscheidungssituation der fiktiven Entscheidungsträger beschreiben. Für die Akzeptanz eines vertragstheoretischen Gedankenexperiments kommt es nicht darauf an, daß die Annahmen im Explanans wahr bzw. (im Falle strikt universaler Hypothesen) in hohem Maße bewährt sind; was zählt, ist allein ihre Funktion in dem konstruierten normativen Argument. Die Falschheit einiger oder aller Annahmen im Explanans dient nicht lediglich einer vorübergehenden Heuristik, sondern ist konstitutives Kennzeichen des vertragstheoretischen Arguments.

Natürlich könnte man der Vertragstheorie auch eine alternative und „realitätsnähere“ Hypothese über menschliches Verhalten zugrunde legen, etwa eine

---

<sup>62</sup> Die Falschheit der Rationalitätsannahme ist Ausgangspunkt der modernen ökonomischen „Anomalien“-Forschung. Unter einer „Anomalie“ wird eine Abweichung von empirisch beobachtetem von solchem Verhalten verstanden, das nach Maßgabe des ökonomischen Rationalwahlansatzes zu erwarten wäre. Übersichten über diese Literatur geben *Schoemaker* (1982); *Machina* (1987); *Frey/Eichenberger* (1989a); *dies.* (1989b).

<sup>63</sup> Vgl. zu diesen Begriffen von *Hayek* (1967).

<sup>64</sup> Vgl. *Tietzel* (1985), Kap. 9; *ders.* (1986).

<sup>65</sup> Vgl. *Tietzel* (1985), S. 118; *ders.* (1986), S. 320.

Hypothese „begrenzter Rationalität“<sup>66</sup>. Eine solche universale Gesetzesaussage hätte aber aus vertragstheoretischer Sicht den Nachteil, daß solche Theorien begrenzter Rationalität ihre größere „Realitätsnähe“ zwangsläufig mit einem - im Vergleich zur Maximierungshypothese - geringeren Informationsgehalt erkaufen.<sup>67</sup> „Verbietet“ eine solche Hypothese jedoch weniger als alle bis auf eine einzige Handlungsalternative, so wird das konstitutionelle Entscheidungsergebnis unbestimmt; ist es nämlich nach Maßgabe einer anderen Rationalitätshypothese möglich, unter bestimmten Entscheidungsbedingungen nicht nur eine einzige Alternative zu wählen, so wäre nicht mehr gesichert, daß wirklich die „gewünschte Lösung herauskommt“. Wenn zwei verschiedene Individuen, obwohl rational, in ein und derselben Situation unterschiedliche Entscheidungsalternativen präferieren, wäre nicht einmal mehr garantiert, daß überhaupt eine Einigung zwischen ihnen zustande kommt.

Darüber hinaus weist der neoklassische *Homo oeconomicus* den Vorzug auf, daß er die Interessen der realen Individuen, die unter den Regeln, auf die sich die fiktiven Entscheider einigen, gegebenenfalls leben müssen, besonders gut vertritt. Wenn es schon nicht möglich ist, daß die wirklich von einer Kollektiventscheidung Betroffenen ihre Wünsche in dem fiktiven Verfassungskonvent äußern können, so erscheint es für die Theorieadressaten einfacher, einen Ansatz zu akzeptieren, in dem die Interessen ihrer jeweiligen Gesellschaftsposition - die des Reichen oder des Armen, die des Politikers oder des Steuerzahlers - auf bestmögliche Weise eigeninteressiert vertreten wird. Nachgiebigkeit, unzureichende Informiertheit und gelegentliche Irrationalität wären zwar Eigenschaften, die realen Menschen eher entsprächen als jene des ökonomischen Rationalakteurs. Wenn es aber um die Vertretung unserer Interessen gegenüber anderen - erst recht in einer so wichtigen Angelegenheit wie der Verfassungswahl - geht, wünschen wir uns doch wohl alle lieber verhandlungsstarke Egoisten zu sein, die alle verfügbaren Informationen optimal zu Entscheidungen von bestechender Rationalität ausnutzen, als die fehlbaren Menschen, die wir sind. Die hypothetische Vertragstheorie legt uns die Attraktivität der zu „begründenden“ Norm also mit Hilfe eines *A-fortiori*-Arguments ans Herz: Wenn eine Regel im Eigeninteresse solch kühl kalkulierender Egoisten liegt, dann liegt sie erst recht im Eigeninteresse beschränkt rationaler realer Menschen.

Von den idealisierten Erklärungen zielintendierten, rationalen Verhaltens, wie sie in der ökonomischen Theorie üblicherweise verwendet werden, unterscheiden sich die vertragstheoretischen Erklärungs- bzw. Prognoseargumente dadurch, daß sie gar nicht erst den Versuch unternehmen, das Verständnis realer Phänomene zu befördern. Idealisiert sind die vertragstheoretischen Argumente

<sup>66</sup> Vgl. zur Hypothese begrenzter Rationalität Simon (1955); ders. (1993 [1983]).

<sup>67</sup> Vgl. Tietzel (1985), S. 90.

nicht, wie die meisten sozialwissenschaftlichen Rationalwahlerklärungen, um ein ansonsten zu komplexes und nicht handhabbares Erklärungs- oder Prognoseargument zu vereinfachen und wenigstens „approximative Erklärungen“<sup>68</sup> realer singulärer Ereignisse liefern zu können. Vielmehr dienen die Annahmen im Explanans hypothetischer Kontraktargumente allein dazu, vom jeweiligen Vertragstheoretiker gewünschte Gerechtigkeitsnormen dem Theorieadressaten dadurch näher zu bringen, daß dargelegt wird, daß auch vollkommen rationale Individuen unter fairen Entscheidungsbedingungen sich nach Maßgabe dieser Prinzipien verhalten würden. Anders also als die idealisierten approximativen *Rational-choice*-Erklärungen realer Ereignisse lassen sich hypothetisch-kontraktualistische Argumente grundsätzlich nicht durch Vervollständigung oder Konkretisierung in adäquate Erklärungen überführen.<sup>69</sup>

Seinen Sinn bezieht ein fiktives vertragstheoretisches Gedankenexperiment gerade aus der Falschheit seiner Prämissen: Wären die Annahmen des Explanans tatsächlich wahr, so würden sich die realen Individuen ohne Probleme auf eine Lösung ihres Zusammenlebens einigen; angesichts dieser tatsächlichen Legitimation wäre eine hypothetische Rechtfertigung, wie sie die Vertragstheorien anstreben, überflüssig. Der Kalkül der fiktiven Rationalentscheider in ihrer fiktiven Situation der Verfassungswahl macht, wenn überhaupt, nur dort Sinn, wo reale Individuen unter realen Entscheidungsbedingungen einen tatsächlichen Konsens verfehlen. Die Vertragstheoretiker lassen sich hierdurch nicht entmutigen; unbeirrt suchen sie nach der fiktiven Einigung der faktisch Uneinigen. Demjenigen aber, der nicht wahrhaben will, was wahr ist, bleibt nur die Flucht aus der Realität. Die Falschheit der Annahmen im Explanans des kontraktualistischen Arguments ist nicht eine unintendierte Imperfekteit der Theorie des hypothetischen Gesellschaftsvertrags, sondern ihre konstitutive Voraussetzung.

### III. Fiktive Verträge und faktische Verpflichtungen

Aufgrund der faktischen Falschheit mindestens einiger Annahmen in seinem Explanans ist das Gedankenexperiment der Theorien des hypothetischen Gesellschaftsvertrags stets ein sehr schwaches Argument. Selbst dann, wenn es im Idealfall alle *logischen* Adäquatheitsbedingungen erfüllt, die erklärende bzw. prognostizierende Elemente sinnvollerweise erfüllen sollten, vermag es bestenfalls nachzuweisen, daß es denkbare, von Natur- und Sozialgesetzen nicht

<sup>68</sup> Diesen Begriff verwendet *Barr* (1974), S. 48.

<sup>69</sup> Zu dieser Eigenschaft idealisierter approximativer Erklärungen vgl. *Tietzel* (1986), S. 317.

grundsätzlich ausgeschlossene Welten gibt, in denen selbst rationale Egoisten sich so verhalten würden, als ob sie eine bestimmte Norm teilten.

Psychologisch mag es einen Unterschied machen zu wissen, daß es eine solche Situation gibt; der einzelne kann sich dann dazu angehalten fühlen, auch in anderen, realen Situationen so zu entscheiden, wie er im Urzustand handeln würde. In logischer Hinsicht beruht eine solche Folgerung jedoch auf einem unzulässigen Induktionsschluß: Aus der Rationalität einer Handlung in einer *bestimmten* Situation kann nicht auf die Rationalität dieser Handlung in *allen* Situationen geschlossen werden.

Alle Theorien des hypothetischen Sozialkontrakts stoßen damit auf die grundlegende Einschränkung, daß ein noch so elaboriertes vertragstheoretisches Argument nicht ein reales Recht oder einer reale Verpflichtung realer Individuen bewirken kann. Denn trivialerweise kann ein fiktiver Vertrag zwischen fiktiven Individuen nur fiktive Individuen verpflichten.<sup>70</sup> Ein hypothetischer Vertrag ist nämlich „not simply a pale form of an actual contract; it is no contract at all.“<sup>71</sup> Eine Theorie des fiktiven Gesellschaftsvertrags kann daher höchstens Bedingungen formulieren, denen sich möglichst viele anschließen können (sollten), um diese schließlich auf bestimmte Folgerungen hieraus festlegen zu können. Der hypothetische Gesellschaftsvertrag ist damit

„... bestenfalls ein zwingendes Argument, das in Ausübung der Macht der logischen Schlußregeln jedem die Freiheit nimmt, die Konklusion zu verwerfen, wenn er die Prämissen akzeptiert. Und wenn er die Prämissen akzeptiert und sich somit hinsichtlich der Konklusion festgelegt hat, dann hat er einen guten Grund, sich entsprechend der in diesem Argument elaborierten Prinzipien zu verhalten. Aber damit ist er beileibe nicht im Sinn einer selbstaufgelegten Pflicht gebunden.“<sup>72</sup>

Zumindest Buchanan scheint Kritik dieser Art an den Vertragstheorien dadurch begegnen zu wollen, daß er die Vorläufigkeit seiner vertragstheoretischen Überlegungen betont. Anders als in den übrigen Theorien des hypothetischen Gesellschaftsvertrags, sollen die von ihm vorgeschlagenen konstitutionellen Tests *keine Legitimitätstests* sein, deren Erfüllung die praktisch Handelnden dazu ermächtigte, das „legitimierte“ Ergebnis herbeizuführen. Nach Buchanan ist statt dessen der real existierende Status quo stets zu respektieren.<sup>73</sup> Der gedankenexperimentelle Nachweis der Möglichkeit eines hypothetischen Gesellschaftsvertrags kann nach dieser Position lediglich eine „mutmaßliche Legitimität“ begründen. Denn nach Maßgabe des normativen Individualismus könne es nicht darauf ankommen, ob fiktive *Homines oeconomici* eine kollektive Ent-

<sup>70</sup> Vgl. Sumner (1977), S. 43; Kliemt (1986), S. 227.

<sup>71</sup> Dworkin (1975 [1973]), S. 18.

<sup>72</sup> Kersting (1994), S. 33.

<sup>73</sup> Vgl. Buchanan (1977b), S. 138.

scheidung für paretoverbessernd halten, sondern nach Buchanan zählt nur die Einschätzung der realiter von einem institutionellen Arrangement Betroffenen.<sup>74</sup>

Das zentrale Mittel, um diesen realen und nicht lediglich hypothetischen Konsens zu erreichen, ist nach Buchanan das der *Kompensation* der im Status quo privilegierten Personen oder Gruppen. Das Argument der Theorie des hypothetischen Gesellschaftsvertrags - dies reklamiert Buchanan auch für die Theorie von Rawls<sup>75</sup> - soll dabei nur die Aufgabe haben, die erreichbaren Pareto-Verbesserungen aufzuzeigen. Es soll die zu kompensierenden Verlierer einer Kollektiventscheidung zur eigenen Einsicht in die Notwendigkeit des hypothetisch legitimierte Politikwechsels bringen und so nicht zuletzt dazu beitragen, das Ausmaß der tatsächlich erforderlichen Kompensationszahlungen zu reduzieren.<sup>76</sup>

---

<sup>74</sup> So schreibt *Buchanan* (1977b), S. 139 f.: „Diagnosis based on hypothetical contract can be useful in making initial evaluations of existing institutions, evaluations that can become inputs in attaining consensus upon suggestions or proposals for change. But hypothetical contract provides no justification for the imposition of change nonvoluntarily. In this respect hypothetical contract has no advantage over the more familiar transcendental norms - economic efficiency, natural law, reason, truth, God's will, and so on. The contractarian construction used in evaluative diagnosis finds its advantage only when attempts are made to reach agreement on change.“ In diesem Sinne formulieren auch *Brennan/Buchanan* (1988 [1980]), S. 15, in bezug auf die von ihnen konstitutionell begründeten Besteuerungsregeln: „Der letztliche Test für die Wünschbarkeit kann natürlich nur die Übereinstimmung selbst sein. Man kann nicht hoffen, aufgrund von Überlegungen, die allein auf Mutmaßungen beruhen, ein System von Steuerregeln festzulegen, die *tatsächlich* allseits vorteilhaft sind, sondern nur ein Regelsystem, von dem man dies *erwartet*. In diesem Sinn haben die normativen Schlußfolgerungen unserer Diskussion einen strikt vorläufigen Charakter. Alles, was man von uns erwarten kann, ist die Vorlage einer Reihe von Steuerregeln, die als Diskussionsgrundlage für eine verfassungsgebende Versammlung über die öffentlichen Finanzen dienen können.“ (Hervorhebungen im Original) - Vgl. zu dieser *Buchanan*-Interpretation auch *Pies* (1994), S. 49.

<sup>75</sup> *Buchanan* (1977b), S. 141, Fußnote 6, behauptet auch für *Rawls*’ „Theorie der Gerechtigkeit“, daß diese von ihrem Autor nicht dazu gedacht sei, unmittelbare Legitimität zu stiften, welche die praktische Umsetzung der *Rawls*’schen Gerechtigkeitskonzeption auch gegen den Willen der Betroffenen rechtfertige.

<sup>76</sup> „To the extent that argument, debate, and discussion can generate consensus on change, on reform, the need to resort to more overt means of compensation is reduced or eliminated.“ *Buchanan* (1977b), S. 141.

## § 2 Die starke Vertragstheorie

Das vertragstheoretische Argument, so wurde im vorangegangenen § 1 gezeigt, verletzt stets die empirische Adäquatheitsbedingung deduktiv-nomologischer Erklärungen oder Prognosen. Nur dann, wenn die wirklich von einer Regel betroffenen Individuen sich realiter *nicht* einigen können, vermag der Nachweis, daß sich fiktive Individuen in einem fiktiven Naturzustand auf eine bestimmte Lösung ihres Zusammenlebens einigen würden, einen gewissen Sinn zu haben. Dieser allenfalls psychologisch, aber niemals logisch zwingende Appell der Vertragstheorie, in der Realität die gleiche Regel zu wählen, die auch die fiktiven Entscheider in ihrer fiktiven Situation der Verfassungswahl wählen würden, setzt jedoch voraus, daß die von der jeweiligen Vertragstheorie behauptete einstimmige Akzeptanz einer bestimmten Regel in der fiktiven Situation der Verfassungswahl nach den Regeln der Logik überhaupt eine zulässige Folgerung aus den Annahmen im Explanans des Kontraktansatzes ist. Das vertragstheoretische Argument muß also wenigstens die *logischen* Adäquatheitsbedingungen erklärender Argumente erfüllen.

In den folgenden Kapiteln dieser Arbeit werde ich jedoch argumentieren, daß die hier untersuchten Kontrakttheorien von Buchanan und Rawls selbst diesem sehr bescheidenen Anspruch nicht gerecht werden. In § 2 werde ich zunächst für die paretianische, sog. starke Vertragstheorie, die Buchanan in „Die Grenzen der Freiheit“ vorgelegt hat, die These zu begründen versuchen, daß ihre Annahmen logisch nicht hinreichend sind, um die als wahrscheinlich ausgegebene konstitutionelle Wahl eines Rechtsschutzstaates zu behaupten. In den folgenden §§ 3 bis 5 werde ich darüber hinaus nachzuweisen versuchen, daß die vertragstheoretische Analyse auch dann ein in logischer Hinsicht unvollständiges Argument bleibt, wenn - wie von Rawls und, in bestimmten Teilen seines Werkes, auch von Buchanan - angenommen wird, daß die Individuen unter einem „Schleier der Unkenntnis“ ihre Verfassungswahl treffen.

Die Konklusion einer Vertragstheorie ist von den jeweiligen Annahmen abhängig, die sie über den Urzustand trifft. Die konstitutionellen Entscheidungsbedingungen legen die Menge der nach dem konstitutionellen Wohlfahrtskriterium möglichen Wohlstandsverbesserungen fest und damit zugleich die Menge der legitimerweise herbeizuführenden gesellschaftlichen Endzustände. Jene Position, die auf eine Explikation der Ausgangsverteilung verzichtet und für legitim hält, was rationale Individuen - unter welchen Bedingungen auch immer - einstimmig entscheiden, soll im folgenden eine *strikt individualistische oder*

*paretianische Gerechtigkeitsauffassung* genannt werden. Sieht man demgegenüber nur solche Entscheidungen als „gerecht“ an, die bestimmten Anforderungen an die Machtverteilung, die Freiwilligkeit oder die Präferenzen der Individuen im Urzustand erfüllen, so gestaltet man die Vertragssituation nach Maßgabe zusätzlicher, „externer“ Werturteile über das Pareto-Prinzip hinaus. Legitimationsansätze dieser Art werden hier als *schwach individualistische* oder *inhaltliche Gerechtigkeitstheorien* bezeichnet.<sup>1</sup>

Strikt individualistische Gerechtigkeitstheorien basieren allein auf dem Postulat des normativen Individualismus bzw. dem Einstimmigkeitsideal als seiner Entsprechung in der Sphäre der Kollektiventscheidung. Sie vertreten die Extremposition der Verfahrensgerechtigkeit. Was immer Individuen im Urzustand - wie immer dieser aussehen mag - entscheiden würden, gilt als individualistisch „legitimiert“. Anforderungen an die Ausgangssituation werden nicht gestellt. Die Entscheider mögen als Gleiche oder als Ungleiche, freiwillig oder unter Zwang dem Vertrag zustimmen; alle diese Lösungen werden als „gerecht“ betrachtet, sofern sie das Ergebnis einer einstimmigen Kollektiventscheidung sind (oder zumindest sein könnten). Kontraktualistische Rechtfertigungsansätze dieser Art werden auch als *starke* bzw. *echte Vertragstheorien*<sup>2</sup> bezeichnet.

Die Position einer strikt individualistischen Vertragstheorie nimmt James Buchanan in seinem Buch „Die Grenzen der Freiheit“ - keineswegs aber in allen seinen Werken - ein.<sup>3</sup> Er geht darin von einem völlig rechtlosen „Naturzustand“ aus, in dem die Individuen ihre Anstrengungen auf die „eigene“ Produktion, auf den Raub der von anderen Menschen produzierten Produkte und auf die Verteidigung gegen den Raub der „eigenen“ Produkte allozieren. Der gleichgewichtige Zustand der „natürlichen Verteilung“, in dem jedes Individuum - gegeben die Handlungen aller anderen Naturzustandsbewohner - die optimale Mischung aller dieser drei Verhaltensweisen wählt, bildet die Legitimitäts-

---

<sup>1</sup> Die hier verwendete Unterscheidung strikt-individualistischer und inhaltlicher Gerechtigkeitstheorien entspricht weitgehend der Unterscheidung einer individuellen und einer sozialen Dimension von Gerechtigkeitskonzeptionen bei *Schmidt* (1995).

<sup>2</sup> Zur Unterscheidung zwischen „echten“ und „unechten“ Vertragstheorien vgl. *Zintl* (1983), S. 31, zu jener zwischen „starken“ und „schwachen“ Vertragstheorien vgl. *Sugden* (1993), S. 175. Weniger glücklich trennt *Arneson* (1992), S. 201 f., zwischen „egoistischen“ und „nicht-egoistischen“ Vertragstheorien; dies ist insofern irreführend, als auch eine in seinem Sinne „nicht-egoistische“ Vertragstheorie wie die Rawlssche „Theorie der Gerechtigkeit“ die Annahme egoistischer Rationalentscheider verwendet (auch wenn *Rawls* sie so nicht nennen mag); allein durch den „Schleier des Nichtwissens“ (s.u.) werden die Individuen gezwungen so zu handeln, als ob sie keine Egoisten wären.

<sup>3</sup> Vgl. *Buchanan* (1984 [1975]). *Buchanan/Congleton* (1998) konstruieren demgegenüber eine „schwache“ Vertragstheorie, die auf dem Generality-Prinzip als „externem“ Werturteil basiert.



stiftende Ausgangssituation in Buchanans Vertragstheorie. Welche natürliche Verteilung an Verfügungsrechten über Güter sich konkret in der Anarchie herausbilden wird, prognostiziert Buchanan nicht. Die einzige konkrete Situationsannahme, die Buchanan über die anarchische Ausgangsverteilung aufstellt, ist, daß sie paretoinferior sein werde, da *alle* Individuen sich besser stellen würden, wenn sie die bisher für unproduktive Rüstung aufgewendeten Ressourcen fortan in der Produktion verwenden könnten.<sup>4</sup> Allgemein zustimmungsfähig ist in dieser Situation ein Gesellschaftsvertrag, der die unter welchen Umständen auch immer zustande gekommene natürliche Verteilung legalisiert.

Buchanans Ziel ist es, eine Vertragstheorie zu formulieren, welche die Bewertung realer gesellschaftlicher Zustände *allein* von den Bewertungen der von diesen Zuständen betroffenen Individuen abhängig macht. Vom Pareto-Prinzip abgesehen, habe die Analyse daher völlig „wertfrei“ zu sein.<sup>5</sup> Wertfrei ist eine Analyse, wenn sie bestenfalls Aussagen *über* Werthaltungen von Menschen enthält, nicht dagegen Normen selbst.<sup>6</sup> Nach Buchanan haben deshalb alle kontrafaktischen Aussagen über das Bestehen von Machtverhältnissen, Gleichheit oder bestimmte Informationsannahmen im Naturzustand zu unterbleiben, da diese ihrerseits - wie ausgeführt - regelmäßig Ausfluß der persönlichen Gerechtigkeitsnormen des jeweiligen Vertragstheoretikers sind. Unter diesen Umständen wären die Individuen im Urzustand gezwungen, so zu entscheiden, als ob sie die Werturteile, nach denen ihre Entscheidungssituation gestaltet ist, selbst teilen. Was die vertragstheoretische Argumentation nach Buchanan aber ermitteln soll, sind nicht Politikempfehlungen, die den privaten Präferenzen des Theoretikers entsprechen, sondern jene, die im Interesse der betroffenen Individuen liegen.

Buchanans Ziel ist es, seiner Staatsgründungstheorie eine „realistische“ Ausgangssituation zugrunde zu legen.<sup>7</sup> Dieser Zustand ist nicht mit dem realen Status quo zu verwechseln. Auch der „realistische“ Naturzustand Buchanans ist rein fiktiv; er ist eine gedachte Situation völliger Rechtlosigkeit, die der realen

---

<sup>4</sup> Und selbst diese Situationsannahme, daß aus der natürlichen Verteilung heraus stets eine Pareto-Verbesserung möglich sei, relativiert *Buchanan* (1984 [1975]), S. 79, wenn er seinen Anarchiezustand von der „natürlichen Gleichheit“ der Menschen im *Hobbes*-schen Naturzustand durch die Annahme möglicher „Ungleichheit zwischen Personen“ abzugrenzen sucht. Natürliche Verteilungen aber, in denen Menschen sich nicht einmal mehr dadurch gleichen, daß sie alle unter dem Zustand der Rechtlosigkeit leiden, sind selbst paretooptimal, weil durch den Übergang in einen Zustand der Rechtsordnung einige schlechter gestellt würden als in der Anarchie. In einer solchen Situation könnte aber überhaupt kein (paretoverbessernder) Gesellschaftsvertrag geschlossen werden.

<sup>5</sup> Vgl. z. B. *Buchanan/Tullock* (1962), S. vii; *Buchanan* (1968), S. 188.

<sup>6</sup> Vgl. zum Prinzip der Wertfreiheit allgemein *Albert* (1967).

<sup>7</sup> Vgl. *Buchanan* (1984 [1975]), S. 76 ff.



Gesellschaftsordnung, die wir täglich erleben, als fiktive Gegenwelt gegenüber gestellt wird. Da Buchanan jedoch darauf verzichtet, konkrete Situationsannahmen zu treffen, ist der Naturzustand völlig *offen*, d.h., jede denkbare Variante der Rechtlosigkeit soll eine mögliche Ausgangssituation darstellen.

Legt man jedoch der Vertragstheorie einen solchermaßen offenen Urzustand zugrunde, so muß man bereit sein zu akzeptieren, daß legitimitätsstiftende Sozialkontrakte buchstäblich in *jeder* Situation geschlossen werden können. Buchanan schreckt nicht einmal davor zurück, selbst unter Zwang getroffene Entscheidungen als gültige „Quellen der Bewertung“ anzusehen. Ausdrücklich will er sogar einen „Sklavereivertrag“ als „legitimes“ bzw. „gerechtes“ Resultat einstimmiger Entscheidungsfindung akzeptieren, solange dieses Verhältnis auf einer „beiderseitig vorteilhaften“ Übereinkunft beruhe und damit dem Pareto-Kriterium der Nichtschlechterstellung von Beteiligten genüge; selbst die Sklaven hätten ihren Vorteil davon, „für die ‘Starken’ Güter zu produzieren im Austausch gegen ein wenig mehr als die nackte Existenz, die ihnen unter anarchistischen Verhältnissen nicht sicher ist.“<sup>8</sup>

Die einzige erlaubte Situationsannahme in Buchanans starker Vertragstheorie soll jene sein, *daß* die Individuen einstimmig entscheiden; sie beruht auf dem Pareto-Prinzip selbst als dem einzigen verbleibenden Werturteil der Analyse.<sup>9</sup> Der wesentliche Unterschied zwischen der Konsensnorm und jedem anderen Werturteil, nach Maßgabe dessen die konstitutionelle Entscheidungssituation gestaltet werden könnte, liegt in ihrer genuinen *Ergebnisoffenheit*; sie macht das Pareto-Prinzip zum einzigen Werturteil, dessen Einführung in die vertragstheoretische Analyse das Legitimationsergebnis nicht bereits vorwegnimmt. Dies ist jedoch nicht nur ein Vorteil.

Kombiniert mit der ökonomischen Rationalverhaltenshypothese, läßt sich aus einer Vertragstheorie mit einem so geringen normativen Gehalt nämlich nur begründen, *daß* das in der konstitutionellen Situation der Verfassungswahl beschlossene Vertragsergebnis paretooptimal sein werde. Da das Pareto-Kriterium selbst nicht in der Lage ist, zwischen verschiedenen Pareto-Optima zu wählen, kann eine strikt paretianische Vertragstheorie jedoch nicht prognostizieren, *welcher* gesellschaftliche Endzustand jener sein wird, auf den sich die Individuen im Urzustand einigen werden. Das Ziel einer individualistischen

---

<sup>8</sup> Buchanan (1984 [1975]), S. 86; Buchanan selbst räumt dort ein, daß die „Deutung der Sklaverei als Institution etwas gezwungen wirken [mag; C. M.]. Sie dient aber ausschließlich dem Zweck, den hier entwickelten Begriffsrahmen so allgemein wie möglich zu halten.“

<sup>9</sup> Vgl. Zintl (1983), S. 31. Ganz im Sinne Buchanans schreibt etwa auch Tollison (1972), S. 4, über die Wertbasis der Konstitutionenökonomik: „Consensus is the *only* standard by which ‘rightness’ can be discerned under this approach.“ (Hervorhebung von mir; C. M.)

Legitimation realer politischer Handlungsalternativen, mit dem die konstitutionelle Analyse eigentlich angetreten war, kann die starke Vertragstheorie deshalb nicht einlösen.

Buchanans starker Kontraktansatz stößt damit auf das gleiche Problem wie eine strikt paretianische Wohlfahrtsökonomik<sup>10</sup>, soweit diese auf dem Pareto-Kriterium als einziger Wertprämisse basiert: Für eine Formulierung praktischer Politikempfehlungen ist sie untauglich. Das wohlfahrtsökonomische Modell des statischen totalen mikroökonomischen Wettbewerbsgleichgewichts unterstellt eine Dichotomie von nutzenmaximierenden Konsumenten, die - gegeben eine bestimmte Erstausrüstung an Gütern und Faktoren - als Anbieter und Nachfrager auftreten, und Unternehmen, die nach Maßgabe einer Produktionsfunktion auf gewinnmaximierende Weise Güter produzieren. Unter bestimmten Annahmen über die Nutzen- und Produktionsfunktionen (z. B. das Fehlen externer Effekte) existiert ein eindeutiges System von Güter- und Faktorpreisen, von nutzenmaximierenden Abgabe- und Verbrauchsmengen bzw. gewinnmaximierenden Einsatz- und Ausbringungsmengen. Nach dem Ersten Theorem der Wohlfahrtsökonomik<sup>11</sup> ist das wohlfahrtsoptimale Allokations-Preis-Paar paretoeffizient, da erstens die Produktion eines Gutes nicht erhöht werden kann, ohne daß mindestens die Produktion eines anderen Gutes eingeschränkt werden muß, und zweitens der Nutzen eines Haushaltes nicht erhöhbar ist, ohne den mindestens eines anderen Haushaltes zu reduzieren.<sup>12</sup>

Mit der Information, daß das Wohlfahrtsmaximum der betrachteten Gesellschaft ein Pareto-Optimum sein müsse, ist indes wenig gewonnen. Für die praktische Politik interessant wäre nämlich die Frage, *welche* der möglichen paretoeffizienten Situationen schließlich gewählt werden müßte, um die gesellschaftliche Wohlfahrt zu maximieren. Dazu aber kann eine strikt paretianische Wohlfahrtsökonomik nichts aussagen.

Der Grund für diese „Sprachlosigkeit“ der starken Vertragstheorie wie der strikt paretianischen Wohlfahrtsökonomik sei anhand von Abbildung 2.1. verdeutlicht, die ein Zwei-Personen-Nutzendiagramm enthält. Auf der Abszisse

---

<sup>10</sup> Vgl. etwa den Überblick bei *Johansson* (1991); *Schumann* (1994).

<sup>11</sup> Vgl. *Varian* (1994 [1992]), S. 326 ff.

<sup>12</sup> Drei Marginalbedingungen müssen hierfür erfüllt sein (vgl. grundlegend *Samuelson* 1947, Kapitel 8): Erstens müssen die Grenzzraten der Faktorsubstitution für sämtliche Güter identisch sein (Produktionsoptimum); zweitens verlangt die Herbeiführung eines wohlfahrtsmaximalen Zustandes den Ausgleich der Grenzzraten der Gütersubstitution für sämtliche Individuen (Tauschoptimum); und drittens setzt ein Wohlfahrtsoptimum voraus, daß die Grenzrate der Transformation der Grenzrate der Gütersubstitution gleich ist (simultanes Produktions- und Tauschoptimum). Alle drei dieser wohlfahrtsökonomischen Marginalbedingungen sind lediglich logische Implikationen aus den Prämissen des Modells der vollkommenen Konkurrenz; vgl. *Bohnen* (1964), S. 87.

und der Ordinate ist jeweils der Nutzen eines Individuums  $A$  und  $B$ ,  $U_A$  und  $U_B$ , abgetragen. Die Wohlstandsgrenze  $WG$  gibt die möglichen maximalen Nutzenverteilungen der Individuen  $A$  und  $B$  an. In der Wohlfahrtsökonomik ergibt sie sich als die Umhüllende der Nutzenmöglichkeitskurven der betroffenen Individuen. Im Abszissenschnittpunkt  $U_A^*$  von  $WG$  erzielt  $A$  seinen maximalen,  $B$  überhaupt keinen Nutzen; das Umgekehrte gilt im Ordinatenschnittpunkt  $U_B^*$ . Verteilungen unterhalb von  $WG$  sind paretoinferior, weil, von ihnen ausgehend, mindestens ein Individuum besser gestellt werden könnte, ohne das andere schlechter zu stellen; Verteilungen oberhalb von  $WG$  sind demgegenüber bei gegebener Technologie nicht erreichbar. Alle Verteilungen indes, die auf der Wohlstandsgrenze  $WG$  liegen, sind paretooptimal; markieren diese die Ausgangsverteilung einer einstimmigen Entscheidung, so ist es nicht möglich, ein Individuum besser zu stellen, ohne das andere schlechter zu stellen.

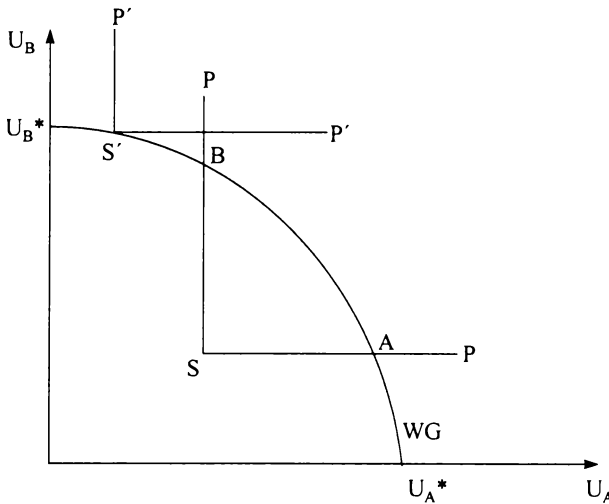


Abbildung 2.1.: Nutzendiagramm im Zwei-Personen-Fall

Die „Sprachlosigkeit“ einer strikt paretianischen Vertragstheorie oder Wohlfahrtsökonomik ergibt sich unmittelbar aus dem Pareto-Prinzip selbst, das nur eine (unvollständige) „*strikte partielle Ordnung*“ gesellschaftlicher Zustände definiert.<sup>13</sup> Gegeben eine bestimmte Ausgangsverteilung der Wohlstandsniveaus

<sup>13</sup> Vgl. Sen (1970), S. 29. Die in § 1 B. I. definierte Pareto-Relation  $\bar{P}$  ist eine „*strikte partielle Ordnung*“ über die Menge  $S$  aller sozialen Zustände, weil sie *transitiv* (für alle  $x, y, z \in S$ :  $(x \bar{P} y \wedge y \bar{P} z) \rightarrow x \bar{P} z$ ) und *asymmetrisch* (für alle  $x, y \in S$ :  $x \bar{P} y \rightarrow \neg(y \bar{P} x)$ ) ist; Asymmetrie impliziert *Anti-Symmetrie* (für alle  $x, y \in S$ :

der betroffenen Individuen, zeigt es auf, wie einige Beteiligte ihre Wohlfahrtsposition verbessern können, ohne daß die jeweils anderen Kollektivmitglieder sich verschlechtern. Beschreibt etwa Punkt  $S$  in Abbildung 2.1. die anfängliche Nutzenverteilung in der Anarchie oder im Markt, so ist es möglich, die Individuen durch Herbeiführen von Punkten innerhalb der Pareto-Grenzen  $PP$  besser zu stellen. Ist indes einmal ein Pareto-Optimum auf dem Kurvenabschnitt  $AB$  erreicht, so kann das Einstimmigkeitskriterium selbst keine Handlungsempfehlung geben, wie diese Situation möglicherweise wohlfahrtssteigernd verlassen werden kann.

Damit kann es mit Hilfe des Pareto-Kriteriums nur im Ausnahmefall gelingen, praktische Politikempfehlungen zu formulieren, die das Prädikat „individualistisch legitimiert“ verdienen. Denn eine strikt paretianische Gerechtigkeitstheorie kommt nur dann zu einer eindeutigen Aussage, wenn es, verglichen mit der Ausgangssituation, genau ein einziges Pareto-Optimum gibt, das herbeizuführen im Interesse aller Beteiligten liegt. Immer dann aber, wenn ein intersubjektiver Nutzenvergleich - d.h. der Vergleich von paretoeffizienten Situationen - erforderlich wäre, ist die starke Vertragstheorie zum Schweigen verurteilt. Politik aber bedeutet sehr häufig, *zwischen Pareto-Optima zu wählen*.<sup>14</sup>

Ein Vergleich von Pareto-Optima könnte etwa erforderlich sein, wenn die betrachtete Entscheidungssituation selbst paretooptimal ist, in Abbildung 2.1. also auf der Wohlstandsgrenze  $WG$  läge. Mit dem Einstimmigkeitsprinzip allein läßt sich dann eine Veränderung dieser Ausgangssituation nicht legitimieren. Eine Umverteilungspolitik, mit Hilfe derer einigen Individuen Einkommen oder Vermögen entzogen werden, um sie notleidenden Individuen zuzuweisen, wäre damit nur im Ausnahmefall zu legitimieren. Selbst dann, wenn die Ausgangsverteilung so ungleich wäre wie im Punkt  $S'$ , wäre eine Redistribution nicht zu rechtfertigen, da die Menge der von dort aus möglichen Pareto-Verbesserungen, die durch die Pareto-Grenze  $P'P'$  angegeben wird, leer ist.

Wie groß auch immer das menschliche Elend im einzelnen sein mag: wenn für die Besserstellung des Notleidenden eine andere Person schlechter gestellt werden muß, ist nach dem Pareto-Kriterium daran nicht legitim zu rütteln; „im

---

$(x \bar{P} y \wedge y \bar{P} x) \rightarrow x = y$ ). Die Relation  $\bar{P}$  ist jedoch keine „starke Ordnung“ über  $S$ , da ihr die Eigenschaft der *Vollständigkeit* (für alle  $x, y \in S$ :  $(x \neq y) \rightarrow (x \bar{P} y \vee y \bar{P} x)$ ) fehlt. Vgl. zu diesen Eigenschaften Sen (1970), S. 8 f.

<sup>14</sup> Nach Hicks (1939), S. 706, gilt dies sogar für *alle* wirtschaftspolitischen Maßnahmen: „Under private enterprise, any ordinary change in economic policy involves a change in the price-system, and any change in prices benefits those on one side of the market, and damages those on the other. Thus ... it *always* inflicts a loss of some sort upon some people.“ (Hervorhebung von mir; C. M.)

Extremfall ist es sogar möglich, daß man paretooptimal verhungert.<sup>15</sup> Umverteilung ist nach dem Pareto-Kriterium nur dann zu rechtfertigen, wenn die Geber entweder altruistisch sind oder durch die Redistribution eigennützige Zwecke zu verfolgen in der Lage sind.<sup>16</sup> Mit der Auszeichnung eines Punktes als paretooptimal ist somit im allgemeinen nicht viel gewonnen, weil diese Eigenschaft eine beliebige gesellschaftliche Situation sehr oft kennzeichnen dürfte.

Aber selbst dann, wenn der jeweils zugrunde gelegte Status quo - wie die natürliche Verteilung in Buchanans starker Vertragstheorie - nicht paretooptimal sein sollte, kann ein Vergleich zwischen Pareto-Optima nötig sein. Dies ist dann der Fall, wenn es mehrere Möglichkeiten gibt, die paretoinferiore Entscheidungssituation zu verlassen. So kann etwa, ausgehend von Punkt *S* in Abbildung 2.1., keine Aussage darüber getroffen werden, welches der auf dem Kurvenabschnitt *AB* liegenden, erreichbaren Pareto-Optima herbeigeführt werden soll. In einer solchen Situation wie Punkt *S* gibt es daher, wie Arrow pointiert formuliert, nach Maßgabe des Pareto-Kriteriums „no way of knowing which of these other alternatives to change to, so that community will find itself in much the same position as Buridan's ass.“<sup>17</sup>

Buchanan hält die Unfähigkeit seiner starken Vertragstheorie, dieses Optimum-Auswahlproblem zu lösen, für deren eigentlichen Vorzug:

„... the very essence of contract is the *nonspecification* of outcome by external observers. Traders trade; agreement is reached, agreement that is presumed to be mutually beneficial to the parties. Conceptually at least, there is a subinfinity of possible equilibria along some generalized contract locus.“<sup>18</sup>

Von Kritikern wurde Buchanans in „Die Grenzen der Freiheit“ vorgelegter Staatsgründungstheorie wegen dieser Position mit Recht „vertragstheoretischer Nihilismus“<sup>19</sup> vorgeworfen: Da er die anarchische Ausgangssituation in keiner Weise expliziert - bei Buchanan ist nur klar, daß sich *irgendeine* „natürliche Verteilung“ einstellen wird -, verbietet sie kein mögliches Pareto-Optimum als Vertragsergebnis. Unter entsprechenden Umständen läßt Buchanans in „Die Grenzen der Freiheit“ formulierter starker Staatsgründungsansatz

---

<sup>15</sup> Nutzinger (1975), S. 8. Vgl. auch die Kritik am Pareto-Prinzip bei Albert (1998), S. 11 f.

<sup>16</sup> Vgl. zu dieser sog. paretooptimalen Umverteilung Hochman/Rodgers (1969) und Brennan (1973).

<sup>17</sup> Arrow (1963), S. 37.

<sup>18</sup> Buchanan (1972), S. 125 (Hervorhebung im Original).

<sup>19</sup> Zintl (1983), S. 90. Vgl. in diesem Sinne auch Barry (1984), S. 591; Zintl (1986), S. 122 und S. 125.

„... ein sozialistisches Gesellschaftssystem, eine Feudalordnung oder eine Sklavenhaltergesellschaft ebenso gut möglich erscheinen ... wie eine faschistische Militärdiktatur, den liberalen Nachtwächterstaat oder die wohlfahrtsstaatliche Demokratie“<sup>20</sup>.

Damit scheitert aber zugleich Buchanans in „Die Grenzen der Freiheit“ unternommener „Versuch zu erklären, wie das ‘Recht’, ‘die Eigentumsrechte’ und ‘die Verhaltensregeln’ aus nichtidealistischem, vom Eigeninteresse bestimmten Verhalten der Menschen abgeleitet werden können, ohne vorauszusetzen, daß in einer Art Naturzustand Gleichheit bestünde“<sup>21</sup>. Aus dem Explanans des Buchananschen vertragstheoretischen Arguments folgt die Errichtung eines Rechtsstaats als Einigungsergebnis der Individuen auf der konstitutionellen Ebene nur dann notwendig, wenn diese - nach Maßgabe eines externen Werturteils - um entsprechende Annahmen über die Entscheidungssituation im Urzustand konkretisiert werden.<sup>22</sup> So, wie er sich in „Die Grenzen der Freiheit“ darstellt, liefert der Buchanansche Ansatz daher nicht nur eine fiktive, sondern auch eine lediglich *partielle Erklärung*: Neben der empirischen Adäquatheitsbedingung, gegen welche alle hypothetischen Vertragstheorien verstoßen, verletzt Buchanans Staatsgründungstheorie, soweit sie den Rechtsstaat als konkretes Legitimationsergebnis des Gedankenexperiments dartun möchte, auch die logischen Anforderungen Hempels und Oppenheims an wissenschaftliche Erklärungs- bzw. Prognoseargumente. Aufgrund der fehlenden Spezifizierung der Situationsannahmen im Naturzustand ist das Explanans zu schwach, um die Explanandum-Aussage - die Einigung auf einen Rechtsschutzstaat - zu behaupten.

---

<sup>20</sup> Koller (1987), S. 242. Für eine ähnliche Kritik vgl. bereits Goldberg (1974), S. 572.

<sup>21</sup> Buchanan (1984 [1975]), S. 78.

<sup>22</sup> Vgl. Goldberg (1974).

### § 3 Die schwache Vertragstheorie (I.): Die Situationsannahmen

Das Scheitern der Buchanan'schen Staatsgründungstheorie, den Rechtsstaat hypothetisch zu begründen, kann, bei näherer Betrachtung, kaum überraschen: Aus einer offenen Ausgangssituation folgen offene Entscheidungsergebnisse; aus einer konkreten Ausgangssituation folgen konkrete Entscheidungsergebnisse. Die beklagte Unbestimmtheit der starken Vertragstheorie ist nur die Kehrseite der gerühmten Ergebnisoffenheit des Pareto-Prinzips. Will die Theorie des hypothetischen Gesellschaftsvertrags Empfehlungen für die praktische Politik geben, so wird sie nicht umhin kommen, den Urzustand in irgendeiner Weise zu konkretisieren, um die Menge möglicher Entscheidungsergebnisse einzuschränken. In der Rawlsschen „Theorie der Gerechtigkeit“, aber wiederum auch in Teilen von Buchanan's Werk, erfolgt diese Konkretisierung durch die zentrale Situationsannahme eines „Schleiers der Unkenntnis“.

Bei Rawls wie bei Buchanan besteht das mit einem Schleier der Unkenntnis arbeitende vertragstheoretische Argument aus zwei (üblicherweise nicht getrennten) Argumentationsstufen: In einem *ersten Schritt*, der im Mittelpunkt der Analyse von § 3 steht, wird dargelegt, daß der Schleier der Unkenntnis die rationalen Individuen im Urzustand dazu veranlasse, bei ihrer Regelentscheidung ihr Verhalten zu „universalisieren“. Die Informationsbeschränkungen auf der konstitutionellen Entscheidungsebene zwingen die Beteiligten dazu, sich rationalerweise in die Rolle der übrigen Individuen zu versetzen. Die Rawlssche „Theorie der Gerechtigkeit“ (Abschnitt A.) postuliert als eine inhaltliche Gerechtigkeitsnorm, daß Personen, die über die grundlegenden Institutionen einer Gesellschaft entscheiden, sich in die Lage der anderen Individuen versetzen und prüfen sollen, ob sie auch von deren Standpunkt aus eine zur Diskussion stehende Regel befürworten könnten. Der Rawlssche Urzustand mit der Informationsbeschränkung durch den Schleier des Nichtwissens expliziert in dieser ersten Stufe lediglich Entscheidungsbedingungen, welche fiktive Rationalentscheider dazu bringen, sich so zu entscheiden, als teilten sie die Universalisierungsnorm, die der Gestaltung ihrer Entscheidungssituation zugrunde liegt. Rawls' „Theorie der Gerechtigkeit“ ist damit eine schwache Vertragstheorie. In der Literatur wird der fiktive Vertragsschluß unter einem Schleier des Nichtwissens häufig allein im Sinne dieser ersten Stufe - als Metapher für ein moralisches Prinzip - verwendet.<sup>1</sup>



Auch der Buchanan'sche Schleier der Unsicherheit führt zu einem vergleichbaren gedanklichen Rollentausch; im Unterschied zur Rawlsianischen Theorie wird dieser jedoch nicht mit einer Norm begründet, sondern als eine empirische Tatsache behauptet, die typischerweise bei Verfassungsentscheidungen auftritt. Dem Anspruch nach ist also auch dieser Teil der Buchanan'schen Theorie eine starke Vertragstheorie, die ohne jede „moralische Bereinigung“ des Urzustandes auskommt. Buchanan glaubt mit dieser in „The Calculus of Consent“ entwickelten Konstruktion ein Dilemma vermeiden zu können, vor dem bereits die moderne Wohlfahrtsökonomik stand: Argumentiert die Legitimationstheorie rein „paretianisch“, so muß sie, wie gezeigt, die ihr zugewiesene Aufgabe, Empfehlungen für die praktische Politik zu formulieren, in den meisten Fällen verfehlen. Versucht sie hingegen konkrete Politikempfehlungen zu geben, so muß sie zusätzliche, inhaltliche Gerechtigkeitsnormen über die vertragstheoretische Konsensnorm hinaus einführen, was jedoch die faktische Aufgabe des Pareto-Prinzips und damit des vertragstheoretischen Legitimationskriteriums selbst bedeutet.

In Abschnitt B. werde ich argumentieren, daß die Buchanan'sche Behauptung, dieses Dilemma durch den konstitutionellen Ansatz auf der Basis eines empirischen Schleiers der Unsicherheit umgehen zu können, falsch ist. Ich werde die Vermutung äußern, daß die Behauptung, der Schleier der Unsicherheit sei ein empirisches Phänomen, nur deshalb in die Buchanan'sche Vertragstheorie eingeführt wurde, um Ergebnisse herzuleiten, die mit der Universalisierungsnorm in Einklang zu bringen sind. Die Buchanan'sche Vertragstheorie basiert dann - uneingestanden - auf dem Universalisierungspostulat als inhaltlicher Gerechtigkeitsnorm; auch sie ist eine schwache Vertragstheorie.

Der *zweite Schritt* des vertragstheoretischen Arguments bei Rawls und Buchanan besteht darin, aus den Modellprämissen des Urzustandes eine konkrete Gerechtigkeitsvorstellung - üblicherweise repräsentiert in Form einer sozialen Wohlfahrtsfunktion - zu deduzieren. Wer die ethische Universalisierungsnorm teilt und die Behauptung akzeptiert, der fiktive Vertragsschluß unter dem Schleier der Unkenntnis sei eine zutreffende Metapher für in diesem Sinne normanaloges Verhalten, der muß - so die Folgerung - auch wollen, daß eine Gesellschaft nach Maßgabe einer bestimmten sozialen Wohlfahrtsfunktion reguliert wird. Dieser Konklusion des kontraktualistischen Arguments werde ich mich in § 4 widmen.

---

<sup>1</sup> Vgl. allein in jüngster Zeit z. B. den Entwurf eines Konzepts sozialpolitischer Gefahrvorsorge durch *Wagner* (1998) oder die Analyse meritatorischer Güter bei *Tietzel/Müller* (1998).



### A. Rawls' „Theorie der Gerechtigkeit“ und der „Schleier des Nichtwissens“

Lange vor der Veröffentlichung der Buchananschen „Grenzen der Freiheit“ hat Rawls das Problem, daß eine allein auf dem Konsensprinzip aufbauende Vertragstheorie keine konkreten Entscheidungsergebnisse vorhersagen kann, klar erkannt. Ziel der Rawlsschen „Theorie der Gerechtigkeit“ ist es, eine ganz konkrete Vorstellung von Gerechtigkeit - die sog. „Maximin-Lösung des Problems der sozialen Gerechtigkeit“<sup>2</sup> - dadurch zu begründen, daß er zeigt, wie sie sich als Entscheidung rationaler Individuen in einer wohldefinierten, als gerecht empfundenen Vertragssituation ergibt. Die betrachteten konstitutionellen Entscheidungsträger setzt er dazu - seit der ersten Vorstellung seiner Theorie im Jahr 1958 - immer neuen Entscheidungsbedingungen aus; auch die Veröffentlichung seines Hauptwerkes im Jahr 1971 brachte diesen Revisions- und Fortbildungsprozeß seiner Theorie nicht zum Stillstand<sup>3</sup>, wie die erst 1993 erschienene partielle Neufassung der Gerechtigkeitstheorie in dem Buch „Political Liberalism“<sup>4</sup> deutlich macht. Die wichtigste Modifikation, die Rawls im Laufe der Jahre in seine Theorie einfügte, ist die drastische Informationsbeschränkung auf der konstitutionellen Ebene durch die berühmte Annahme eines „Schleiers des Nichtwissens“, die erst 1967 Eingang in sein Vertragsargument fand.<sup>5</sup> Alle diese Situationsannahmen über den Urzustand sollen, wie Rawls ausdrücklich vermerkt, dazu dienen, das Problem der Unbestimmtheit und Ergebnisoffenheit einer starken, allein auf der Konsensnorm aufbauenden Vertragstheorie zu vermeiden.<sup>6</sup>

---

<sup>2</sup> Rawls (1994 [1971]), S. 177. Eine ausführlichere Diskussion des Inhalts dieser Gerechtigkeitseinstellung erfolgt in § 4.

<sup>3</sup> Einen Überblick über die bis dahin vorgenommenen Änderungen gibt Wolff (1977), Part Two.

<sup>4</sup> Vgl. Rawls (1993).

<sup>5</sup> Vgl. erstmals Rawls (1967), S. 60 ff., womit Rawls auf vielfach geäußerte Kritik an früheren Versionen seiner Theorie (v.a. in Rawls, 1958) begegnete. Für eine Kritik dieses frühen Rawls-Modells vgl. z. B. Wolff (1977), S. 35-56.

<sup>6</sup> So schreibt Rawls (1994 [1971]), S. 164: „Die Beschränkungen der Einzelkenntnisse im Urzustand sind ... von grundlegender Bedeutung. Ohne sie könnte man überhaupt keine *bestimmte* Theorie der Gerechtigkeit aufstellen. Man müßte sich mit der unbestimmten Formel zufrieden geben, die Gerechtigkeit sei das, worauf man sich einigen würde, wobei man nur wenig oder gar nichts über den Inhalt der Vereinbarung sagen könnte. ... Der Schleier des Nichtwissens ermöglicht die einstimmige Annahme einer *bestimmten* Gerechtigkeitseinstellung. Ohne diese Beschränkungen des Wissens wäre das Verhandlungsproblem im Urzustand hoffnungslos verwickelt.“ (Hervorhebungen von mir; C. M.)

Ziel der Rawlsschen Theorie - so wie sie sich heute darstellt - ist die „Begründung“ einer Gerechtigkeitskonzeption, durch die sich die optimale „Grundstruktur“ einer Gesellschaft - d.h. eine politische Verfassung, die Wirtschaftsordnung und andere grundlegende Einrichtungen - ermitteln läßt.<sup>7</sup> In jüngerer Zeit geht Rawls bei diesem Vorhaben ausdrücklich vom Faktum eines weltanschaulichen Pluralismus in der modernen Gesellschaft aus, die durch die verschiedenen religiösen, philosophischen und moralischen Lehren tief gespalten ist. Seine Gerechtigkeitstheorie will Rawls daher nicht (mehr) als eine „allgemeine moralische Konzeption“ verstanden wissen, da eine solche Doktrin notwendig mit konkurrierenden Anschauungen kollidieren müßte.<sup>8</sup> Chancen, eine stabile Grundlage einer gegebenen Gesellschaft zu bieten, hat demgegenüber nach Rawls nur eine *politische* Gerechtigkeitskonzeption, welche die Berufung auf ethische Positionen so weit wie möglich vermeidet und deshalb mit einem breiten Spektrum von Weltanschauungen vereinbar ist. In seinen neueren Veröffentlichungen fordert Rawls daher, daß die gesuchte Gerechtigkeitskonzeption dazu geeignet sein müsse, Gegenstand eines *übergreifenden* („*overlapping*“) Konsenses zu sein, also in der Schnittmenge der verschiedenen moralischen, religiösen und philosophischen Positionen zu liegen.<sup>9</sup>

Der Gegenstand von Gerechtigkeitskonzeptionen ist nach Rawls die Regelung des Problems der gerechten Verteilung sog. gesellschaftlicher Grundgüter in einer wohlgeordneten Gesellschaft, worunter er eine durch eine öffentliche Gerechtigkeitskonzeption regulierte Gesellschaft von Individuen versteht, die sich gegenseitig als freie und gleiche moralische Personen respektieren.<sup>10</sup> Grundgüter sind solche Güter, die alle Gesellschaftsmitglieder annahmegemäß in gleicher Weise erstreben. Im einzelnen handelt es sich dabei um (1) Grundfreiheiten wie die Gedanken- oder Gewissensfreiheit, (2) die Freizügigkeit und die Freiheit der Berufswahl, (3) die mit öffentlichen Ämtern verbundenen Vorteile, (4) Einkommen und Besitz und (5) die sozialen Grundlagen der Selbstachtung.<sup>11</sup> Von natürlichen Gütern wie etwa Gesundheit und Geisteskraft, die in unserer Welt häufig von Geburt an ungleich verteilt sind, unterscheiden sich

<sup>7</sup> Zu Rawls' Beschränkung auf die Grundstruktur vgl. *Rawls* (1992 [1977]); siehe auch *Beauchamp* (1980).

<sup>8</sup> Vgl. *Rawls* (1992 [1985]), S. 257; *ders.* (1993), S. 175.

<sup>9</sup> Vgl. *Rawls* (1992 [1987]); *ders.* (1992 [1989]); *ders.* (1993), S. 133-172. Zur Unterscheidung umfassender moralischer und politischer Gerechtigkeitskonzeptionen vgl. *Pogge* (1994), S. 39-43, sowie kritisch *Ricken* (1997).

<sup>10</sup> Der Begriff der wohlgeordneten Gesellschaft, der schon in *Rawls* (1994 [1971]), S. 21, verwendet wurde, findet sich erstmals in *Rawls* (1992 [1980]), S. 88 ff., eingehend expliziert.

<sup>11</sup> Vgl. *Rawls* (1982), S. 162 und S. 165 f.; *ders.* (1992 [1980]), S. 94 f.; *ders.* (1993), S. 181 f. *Rawls* hat seine Grundgüterkonzeption im Laufe der Zeit erheblich geändert; vgl. hierzu *Hinsch* (1992), S. 36-44.

diese Grundgüter in der Rawlsschen Konzeption dadurch, daß ihre Bereitstellung, direkt oder indirekt, durch die gesellschaftliche Grundstruktur erfolgt, auf die allein sich die Rawlssche Theorie beziehen soll.

Die Grundgüter, die in „Eine Theorie der Gerechtigkeit“ noch als für die Verwirklichung der konkreten Lebenspläne der Gesellschaftsmitglieder notwendig betrachtet wurden, sieht Rawls heute als Voraussetzung dafür an, daß Individuen moralische Personen sein können; der Begriff der moralischen Person determiniert, was im Einzelfall ein Grundgut ist.<sup>12</sup> Eine moralische Person hat nach Rawls drei Merkmale<sup>13</sup>: *Erstens* verfügt eine moralische Person über die Anlage zu einem wirksamen Gerechtigkeitssinn, womit die Fähigkeit gemeint ist, Gerechtigkeitsgrundsätze zu verstehen, sie anzuwenden und aus ihnen heraus zu handeln. *Zweitens* besitzt sie die Befähigung, eine Konzeption des Guten - eine Rangordnung letzter Ziele - auszubilden, zu revidieren und rational zu verfolgen. Und neben diesen beiden Merkmalen, die auch als die beiden moralischen Vermögen bezeichnet werden, hat eine moralische Person annahmegemäß *drittens* eine persönliche Konzeption des Guten. Diesen drei Merkmalen entsprechen drei übergeordnete Interessen moralischer Personen: zum einen die beiden höchstrangigen Interessen, die beiden moralischen Vermögen zu verwirklichen und auszuüben, zum anderen das höher- (jedoch nicht höchst-) rangige Interesse daran, die persönliche Konzeption des Guten, welchen konkreten Inhalt sie auch haben mag, zu schützen und zu fördern. Der Inhalt des dritten Interesses soll abstrakt und offen bleiben, um die allgemeine Akzeptierbarkeit der Rawlsschen Theorie nicht zu gefährden.

Gerecht sind nach Rawls solche gesellschaftlichen Verteilungsprinzipien - sog. Gerechtigkeitsprinzipien - die *rational* autonome Bürger in einem fiktiven, als gerecht empfundenen Urzustand auswählen würden. Die - moralisch neutrale - rationale Autonomie der Bürger in der konstitutionellen Entscheidungssituation entspricht dem (Zweck-)Rationalitätsbegriff der neoklassischen Wirtschaftstheorie, nach dem zu gegebenen Zielen die je wirksamsten Mittel eingesetzt werden.<sup>14</sup> Entsprechend beschreibt Rawls die Akteure im Urzustand als

<sup>12</sup> Vgl. ursprünglich *Rawls* (1994 [1971]), S. 112 ff.; die Änderung findet sich in *Rawls* (1992 [1980]), S. 96.

<sup>13</sup> Vgl. *Rawls* (1992 [1980]), S. 93 f.; *ders.* (1992 [1985]), S. 267 f. und S. 268 f.; *ders.* (1993), S. 18 f.

<sup>14</sup> So ausdrücklich *Rawls* (1994 [1971]), S. 31 und S. 166-169, wo in der deutschen Übersetzung unglücklicherweise durchgehend von „Vernünftigkeit“ statt von „Rationalität“ die Rede ist, obwohl „das Vernünftige“ und „das Rationale“ bei *Rawls* unterschiedliche Bedeutungen haben; vgl. auch *Rawls* (1992 [1980]), S. 88.

rational und gegenseitig desinteressiert<sup>15</sup>, als Menschen, die „kein Interesse an den Interessen anderer nehmen.“<sup>16</sup>

*Vollständige* - d.h. moralisch gehaltvolle - Autonomie können nach Rawls hingegen nur die Bürger einer wohlgeordneten Gesellschaft besitzen, welche die obersten Gerechtigkeitsgrundsätze, auf die man sich einigen würde, gutheißen und nach ihnen handeln. Im Urzustand sollen die wesentlichen Merkmale der vollständigen Autonomie realisiert sein, wozu insbesondere die fairen Bedingungen der Kooperation gehören. Fair sind diese Bedingungen der Zusammenarbeit, wenn sie der Forderung nach Gegenseitigkeit entsprechen, d.h. alle Kooperationsteilnehmer müssen aus der Zusammenarbeit einen Vorteil ziehen und die gemeinsamen Belastungen in angemessener Form teilen. Rawls nennt diesen Reziprozitätsaspekt sozialer Kooperation auch *das Vernünftige* - im Unterschied zum *Rationalen* der Entscheidungsträger im Urzustand.<sup>17</sup>

Den Urzustand expliziert Rawls so, daß er zum einen den Begriff freier und gleicher moralischer Personen repräsentiert und zum anderen das Vernünftige dem Rationalen vorgeordnet ist.<sup>18</sup> Dies erreicht er dadurch, daß er die fiktiven Rationalakteure zwar als Repräsentanten bestimmter sozialer Positionen - die des Reichen oder Armen, des Schwarzen oder Weißen, einer Frau oder eines Mannes - beschreibt, sie aber unter einem „*Schleier des Nichtwissens*“ entscheiden läßt, der sie in Unkenntnis all jener Partialinteressen versetzt, die aus der von ihnen repräsentierten sozialen Position resultieren können. Im Urzustand sind den Parteien

<sup>15</sup> Vgl. Rawls (1994 [1971]), S. 168 und S. 286.

<sup>16</sup> Rawls (1994 [1971]), S. 30.

<sup>17</sup> Zum Begriff des Vernünftigen im Unterschied zu dem des Rationalen vgl. Rawls (1992 [1980]), S. 88 und S. 97 ff; ders. (1993), S. 48-54 und S. 72-81. Siehe auch Kley (1989), S. 18; Ricken (1997), S. 423.

<sup>18</sup> Vgl. Rawls (1992 [1980]), S. 89; zur Priorität des Vernünftigen über das Rationale vgl. Rawls (1992 [1980]), S. 100 und S. 103. - Wegen dieses Vorrangs des Vernünftigen vor dem Rationalen hat Rawls es in einer jüngeren Veröffentlichung ausdrücklich bedauert, den Urzustand als Entscheidung ökonomischer *Homines oeconomici* expliziert zu haben. So war es nach Rawls (1992 [1985]), S. 273 f., Fußnote 20, „ein (sehr irreführender) Fehler, in *Eine Theorie der Gerechtigkeit* eine Gerechtigkeitstheorie als Teil einer Theorie rationaler Entscheidung darzustellen ... Ich hätte sagen sollen, daß *Gerechtigkeit als Fairneß* eine Theorie rationaler Entscheidung, die den Bedingungen des Vernünftigen unterliegt, dazu benutzt, die Überlegungen der Parteien als freier und gleicher Personen zu beschreiben, und all das innerhalb einer Gerechtigkeitskonzeption, die natürlich eine moralische Konzeption ist. Ich habe nie daran gedacht, zu versuchen, den Gehalt der Gerechtigkeit innerhalb eines Rahmens herzuleiten, der sich auf den Gedanken des Rationalen als einziger normativer Idee beschränkt. Dies wäre unvereinbar mit jeglicher Form kantischen Denkens.“ (Hervorhebungen im Original)

„... bestimmte Arten von Einzeltatsachen unbekannt ... Vor allem kennt niemand seinen Platz in der Gesellschaft, seine Klasse oder seinen Status; ebenso wenig seine natürlichen Gaben, seine Intelligenz, Körperkraft usw. Ferner kennt niemand seine Vorstellung vom Guten, die Einzelheiten seines vernünftigen Lebensplanes, ja nicht einmal die Besonderheiten seiner Psyche wie seine Einstellung zum Risiko oder seine Neigung zu Optimismus oder Pessimismus. Darüber hinaus setze ich noch voraus, daß die Parteien die besonderen Verhältnisse in ihrer eigenen Gesellschaft nicht kennen, d.h. ihre wirtschaftliche und politische Lage, den Entwicklungsstand ihrer Zivilisation und Kultur. Die Menschen im Urzustand wissen auch nicht, zu welcher Generation sie gehören.“<sup>19</sup>

Die Annahme des Schleiers des Nichtwissens hat im vertragstheoretischen Argument von Rawls zwei zentrale Aufgaben zu erfüllen, die man die Fairneß- und die Konsensfunktion des Unwissenheitsschleiers nennen könnte. Die *Fairneßfunktion* der Informationsbeschränkungen besteht darin, faire Bedingungen für die Wahl von Gerechtigkeitsgrundsätzen zu schaffen. In bezug auf die Auswahl distributiver Grundsätze befinden sich die Individuen in völlig symmetrischen Positionen. Vorrechte und Privilegien bestehen nicht. Realiter bestehende asymmetrische Verteilungen von Macht und Einflußmöglichkeiten bleiben für die Wahl im Urzustand ohne Konsequenzen.<sup>20</sup> Auch sind die Beteiligten im Urzustand aller ihrer Partialinteressen beraubt, so daß selbst egoistischste Individuen keinen Platz für Diskriminierung finden: Weiß der einzelne nicht, ob er Mann oder Frau ist, wird er nicht für Regeln stimmen, die eines der beiden Geschlechter diskriminieren. Weiß er nicht, ob er Generaldirektor oder Müllmann ist, wird er gegen Klassenschranken votieren. Kennt er nicht einmal die Generation, der er angehört, wird er sich gegen Umweltzerstörung und für eine nachhaltige Ressourcennutzung einsetzen. Entscheidungen, die unter diesen extremen Informationsbedingungen zu treffen sind, sollen damit stets als unparteiisch angesehen werden können:

„Unsere Stellung in der Gesellschaft unter diesem Blickpunkt zu sehen heißt ... sie sub specie aeternitatis sehen: Es bedeutet, daß die Situation des Menschen nicht nur unter allen gesellschaftlichen Gesichtspunkten, sondern von allen Zeiten her gesehen wird.“<sup>21</sup>

Der Schleier des Nichtwissens bewirkt nach Rawls, „daß Personen ausschließlich als moralische Personen repräsentiert werden und nicht als Personen, die durch die Zufälligkeiten ihrer gesellschaftlichen Stellung, die Verteilung natürlicher Fähigkeiten oder durch das Schicksal und historische Zufälle im Laufe ihres Lebens begünstigt oder benachteiligt wurden.“<sup>22</sup> Zwar strebt jeder rationale Entscheidungsträger im Urzustand danach, seine eigenen Inter-

<sup>19</sup> Rawls (1994 [1971]), S. 160.

<sup>20</sup> So auch Hinsch (1997a), S. 60 f.

<sup>21</sup> Rawls (1994 [1971]), S. 637.

<sup>22</sup> Rawls (1992 [1980]), S. 99.

essen, so gut er kann, zu fördern. Gleichwohl „verschleiern“ die Informationsbeschränkungen alles diskriminierungsrelevante Wissen. Die Bedingungen des Urzustandes zwingen die Rationalentscheider dazu, sich allein von ihren höchstrangigen Interessen an der Ausbildung und Ausübung ihrer beiden moralischen Vermögen - ihrem Gerechtigkeitssinn und ihrer Befähigung zu einer Konzeption des Guten - leiten zu lassen, die allen Individuen als moralischen Personen gemein sind. Die Gerechtigkeitsgrundsätze, die sie unter dem Schleier der Unwissenheit wählen, sind solche, die ihnen die für die Entwicklung und Ausübung ihrer beiden moralischen Vermögen notwendigen institutionellen und materiellen Bedingungen gewähren.

Neben dieser Fairneß- hat der Schleier des Nichtwissens auch eine *Konsensfunktion*. Die Informationsbeschränkungen im Urzustand versetzen alle Rationalentscheider in *dieselbe* Entscheidungssituation. Da ihnen unter diesen Bedingungen allein ihre höherrangigen Interessen als Entscheidungsgrundlage bleiben und diese bei allen moralischen Personen identisch sind, ist jeder von den gleichen Argumenten überzeugt und stellt denselben Rationalkalkül an; folglich treffen alle Akteure im Zeitpunkt ihrer fiktiven Verfassungsentscheidung ohne jede Kommunikation die gleiche Entscheidung. Die Individuen haben „keinen Anlaß zu Verhandlungen im üblichen Sinne“<sup>23</sup>; sie finden einen „monologischen“ Konsens<sup>24</sup>. Die Wahl im Urzustand läßt sich damit vom Standpunkt einer willkürlich ausgewählten Person interpretieren, die als repräsentativ für alle anderen betrachtet werden kann.<sup>25</sup>

Die Gerechtigkeitsnorm, die der Rawlsschen Konzeption der moralischen Person und den Situationsannahmen über den Urzustand zugrunde liegt, ist das aus der Ethik bekannte *Universalisierungspostulat*, welches fordert, „daß man sich in die Lage des anderen versetzt und sich fragt, ob man ... auch im Fall, daß man selbst der Betroffene ist, zu der Handlungsanweisung stehen würde“<sup>26</sup>. Nach Rawls stellt die Konstruktion des Schleiers des Nichtwissens den Versuch dar, den kantischen „kategorischen Imperativ“ bildhaft zu explizieren, der verlangt: „Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde.“<sup>27</sup> „Der Urzustand“, schreibt Rawls, „läßt sich ... auffassen als eine verfahrensmäßige Deutung von Kants Begriff der Autonomie und des Kategorischen Imperativs im Rahmen einer

<sup>23</sup> Rawls (1994 [1971]), S. 163.

<sup>24</sup> Vgl. zum Begriff des monologischen Konsenses im Unterschied zu jenem des diskursiven Konsenses der Diskursethik Habermas (1983), S. 76.

<sup>25</sup> Vgl. Rawls (1994 [1971]), S. 162.

<sup>26</sup> Mackie (1981 [1977]), S. 114. Zur Universalisierungsnorm in der Ethik vgl. z. B. Mackie (1981 [1977]), Kap. 4; Hare (1981), Kap. 6; Wimmer (1980).

<sup>27</sup> Kant (1785), Band 6, S. 51.

empirischen Theorie.“<sup>28</sup> Um festzustellen, ob eine Maßnahme ein allgemeines Gesetz werden könne, müsse sich jeder einzelne in die Schuhe eines jeden Betroffenen versetzen und damit genau jenen fiktiven Rollentausch vornehmen, für den der Schleier des Nichtwissens steht.<sup>29</sup>

In dieser Variante der Vertragstheorie wird also nicht behauptet, daß Menschen im Zeitpunkt von Regelentscheidungen tatsächlich unsicher sind über ihre eigene gesellschaftliche Position; vielmehr fordert die so verstandene Legitimationstheorie, sich hypothetisch in eine solche Situation des konstitutionellen Nichtwissens zu versetzen und reale (Regel-)Entscheidungen so zu treffen, als ob man sich in einer Situation konstitutioneller Unkenntnis befände. Jeder kann den Urzustand zu jedem Zeitpunkt einfach dadurch „betreten“, daß er sich gedanklich den spezifischen Entscheidungsrestriktionen unterwirft und seine Handlungen allein aus diesem Blickwinkel wählt.<sup>30</sup> Die Durchführung des Rawlsschen Gedankenexperiments läuft insoweit auf die Formulierung der Norm hinaus: „Handle bei Regelentscheidungen so, als ob Du von Deinem eigenen Handeln in jeder möglichen gesellschaftlichen Position betroffen wärest.“ Das vertragstheoretische Argument ist in dieser Sichtweise lediglich ein Hilfsmittel, um eine Handlungsalternative auszuwählen, die mit der Universalisierungsnorm vereinbar ist.

Das Universalisierungspostulat, für das der Schleier des Nichtwissens metaphorisch steht, ist eine *inhaltliche (externe) Gerechtigkeitsnorm*.<sup>31</sup> Darunter verstehe ich ergebnisbezogene Werturteile über die Bevorzugung eines oder mehrerer Pareto-Optima auf der Basis eines ergebnisbezogenen Vergleichs alternativer Gesellschaftszustände. Das traditionelle Darstellungsmittel solcher Normen in der ökonomischen Theorie ist die Bergsonsche soziale Wohlfahrts-

---

<sup>28</sup> Rawls (1994 [1971]), S. 289.

<sup>29</sup> Vgl. Rawls (1994 [1971]), S. 159 f., Fußnote 11, und S. 164; vgl. insbesondere auch Abschnitt 40. Zur „kantischen Interpretation“ von Rawls vgl. Wolff (1977); Darwall (1980); Wimmer (1980), S. 163-168; Höffe (1984).

<sup>30</sup> Vgl. Nagel (1975 [1973]), S. 6.

<sup>31</sup> Vgl. auch Goldberg (1974), S. 570. Auch Rawls (1992 [1980]), S. 102 f., räumt dies ein, nicht ohne die Charakterisierung als externe Norm gleich wieder zu relativieren: „(W)as ich ‘die Einschränkungen, denen die Parteien im Urzustand unterworfen sind’ genannt habe, ist in der Tat den Parteien als rationalen Akteuren der Konstruktion äußerlich. Nichtsdestoweniger drücken diese Einschränkungen das Vernünftige aus und daher die formalen Bedingungen, die zu den moralischen Vermögen der Mitglieder einer wohlgeordneten Gesellschaft, die durch die Parteien repräsentiert werden, gehören. ... In *Gerechtigkeit als Fairneß* umrahmt das Vernünftige das Rationale und ist aus einem Begriff der moralischen Person als frei und gleich abgeleitet. Sobald dies verstanden ist, erscheinen die Einschränkungen des Urzustands nicht länger äußerlich.“ (Hervorhebung im Original)



funktion<sup>32</sup>, also eine reellwertige Funktion über die Menge aller gesellschaftlichen Zustände, welche die Ergebnisoffenheit der strikt paretianischen Wohlfahrtsökonomik dadurch auflöst, daß sie - auf der Basis eines über das Pareto-Prinzip hinausreichenden Werturteils - eines der Pareto-Optima, zwischen denen die Pareto-Regel allein nicht unterscheiden kann, zum *Optimum optimorum* erhebt. Das grundsätzlich geltende Pareto-Prinzip wird dadurch weitgehend entwertet: Nicht mehr die Individuen selbst entscheiden, welches aus der Vielzahl der möglichen Pareto-Optima auf der Wohlstandsgrenze gewählt werden sollte, sondern entschieden wird nach Maßgabe der den Betroffenen von außen vorgegebenen Verteilungsnorm.

Die Rawlssche „Theorie der Gerechtigkeit“ geht, wie alle sog. schwachen Vertragstheorien, einen ganz ähnlichen Weg. Will sie nicht darauf verzichten, konkrete Kollektiventscheidungen als legitim auszuzeichnen, muß sie inhaltliche - nicht bloß formale, verfahrensbezogene - Gerechtigkeitsprinzipien einführen, die über das Pareto-Prinzip hinausreichen. Die Entscheidungsbedingungen sind so gestaltet, daß die Individuen sich notwendig so entscheiden, *als ob* sie die zugrunde liegende inhaltliche Gerechtigkeitsnorm - das Universalisierungsprinzip - selbst teilen, und zwar unabhängig davon, ob sie dies tatsächlich tun. Auch die Wohlfahrtsökonomik kennt diese indirekte Art, Verteilungsnormen durch Variation der Entscheidungssituation zu integrieren: So lehrt das Zweite Theorem der Wohlfahrtsökonomik<sup>33</sup>, daß jede zulässige paretoeffiziente Allokation auf der Nutzenmöglichkeitskurve durch eine passende Wahl der Anfangsausstattung der betrachteten Individuen dezentral - d.h. durch (am Markt geschlossene) Verträge - erzeugt werden kann.

Die Auswahl, welche die Individuen unter dem Nichtwissenschleier aus der Menge aller möglichen Pareto-Optima treffen, ist unter Umständen sehr konkret. Ausgangssituationen etwa, die ihrerseits paretooptimal, aber nicht „universalisierbar“ sind, erklärt der Vertragsansatz für illegitim. Verweigert etwa - auf der postkonstitutionellen Ebene - ein Reicher einem Armen das Nötigste zum Leben, so ist diese Situation paretoeffizient. Eine Regel indes, die dieses Pareto-Optimum festschreibt, wäre nicht universalisierbar. Die Universalisierungsnorm stellt hier die Interessen des Armen über die des Reichen. Wüßten beide nicht, in welcher Rolle sie sich wiederfinden werden, wären solche Verteilungsregeln paretoverbessernd, die den Reichen zwar zur Hilfe verpflichten, ihn aber nicht übermäßig schädigen.

<sup>32</sup> Vgl. grundlegend *Bergson* (1938), *ders.* (1948); *Samuelson* (1947), Kap. 8, *ders.* (1981). Überblicke geben *Samuelson* (1981); *Mueller* (1989a), Kap. 19; *Schmidt* (1991), S. 52-80; *Giersch* (1993).

<sup>33</sup> Vgl. z. B. *Varian* (1994 [1992]), S. 329-332.

Auch in Ausgangssituationen, die paretoinferior sind, weist der Schleier des Nichtwissens den Individuen den Weg, welche aus einer Menge mehrerer möglicher Pareto-Verbesserungen auszuwählen sind. Ein „Sklavereivertrag“ etwa, wie Buchanan ihn diskutiert, könnte unter dem Schleier des Nichtwissens - obwohl vielleicht paretoverbessernd - nicht geschlossen werden. Er widerspricht insofern dem kategorischen Imperativ, als der potentielle Sklave nicht freiwillig entscheiden kann und somit als Mensch nur Mittel, nicht aber Selbstzweck wäre.<sup>34</sup> Wüßten beide Beteiligten nicht, wer von ihnen Sklave und wer Herr ist, würden sie sich wohl auf Regeln einigen, die zwar die Erwerbsarbeit des einen für den anderen zulassen, nicht aber dessen würdelose Unterwerfung. Der Schleier des Nichtwissens macht es den Individuen somit unmöglich, einen paretoverbessernden „Sklavereivertrag“ zu schließen, weil nach Maßgabe der Universalisierungsnorm die Interessen des potentiellen Sklaven im Ergebnis über jene des potentiellen Sklavenhalters gestellt werden.

Der Rawlssche Schleier des Nichtwissens führt die Legitimationstheorie also aus dem „vertragstheoretischen Nihilismus“ des starken Kontraktansatzes; aus einer Menge möglicher Pareto-Optima vermag sie einigermaßen scharf zwischen legitimen und illegitimen Vertragsergebnissen zu unterscheiden. Der Preis, den die „Theorie der Gerechtigkeit“ für diese Konkretheit bezahlt, ist indes hoch: Der Schleier raubt den Individuen jene Informationen, die sie benötigten, um einer gegen sie gerichteten Kollektiventscheidung widersprechen zu können; er entzieht ihnen gewissermaßen ihr Vetorecht. Die grundlegende vertragstheoretische Norm - das Pareto-Prinzip, das die Vertragstheorie gerade deshalb eingeführt hatte, damit jedes Individuum sich durch Einlegung seines Vetos einer ihm mißliebigen Kollektiventscheidung widersetzen kann - ist hiermit faktisch aufgegeben. Die Entscheidungen fallen nicht im Urzustand, sondern am Schreibtisch des Vertragstheoretikers. Entscheidend sind dabei nicht die Bewertungen jener Individuen, die von den Folgen der Regelgeltung betroffen sind; diese Personen werden überhaupt nicht gefragt. Maßgebend ist allein eine extern vorgegebene, von den Individuen nicht notwendig geteilte Gerechtigkeitsnorm: der kategorische Imperativ.

Die Rawlssche „Theorie der Gerechtigkeit“ setzt somit die Gerechtigkeitsnormen, die eigentlich erst das Ergebnis der Wahl im Urzustand sein sollen,

---

<sup>34</sup> Die Unmöglichkeit eines „Sklavereivertrags“ deutet *Kant* (1983 [1785]), S. 68, an, wenn er zwar die Arbeitskraft eines Menschen den Marktgesetzen unterworfen sein läßt, nicht aber den Menschen selbst. Ähnlich argumentiert *Rousseau* (1977 [1762]), Buch I, Kapitel 4, S. 11: „Auf seine Freiheit verzichten heißt auf seine Eigenschaft als Mensch, auf seine Menschenrechte, sogar auf seine Pflichten verzichten. Wer auf alles verzichtet, für den ist keine Entschädigung möglich. Ein solcher Verzicht ist unvereinbar mit der Natur des Menschen; seinem Willen jegliche Freiheit nehmen heißt seinen Handlungen jede Sittlichkeit nehmen.“

bereits voraus. Um nicht, wie die starke Vertragstheorie, zu allen verteilungsrelevanten Kollektiventscheidungen schweigen zu müssen, bedarf es der Einführung starker Restriktionen wie der Informationsbeschränkungen durch den Schleier des Nichtwissens; diese Beschränkungen aber wären, da sie den Handlungsspielraum der Individuen variieren, ihrerseits einstimmig zu legitimieren. Ein Unwissenheitsschleier kann daher die Universalisierungsnorm bestenfalls illustrieren, er kann sie jedoch nicht begründen. Der Kalkül unter dem Schleier der Unkenntnis kann niemals mehr sein als ein „Instrument der Normdiskussion“<sup>35</sup>. Statt zu fragen, wie sich fiktive Individuen in einer fiktiven Urvertragssituation mit fiktiven Informationsbedingungen entscheiden würden, könnte man auch unmittelbar fragen, wie die Grundstruktur einer Gesellschaft auszusehen hätte, wenn diese der Universalisierungsnorm entsprechen soll. Das Rawlssche Gedankenexperiment leistet insofern nichts, was man nicht grundsätzlich auch ohne es ableiten könnte: der Umweg über die Vertragstheorie ist entbehrlich.<sup>36</sup>

### **B. Buchanans Konstitutionenökonomik und der „Schleier der Unsicherheit“**

Die voranstehende Diskussion der Zirkularität des Rawlsschen Legitimationsversuchs deutet auf ein grundlegendes Dilemma individualistischer Legitimationstheorien hin. Auf der einen Seite muß eine Theorie, welche die von einer Kollektiventscheidung betroffenen Individuen selbst entscheiden lassen will, was gut und was schlecht für sie ist, auf kollektiver Ebene das Einstimmigkeitskriterium zugrunde legen. Nur bei Geltung dieser Entscheidungsregel verfügt der einzelne dem Markttausch entsprechend über ein Veto, mit dem er sich vor einem Eingriff in seine individuelle Autonomie wirksam schützen kann.

Auf der anderen Seite bedeutet aber gerade die Akzeptanz des Pareto-Kriteriums als höchstrangiger Norm praktisch den Verzicht, zu irgendeiner Kollektiventscheidung Stellung zu nehmen. Denn das Treffen politischer Entscheidungen verlangt fast ausnahmslos, zwischen Pareto-Optima zu wählen, die nach dem Pareto-Kriterium selbst aber nicht miteinander verglichen werden können. Das Anliegen von Legitimationstheorien, dem Politiker oder Verfassungskonstrukteur - vom Verbleib im Status quo abgesehen - individualistisch gerechtfertigte Handlungsempfehlungen zu geben, wird damit verfehlt.

---

<sup>35</sup> Vgl. Zintl (1983), S. 30.

<sup>36</sup> Vgl. Nagel (1975 [1973]), S. 15; Held (1977), S. 718; Kliemt (1977), S. 69 f.; Schmidtz (1991), S. 9 f.; Homann (1988), S. 213.

Jede konkrete Empfehlung eines bestimmten Effizienzpunktes aus der Menge aller möglichen Pareto-Optima setzt demgegenüber eine inhaltliche Gerechtigkeitsnorm über die ergebnisbezogene Wahl zwischen gesellschaftlichen Endzuständen voraus, zu der das Einstimmigkeitskriterium selbst nur subsidiär wäre. Gerade hierdurch jedoch hört, streng genommen, die Legitimationstheorie auf, individualistisch zu sein, da es eben nicht mehr das einstimmige Votum aller Individuen ist, das zur Wahl eines bestimmten Pareto-Optimums führt, sondern das Werturteil des jeweiligen Wissenschaftlers, der die Empfehlung ausspricht, oder des politischen Entscheidungsträgers, der die Entscheidung trifft. Diese Norm ist notwendig „extern“, weil sie ihrerseits nicht durch den Konsens der betroffenen Individuen legitimiert wird.

Die Analogie zwischen der Wertbildung im Markt und jener in der Politik bricht damit zusammen: Am Markt bilden sich, ohne Rückgriff auf „externe“ Normen, Werte *allein* im Konsens; das Tauschergebnis ist damit strikt paretianisch legitimiert. Der wertbildende Maßstab aber, den inhaltliche Gerechtigkeitstheorien an Kollektiventscheidungen anlegen, ist davon wesentlich verschieden, da ein Konsens hier nur dann als legitimitätsstiftend anerkannt wird, wenn dieser bestimmte normative Zusatzanforderungen erfüllt. Die Wertbildung am Markt und in der Politik basiert damit, anders als das „Politik als Tausch“-Paradigma nahelegt, auf *unterschiedlichen* Legitimitätskriterien.

Das Dilemma des „paretianischen“ Ökonomen besteht somit darin, entweder zu nahezu jeder politischen (Regel-)Entscheidung schweigen oder gegen den eigenen Anspruch einer individualistischen Legitimation verstoßen zu müssen, da jene Annahmen, welche die Entscheidungssituation so spezifizieren, daß eine konkrete Entscheidung möglich wird, notwendig auf Normen basieren, die selbst erst individualistisch zu rechtfertigen wären. Die Entscheidung zwischen der Unbestimmtheit des Einstimmigkeitskriteriums auf der einen Seite und einer konkreteren Politikempfehlung auf der anderen Seite impliziert stets eine Entscheidung zwischen der Akzeptanz oder dem Verstoß gegen die Präferenzen der beteiligten Individuen. Anders formuliert, sind Politikempfehlungen, welche die individuelle Autonomie aller Beteiligten akzeptieren - von dem seltenen Extremfall abgesehen, in dem alle Individuen identische Interessen haben - eine Unmöglichkeit.

Während sich Buchanan, wie beschrieben, in zentralen Teilen seines Werkes mit der Indeterminiertheit des Legitimationsarguments einer starken Vertragstheorie abzufinden scheint, gilt dies offensichtlich nicht für jene seiner Arbeiten, die - wie „The Calculus of Consent“ oder „Besteuerung und Staatsgewalt“ - mit der Prämisse einer extremen konstitutionellen Unsicherheit, dem später sog. Schleier der Unsicherheit, arbeiten. In „The Calculus of Consent“ sahen er und Tullock sehr klar, daß die moderne Wohlfahrtsökonomik à la Bergson und Samuelson vor dem beschriebenen Dilemma stand; was sie offenbar nicht er-

kannten, war, daß die Relevanz dieses Dilemmas nicht auf die Wohlfahrtsökonomik beschränkt ist, sondern *allen* Arten paretianischer Legitimationstheorien zu eigen ist. Buchanan und Tullock beanspruchen nämlich ausdrücklich, mit ihrem konstitutionellen Ansatz einen Weg gefunden zu haben, das beschriebene legitimationstheoretische Dilemma vermeiden zu können:

„The approach taken here has the advantage over the new welfare economics in that it does enable us to discuss the organization of social action beyond the limits of the Pareto rule. Whereas the welfare economist either remains silent on all proposals that involve less-than-unanimous support or falls back on some nonindividualistic ethical ordering as given by a ‘social welfare function,’ we are able to describe the individual calculus on the constitutional level.“<sup>37</sup>

Der konstitutionelle Ansatz Buchanans und Tullochs, der hier als eine Art „dritter Weg“ zwischen einer strikt individualistischen und einer inhaltlichen Gerechtigkeitstheorie aufscheint, basiert auf einer dem Schleier des Nichtwissens ähnlichen Informationsbeschränkung, welche die beiden Autoren noch vor Rawls in die Vertragstheorie einführten. Dabei unterstellen sie, daß die einzelnen Kollektivmitglieder unter den Bedingungen einer extremen Unsicherheit über ihre eigenen Positionen in der postkonstitutionellen Phase der Regelgeltung entscheiden. In Anlehnung an Rawls nennen Brennan und Buchanan in ihrem Versuch, die Idee der konstitutionellen Unsicherheit aus „The Calculus of Consent“ argumentativ stärker zu untermauern, diese Annahme daher auch den „*Schleier der Unsicherheit*“<sup>38</sup>.

Das Hauptanliegen von Buchanan und Tullock ist, entgegen dem von Rawls, nicht die Entwicklung einer ethischen Theorie oder einer allgemeinen politischen Gerechtigkeitskonzeption unter Verwendung des analytischen Instrumentariums der Wirtschaftswissenschaft, sondern die Entwicklung einer ökonomischen Theorie der Verfassung mit dem Ziel einer Entmythologisierung der Abstimmungsregel der einfachen Mehrheit als *der* demokratischen Entscheidungsregel schlechthin. Im Mittelpunkt steht der individuelle Kalkül eines beliebigen Teilnehmers einer verfassungsgebenden Versammlung, deren Problem in der Suche nach einer (kosten-)optimalen Kollektiventscheidungsregel für die postkonstitutionelle Gesellschaft besteht.

Die Kosten, die das betrachtete Individuum kalkuliert, sind eine Funktion des sog. Inklusivitätsgrads  $r$ , der den Anteil der erforderlichen Ja-Stimmen umfaßt, die für ein zur Abstimmung stehendes Kollektiventscheidungsproblem erforderlich sind. Das eine Extrem dieses Kontinuums möglicher Abstimmungsregeln,  $r = 1$ , bezeichnet die „Jedermann-Regel“, die jedem einzelnen Kollektiv-

<sup>37</sup> Buchanan/Tullock (1962), S. 93.

<sup>38</sup> Brennan/Buchanan (1993 [1985]), S. 37 ff. Vgl. ebenso Buchanan/Congleton (1998), S. 5-7.

mitglied gewissermaßen die diktatorische Vollmacht einräumt, für alle anderen Mitentscheider verbindliche Entschlüsse zu fassen. Das andere Extrem,  $r = N$ , bildet die Einstimmigkeitsregel, die verlangt, daß eine Kollektiventscheidung nur dann als bindend gelten kann, wenn alle Mitglieder des Kollektivs die Maßnahme befürworten. Zwischen diesen beiden Extremen findet sich die Regel der absoluten Mehrheit ( $N/2 + 1$ ).

Das betrachtete Individuum berücksichtigt zwei Kostenarten: die erwarteten externen Kosten und die Konsensfindungskosten. Die *erwarteten externen Kosten*  $EK(r)$ <sup>39</sup> sind die „Kosten, überstimmt zu werden“, die ihm „von außen“ („extern“) durch die anderen Kollektivmitglieder dadurch auferlegt werden, daß es mit seiner eigenen Meinung unterliegt und eine Kollektiventscheidung gegen seinen Willen getroffen wird. Dabei dürfte es einem einzelnen Entscheidungsträger gleich sein, ob er etwa von  $N-1$  Kollektivmitgliedern, einer absoluten Mehrheit oder nur von einem einzigen Kollektivmitglied überstimmt wird; die daraus für ihn resultierende Nutzeneinbuße wird für ihn stets dieselbe sein. Dies gilt jedoch nicht für die Wahrscheinlichkeit, mit der er diese Kosten zu ertragen hat. Diese ist bei der Einstimmigkeitsregel  $N$  gleich null, da durch sie jedes Kollektivmitglied ein Vetorecht eingeräumt bekommt, mit dem es jede seinen eigenen Vorstellungen widersprechende Kollektiventscheidung verhindern kann. Kollektiventscheidungen, die nach Maßgabe der Einstimmigkeitsregel getroffen werden, sind paretooptimal, da institutionell verhindert wird, daß sich die in der Kollektiventscheidung Obsiegenden allein dadurch besser stellen, daß sie andere Abstimmungsberechtigte schlechter stellen. Maximal ist die Gefahr, ungewünschte Entscheidungskosten zu tragen, dann, wenn bei Existenz der Jedermann-Regel jeder einzelne das Recht hat, verbindlich für alle zu entscheiden. Es ist daher plausibel anzunehmen, daß die erwarteten externen Kosten mit zunehmendem Inklusivitätsgrad  $r$  einen streng monoton fallenden Verlauf haben.

Die zweite Kostenart, die das betrachtete Individuum kalkuliert, sind die *erwarteten Konsensfindungskosten*  $KK(r)$ .<sup>40</sup> Dabei handelt es sich um alle interpersonellen Kosten, die das Individuum in Form von Kosten an Zeit, für Werbung für die eigene Meinung, Mühe etc. aufwenden muß, um die anderen Kollektivmitglieder zu überzeugen. Diese Kosten sind es vor allem, die dem Wunsch der meisten Entscheidungsträger entgegenstehen dürften, paretooptimale Kollektiventscheidungen zu treffen. Denn diese Nutzeneinbußen zur Herbeiführung einer Kollektiventscheidung dürften dann maximal sein, wenn die Entscheidungsregel Einstimmigkeit aller Abstimmungsberechtigten verlangt. Je weniger Ja-Stimmen jedoch verlangt werden, um für alle verbindlich zu ent-

<sup>39</sup> Vgl. *Buchanan/Tullock* (1962), S. 63 ff.

<sup>40</sup> Vgl. *Buchanan/Tullock* (1962), S. 68 f. und S. 97-116.



scheiden, desto geringer dürften auch die Kosten der interpersonellen Entscheidungsfindung sein, die das betrachtete Individuum in seinem Kostenkalkül ansetzen wird. Buchanan und Tullock schließen hieraus, daß die erwarteten Konsensfindungskosten im Regelfall bei der Jedermann-Regel null sein und von da an streng monoton ansteigen werden.

Optimal aus der Sicht des einzelnen Rationalentscheiders dürfte nach dem bisher Gesagten jene Entscheidungsregel sein, bei der die Summe dieser beiden Kostenarten - die *Interdependenzkosten*  $IK(r)^{41}$  - minimal ist. Die Höhe des geforderten kosteneffizienten Inklusivitätsgrads  $r^*$  wird dabei durch die Krümmung der u-förmigen  $IK$ -Kurve bestimmt, die ihrerseits Ausdruck der Dringlichkeit ist, die das Individuum der jeweils im postkonstitutionellen Stadium zur Entscheidung anstehenden Kollektiventscheidung beimißt. Je flacher diese Kurve verläuft, desto indifferenter ist das Individuum gegenüber der Verfassungsfrage. Je gekrümmter die Kurve der sozialen Interdependenz demgegenüber verläuft, desto wichtiger nimmt das Individuum die Abweichungen von der von ihm für optimal gehaltenen Entscheidungsregel. Je steiler die Kurve der Konsensfindungskosten *ceteris paribus* ansteigt, desto geringer ist der Anteil der effizienten Ja-Voten, je steiler die Kurve der erwarteten externen Kosten fällt, desto mehr Ja-Stimmen wird das Individuum *ceteris paribus* für optimal halten. Weder die Einstimmigkeitsregel noch die Jedermann-Regel dürften demnach bei den meisten Entscheidungsproblemen aus individueller Sicht optimal sein, weil im ersten Fall die erwarteten externen Kosten und im zweiten die Konsensfindungskosten maximal sind.

Das zentrale Theorem Buchanans und Tullocks ist es, daß nicht einmal die Regel der einfachen Mehrheit, die auch in der politökonomischen Literatur vielfach einfach mit „Demokratie“ gleichgesetzt wird<sup>42</sup>, unter den klassischen Entscheidungsregeln grundsätzliche Besonderheiten aufweist:

„... on a priori grounds there is nothing in the analysis that points to any uniqueness in the rule that requires a simple majority to be decisive. The  $(N/2 + 1)$  point seems, a priori, to represent nothing more than one among the many possible rules, and it would seem very improbable that this rule should be 'ideally' chosen for more than a very limited set of collective activities. On balance, 51 per cent of the voting population would not seem to be much preferable to 49 per cent.“<sup>43</sup>

<sup>41</sup> Vgl. *Buchanan/Tullock* (1962), S. 69 ff.

<sup>42</sup> So etwa der Gebrauch des Demokratiebegriffs bei *Barry* (1979). Vgl. auch *Benn/Peters* (1959), S. 336 f.

<sup>43</sup> *Buchanan/Tullock* (1962), S. 81. An der Stichhaltigkeit dieses Theorems sind, worauf hier nicht weiter eingegangen werden kann, gleichwohl Zweifel angebracht. Der gewichtigste Kritikpunkt ist wohl, daß die einfache Mehrheitsregel die kleinste Abstimmungsregel ist, die das Treffen sich widersprechender Entscheidungen verhindert; die Kurve der Konsensfindungskosten dürfte folglich im Regelfall eine Sprungstelle bei  $r =$



Der beschriebene „Kalkül der Konsensfindung“ gilt zunächst nur für ein einziges Individuum. Damit die Verfassung legitim im Sinne der Theorien des hypothetischen Gesellschaftsvertrags ist, ist daher noch in einem zweiten Schritt zu zeigen, daß das hergeleitete Kostenoptimum in dieser oder wenigstens in ähnlicher Form auch für alle anderen Teilnehmer der verfassungsgebenden Versammlung gilt. Um zu einer einstimmigen Verfassungsentscheidung zu gelangen, ist somit darzulegen, daß der Kalkül aller Individuen die gleiche Entscheidungsregel als kostenoptimal erweist. Dies ist die Stelle, an welcher sich der *Schleier der Unsicherheit* über die konstitutionellen Entscheidungsträger legt.<sup>44</sup> Buchanan und Tullock unterstellen,

„... that the individual is *uncertain* as to what his own precise role will be in any one of the whole chain of later collective choices that will actually have to be made. For this reason he is considered not to have a particular and distinguishable interest separate and apart from his fellows. This is not to suggest that he will act contrary to his own interest; but the individual will not find it advantageous to vote for rules that may promote sectional, class, or group interests because, by presupposition, he is unable to predict the role that he will be playing in the actual collective decision-making process at any particular time in the future. He cannot predict with any degree of certainty whether he is more likely to be in a winning or a losing coalition on any specific issue. Therefore, he will assume that occasionally he will be in one group and occasionally in the other. His own self-interest will lead him to choose rules that will maximize the utility of an individual in a series of collective decisions with his own preferences on the separate issues being more or less randomly distributed.“<sup>45</sup>

Trotz der individuellen Nutzenmaximierung eines jeden Rationalentscheiders werden nach Buchanan und Tullock die Entscheidungen, die unter den Bedingungen eines solchen Schleiers der Unsicherheit getroffen werden, tendenziell „fair“ und einstimmig sein:

„At the constitutional level, *identifiable* self-interest is not present in terms of external characteristics. The self-interest of the individual participant at this level leads him to take a position as a ‘representative’ or ‘randomly distributed’ participant in the succession of collective choices anticipated. Therefore, he may tend to act, from self-interest, *as if* he were choosing the best set of rules for the social group. Here the pu-

---

*N/2* aufweisen, so daß die Mehrheitsregel auch in *Buchanans* und *Tullocks* Interdependenzkostenkalkül optimal ist. Vgl. zu dieser Eigenschaft der Mehrheitsregel bereits *Locke* (1992 [1689]), § 96, S. 260; *Riemer* (1951), S. 18; im Kontext des „Calculus of Consent“ *Mueller* (1989a), S. 55-57; *ders.* (1996), S. 155 f. - In seinem jüngsten Buch unternimmt *Buchanan* nicht mehr den Versuch, die Regel der einfachen Mehrheit unmittelbar zu attackieren, sondern beschäftigt sich mit institutionellen Arrangements, die geeignet sein könnten, die negativen Folgen dieser Abstimmungsregel zu minimieren; vgl. *Buchanan/Congleton* (1998).

<sup>44</sup> Vgl. *Buchanan/Tullock* (1962), S. 77-80.

<sup>45</sup> *Buchanan/Tullock* (1962), S. 78 (Hervorhebung im Original).

rely selfish individual and the purely altruistic individual may be indistinguishable in their behavior.“<sup>46</sup>

Es wurde bereits dargelegt, daß Buchanan das allen Vertragstheorien zugrunde liegende Werturteil des normativen Individualismus als eine Art Wertfreiheitspostulat interpretiert. Der Schleier der Unsicherheit darf daher nicht die Explikation einer ethischen Norm sein. Sollen nach Maßgabe des individualistischen Postulats nämlich nur die Präferenzen der Betroffenen zählen, dann darf nach seiner Auffassung der Theoretiker ihnen nicht durch Beschränkungen der Entscheidungssituation im Urzustand oktroyieren, sich so zu verhalten, als ob sie bestimmte inhaltliche Gerechtigkeitsnormen teilten. In seiner Theorie wird daher auch nicht verlangt, daß die Individuen sich in eine Situation konstitutioneller Unsicherheit versetzen sollen, um mit ihren Entscheidungen einer bestimmten Gerechtigkeitsvorstellung - etwa der ethischen Universalisierungsnorm - zu genügen.<sup>47</sup> Anders als bei Rawls, ist bei Buchanan der Schleier der Unsicherheit die Metapher für ein *empirisches* Phänomen:

„We postulate that individuals are wholly uncertain about their prospective income-wealth positions in the periods for which the institutional structure to be chosen is to be applicable. Individuals make their own decisions, each behind a Rawlsian ‘veil of ignorance.’ For our purposes this postulate is a *positive* assumption about the actual states of persons making constitutional choice; it is *not* a *normative* statement concerning how persons should conceive themselves in making choices.“<sup>48</sup>

Schon Buchanan und Tullock beschränkten den Anwendungsbereich dieser Annahme nicht auf die konstitutionelle Wahl von Kollektiventscheidungsregeln; der Schleier der Unsicherheit ist nach ihrer Auffassung ein empirisches Phänomen, das kennzeichnend für *alle* Arten von Verfassungsentscheidungen sein soll.<sup>49</sup> Diese reale Unsicherheit der Abstimmenden über ihre eigene postkonstitutionelle Position begründet sich, wie Brennan und Buchanan mehr als zwanzig Jahre später ausführten, mit den Eigenschaften des konstitutionellen Abstimmungsgegenstands - nämlich Regeln - selbst. Zwei Merkmale von Regeln betrachten sie als wesentlich.<sup>50</sup>

<sup>46</sup> Buchanan/Tullock (1962), S. 96 (Hervorhebungen im Original).

<sup>47</sup> Vgl. ausdrücklich Brennan/Buchanan (1993 [1985]), S. 40.

<sup>48</sup> Buchanan/Bush (1977 [1974]), S. 187 (Hervorhebungen von mir; C. M.).

<sup>49</sup> So formulieren Buchanan/Tullock (1962), S. 78: „The uncertainty that is required in order for the individual to be led by his own interest to support constitutional provisions that are generally advantageous to all individuals and to all groups seems likely to be present at any constitutional stage of discussion.“

<sup>50</sup> Vgl. Brennan/Buchanan (1993 [1985]), S. 38 f.; siehe auch Buchanan/Congleton (1998), S. 6; zur Eigenschaft der Quasi-Permanenz von Regeln als Ursache konstitutioneller Positionsunsicherheit vgl. bereits Buchanan (1972), S. 126 f.

Zum einen verfügen Regeln meist über einen hohen *Allgemeinheitsgrad*, da sie auf verschiedene Kollektiventscheidungsprobleme angewendet werden. Ein Milchbauer etwa, der sich in der postkonstitutionellen Gesellschaft für Mindestpreise auf Agrarmärkten einsetzt, könnte auf der konstitutionellen Ebene trotzdem gegen jede Form von staatlichen Preiseingriffen stimmen, da er sich als Verbraucher hiervon den höheren Nutzen erwarten kann.

Zum anderen sind Regeln im allgemeinen durch eine längere Geltungsdauer gekennzeichnet. Sie haben die Eigenschaft der „*Quasi-Permanenz*“, da die gleiche Regel für eine Vielzahl von Kollektiventscheidungsproblemen gilt, deren Gegenstand im Zeitpunkt der Regelentscheidung noch nicht absehbar ist. Verfassungsregeln sind stets zukunftsgerichtet. Hierin liegt gerade der Sinn der Selbstbindung individueller Handlungen, die durch eine Verfassung angestrebt wird. Bildlich gesprochen, gelten Spielregeln stets für mehrere Spielrunden. Wiederum werden eher allgemeine Überlegungen statt Partialinteressen das Abstimmungsverhalten bestimmen.

Brennan und Buchanan gewinnen hieraus die folgende zweistufige Hypothese, deren beide Teile ich als „Begründungsthese“ und als „Wirkungsthese“ bezeichnen werde:

1. „In dem Maße, in dem der Allgemeinheitsgrad und die Geltungsdauer von Regeln zunimmt, wird das Individuum, das sich mehreren Entscheidungsalternativen gegenüber sieht, unsicher im Hinblick auf die Wirkungen, die die einzelnen Alternativen für seinen Netto-Wohlstand haben.“ (*Begründungsthese*)
2. „In dem Maße, in dem eine Person, die sich einer Verfassungsfrage gegenüber sieht, über ihre eigene Position im Rahmen verschiedener Optionen im unklaren ist, wird eine Tendenz bestehen, solchen Arrangements zuzustimmen, die als ‘fair’ in dem Sinne angesehen werden, daß die Resultate, die mittels dieser Regeln zustandekommen, allgemein akzeptabel sind, und zwar unabhängig davon, wo genau ein Teilnehmer sich am Ende befindet.“<sup>51</sup> (*Wirkungsthese*)

Aus der Allgemeinheit und langen Geltungsdauer von Regeln resultiert nach Brennan und Buchanan also der Schleier der Unsicherheit (Begründungsthese); aus dieser Positionsunsicherheit folgern sie wiederum die Fairneß und die Einstimmigkeit der konstitutionellen Abstimmung (Wirkungsthese).

Auch der Schleier der Unsicherheit führt damit nach Buchanan - ohne daß dies mit einer Norm begründet wird - zur „Universalisierung“ des Verhaltens der Beteiligten. Wie der Rawlssche hat auch der Buchanansche Schleier somit eine Fairneß- und eine Konsensfunktion. Aufgrund der unter beiden Varianten konstitutioneller Unkenntnis erwarteten Verhaltensuniversalisierung betonen Brennan und Buchanan sogar, daß „im Grenzfall ... der Schleier der Unsicher-

<sup>51</sup> Brennan/Buchanan (1993 [1985]), S. 39 f. (beide Zitate).

heit mit dem von Rawls ausführlich beschriebenen Schleier des Nichtwissens identisch“<sup>52</sup> ist.

Während die Wirkungsthese so oder ähnlich auch bei Rawls stehen könnte, besteht das besondere Element von Buchanans Konstruktion im ersten Teil des zweistufigen Konstrukts, in der Begründung für die Existenz des Schleiers der Unsicherheit. Mit dieser Begründungsthese werde ich mich in diesem Abschnitt vorrangig auseinandersetzen; die Wirkungsthese steht demgegenüber im Mittelpunkt des folgenden § 4 sowie der modelltheoretischen Analyse der §§ 5 und 6 dieser Arbeit.

In allen Teilen seines Werkes, in denen er mit der Prämisse des Schleiers der Unsicherheit arbeitet, geht Buchanan von der Annahme aus, daß die konstitutionelle Unkenntnis vollkommen, der Schleier der Unsicherheit also „dick“ ist. Die Begründungsthese ließe sich deshalb dahingehend interpretieren, daß Verfassungsregeln immer und überall die Eigenschaften der Allgemeinheit und der Quasi-Permanenz hätten, so daß stets von einem dicken Schleier der Unsicherheit auszugehen sei. Eine solche Fassung der Begründungsthese aber wäre empirisch offensichtlich falsch. Beispiele von Verfassungsregeln, die - trotz langer Geltungsdauern - nicht allgemein waren, sondern einzelne nach Rasse, Religion, Vermögensposition und anderen persönlichen Merkmalen diskriminierten, sind Legion. Die Verfassung der USA erlaubte, obwohl grundsätzlich demokratisch, die Versklavung der schwarzen Zwangsimmigranten aus Afrika.<sup>53</sup> In Südafrika schloß bis vor wenigen Jahren die weiße Minderheit die schwarze Bevölkerungsmehrheit verfassungsmäßig von wesentlichen Teilen des gesellschaftlichen und wirtschaftlichen Lebens aus. Und in einigen westlichen Demokratien hatten Frauen bis weit ins 20. Jahrhundert hinein kein Wahlrecht; noch heute gibt es dort diskriminierende Verfassungsregeln wie die, daß nur Männer eine geschlechtsspezifische Naturalsteuer in Form von Wehrdienst zu leisten haben, während Frauen keinem äquivalenten Arbeitszwang unterliegen.

Die Allgemeinheit stellt also, wie diese Beispiele zeigen, keineswegs eine zwangsläufige Eigenschaft von Verfassungsregeln dar. Dies allein muß allerdings die Begründungsthese nicht widerlegen. Denn die genannten Beispiele haben gemeinsam, daß der Kreis der Betroffenen der jeweiligen Regel von dem des jeweils regelsetzenden Kollektivs abweicht; Diskriminierungen außenste-

<sup>52</sup> Brennan/Buchanan (1993 [1985]), S. 40 f. Vgl. ähnlich Buchanan (1979), S. 196.

<sup>53</sup> So schreibt etwa Weber (1996), S. 215 Fußnote 151: „Die Situation in Nordamerika war in bezug auf die beiden Gruppen ‘Siedler’ und ‘zwangsimmigrierte Afrikaner’ geradezu eine Gegenwelt zu einer Verfassungsfindung hinter einem Rawls’schen *veil of ignorance*. Eine Gruppe von Menschen konnte eine andere in jeder Hinsicht unterwerfen, ohne befürchten zu müssen, in der postkonstitutionellen Zukunft selbst einmal zu den Unterworfenen zu zählen, denn die Diskriminierung war perfekt.“ (Hervorhebung im Original) Vgl. ähnlich auch Mueller (1996), S. 216.

hender, nicht an der Entscheidung beteiligter Individuen sind insofern nicht überraschend.

Buchanan hingegen geht in seinen Theorien grundsätzlich davon aus, daß die von einer Regel Betroffenen selbst - und vollzählig - die einstimmige Entscheidung über den Gesellschaftsvertrag treffen sollen. Bezogen auf Verfassungsnormen, die von *allen* betroffenen Individuen gewählt werden, hat die Begründungsthese auf den ersten Blick eine gewisse Plausibilität. Hätte man die Schwarzen, die Frauen oder die jungen Männer selbst gefragt, ob sie der sie diskriminierenden Regel zustimmen, wären diese Institutionen - so ließe sich vermuten - nie in Kraft getreten. Die Durchsetzung der paretianischen Norm, daß alle Betroffenen dem Gesellschaftsvertrag zustimmen müssen, könnte also als eine Art Garantie dafür angesehen werden, daß überhaupt nur allgemeine Regeln den Individuen zur Wahl vorgelegt werden. Aufgrund der Allgemeinheit der zu beschließenden Regeln existierte dann ein dicker Schleier der Unsicherheit bei allen Verfassungsentscheidungen.

Doch auch diese Überlegung läßt die Begründungsthese im Kontext der Buchananschen Vertragstheorie nicht plausibler erscheinen. Denn auch Personen, die durch eine bestimmte Regel diskriminiert werden, könnten dieser Institution nichtsdestoweniger zustimmen. Widerspruch gegen eine kollektive Entscheidung hat in großen Kollektiven den Charakter eines öffentlichen Gutes: der Nutzen des Erfolgs eines Einspruchs diffundiert über die Allgemeinheit, seine Kosten fallen jedoch individuell an. Haben Personen einen solchen negativen erwarteten Nettonutzen aus der individuellen Beitragsleistung zu einem Kollektivgut, werden sie selbst dann der Wahl einer Regel nicht widersprechen, wenn sie durch sie diskriminiert werden.

In vielen Fällen entfalten Individuen darüber hinaus sogar eine scheinbare Nachfrage nach Paternalismus: Auch einige Frauen etwa waren lange Zeit gegen die Einführung des Frauenwahlrechts; auch die Opfer des indischen Kastensystems argumentierten oft für diese Ordnung; und noch heute gibt es in allen Staaten glühende Verfechter der Monarchie, die gleichwohl wissen, daß sie unter einer solchen Herrschaftsregel selbst nur Untertanen wären, die von den Privilegien des Monarchen nicht profitieren würden. Es gibt eine Vielzahl solcher Fälle einer möglichen „Präferenzfalsifikation“<sup>54</sup> - einer Divergenz zwischen öffentlich geäußerten und tatsächlichen, „privaten“ Präferenzen. Der in einem Klima öffentlichen Drucks erzielbare Nutzen, eine „politisch korrekte“ Meinung zu äußern, wird hier durch jene Handlungsalternative maximiert, die das Individuum nach Maßgabe seiner privaten Präferenzen in Wahrheit ablehnt.

---

<sup>54</sup> Kuran (1995).

Die Möglichkeit, daß sogar diejenigen Individuen Regeln zustimmen, durch die sie selbst diskriminiert werden, ist nicht lediglich eine spitzfindige und daher zu vernachlässigende Ausnahme; im Kontext des Buchananschen Ansatzes kommt ihr hohe Relevanz zu, da die starke Vertragstheorie keinerlei normative Mindestanforderungen an die Fairneß der konstitutionellen Entscheidungssituation stellt. Selbst Entscheidungen unter Zwang sind danach annahmegemäß legitimitätsstiftend. So stimmt auch der künftige Sklave - vor die „Wahl“ zwischen „Tod oder Unterwerfung“ gestellt - dem Buchananschen „Sklavereivertrag“ zu, der damit eine Pareto-Verbesserung darstellt. Allgemein ist diese Regel jedoch nicht; die beiden Hauptakteure entscheiden in voller Kenntnis ihrer jeweiligen Partialinteressen als Ausbeuter und als überlebender Sklave. Ein Schleier der Unsicherheit existiert nicht.

Besteht aber die Möglichkeit, daß - trotz des konstitutionellen Einstimmigkeitserfordernisses - auch solche Regeln den hypothetischen Verfassungskonvent passieren, die einige der Abstimmenden diskriminieren, so kann - die Richtigkeit der Wirkungsthese vorausgesetzt - nicht davon ausgegangen werden, daß ein dicker Schleier der Unsicherheit ein Phänomen ist, das empirisch immer kennzeichnend ist für konstitutionelle Entscheidungssituationen. Vielmehr wird man realistischerweise verschiedene Allgemeinheitsgrade der zu testenden Institutionen unterscheiden müssen sowie verschiedene „Dicken“ des Schleiers der Unsicherheit. Für die Legitimationstheorie müßte dies bedeuten, daß für jede zu rechtfertigende Regel - in Abhängigkeit von ihrem Allgemeinheitsgrad - grundsätzlich ein anderes Modell der Ursituation zu verwenden wäre: Soll die mutmaßliche Legitimität diskriminierender Verfassungsregeln wie der genannten geprüft werden, so wäre von einem entsprechend dünnen Schleier der Unsicherheit auszugehen, in bestimmten Fällen möglicherweise sogar von *sicheren* Erwartungen im Entscheidungszeitpunkt. Soll jedoch die hypothetische Legitimität höchst allgemeiner Regeln getestet werden, so hätte das Modell größtmögliche konstitutionelle Unkenntnis zu unterstellen. Steht beispielsweise die Einführung rechtsstaatlicher Institutionen zur Debatte, so kann der einzelne Abstimmungsteilnehmer nicht wissen, welche Angriffe auf sein Persönlichkeitsrecht er in der postkonstitutionellen Gesellschaft erleiden oder welche juristischen Auseinandersetzungen er mit anderen Bürgern ausfechten wird. Weil ihm insofern seine späteren Partialinteressen zum Entscheidungszeitpunkt noch verborgen sind, müßte hier ein dichter Schleier der Unsicherheit zugrunde gelegt werden.

Buchanans Werk läßt jedoch einen solchen, nach dem Allgemeinheitsgrad der jeweils zu legitimierenden Regel differenzierenden Gebrauch des Unsicherheitsschleiers nicht erkennen. In jenen konstitutionenökonomischen Arbeiten, in denen er diese Prämisse verwendet, geht er - unabhängig von der zur Wahl



stehenden Regel - von einem dicken Schleier der Unsicherheit aus.<sup>55</sup> Wegen dieser Annahme beschränkt er sich darauf, den Kalkül eines einzigen Individuums, das unter den gewählten Unsicherheitsbedingungen als repräsentativ für alle anderen Abstimmenden angesehen wird, zu untersuchen. Gerade in „Die Grenzen der Freiheit“ jedoch - also in jenem Buch, das eine besonders „realistische“ Beschreibung des Naturzustands zu geben beansprucht<sup>56</sup> - taucht die Annahme des Schleiers der Unsicherheit überhaupt nicht auf, obwohl Buchanan hierin Regeln höchster Allgemeinheit und Quasi-Permanenz - wie vor allem die beschriebene Rechtsstaatsregel - herzuleiten sucht. Wenn der Schleier der Unsicherheit aber ein empirisches Phänomen und nicht lediglich die Metapher für ein Gerechtigkeitsurteil sein soll, so drängt sich die Frage auf, warum auch und gerade an dieser Stelle nicht mit konstitutioneller Unsicherheit zu rechnen sein sollte, zumal der „Calculus of Consent“, wo diese Annahme zum ersten Mal Verwendung fand, mehr als ein Jahrzehnt früher erschienen war.

Doch auch eine ganz andere Überlegung läßt die Begründungsthese für den Schleier der Unsicherheit unplausibel erscheinen. Ist nämlich, neben der Quasi-Permanenz, die Allgemeinheit einer Regel die Begründung für die Unsicherheit der Entscheider über ihre eigenen Partialinteressen, so wird damit genau genommen - als Bedingung der Existenz eines Schleiers der Unsicherheit - vorausgesetzt, was diese Annahme erst herstellen soll: die Allgemeinheit des gewählten institutionellen Arrangements. Denn der Buchanansche Begriff der Allgemeinheit (*generality*) einer Regel kann kaum etwas anderes bedeuten als ihre „Universalisierbarkeit“. Auch Buchanan und Congleton verwenden den Begriff der „Fairneß“ bzw. „Reziprozität“ einer Regel weitgehend austauschbar mit dem der Allgemeinheit.<sup>57</sup>

Die Buchanansche Begründung der „empirischen“ Positionsunsicherheit ist darum zirkulär: die Universalisierbarkeit einer Regel ist die *Voraussetzung* für die Existenz eines Schleiers der Unsicherheit (Begründungsthese); Entscheidungen unter dieser Unsicherheit haben wiederum die Universalisierbarkeit der

<sup>55</sup> Vgl. etwa die Verwendung des Schleiers der Unsicherheit in *Buchanan/Tullock* (1962) oder in *Brennan/Buchanan* (1988 [1980]).

<sup>56</sup> Vgl. *Buchanan* (1984 [1975]), S. 76 ff.

<sup>57</sup> Daß die Universalisierbarkeit bzw. Fairneß einer Regel bei *Buchanan/Congleton* (1998) ihre Allgemeinheit bedeutet, wird etwa in folgenden Passagen deutlich: „... reciprocity ... implies generality. Constraining rules that emerge from general agreement will tend to be *general* in application. Rules that apply to others must also be applied to one's own behavior.“ (S. 6; Hervorhebung im Original) Oder: „... elementary notions of fairness dictate that the same rules apply to all players. ... the generality norm here is ... defined as an attribute of fairness ...“ (S. 10) Schon früher hatte *Buchanan* (1977c), S. 244, betont, daß die Gleichheit vor dem Gesetz (*equality before the law*) vertragstheoretisches Legitimationsergebnis sei und diese im Steuerrecht „Nondiscrimination in prices“ (S. 246) - also Allgemeinheit im obigen Sinne - verlange.



Regel zur *Folge* (Wirkungsthese). Das gewünschte Legitimationsergebnis wird mit der Allgemeinheit als angeblich empirischer Eigenschaft von Regeln bereits vorausgesetzt. Des Umweges über den Schleier der Unsicherheit bedarf es insoweit überhaupt nicht: Es kann nicht verwundern, daß Individuen, denen nur universalisierbare Regeln zur Entscheidung vorgelegt werden, sich nur auf universalisierbare Regeln einigen werden.

Nach dem bisher Gesagten liegt der Verdacht nahe, daß die Behauptung, der Schleier der Unsicherheit sei ein empirisches Phänomen, vor allem eine *ad hoc* - d.h. nicht unabhängig vom gewünschten Legitimationsergebnis - eingeführte Interpretation ist, um einerseits die „externe“ Universalisierungsnorm aus der Verfassungsökonomik herauszuhalten, andererseits aber trotzdem aus dem Rechtfertigungsargument Ergebnisse herzuleiten, die mit dieser Norm in Einklang zu bringen sind.<sup>58</sup> Die eigentliche Motivation Buchanans für die Einführung der Annahme konstitutioneller Unsicherheit scheint sich insofern nicht grundsätzlich von derjenigen Rawls' zu unterscheiden, dessen „Theorie der Gerechtigkeit“ ganz explizit zu im kantischen Sinne universalisierbaren Ergebnissen führen soll. Für eine solche „kantische Interpretation“ auch des Buchananschen Schleiers der Unsicherheit<sup>59</sup> spricht, daß Buchanan immer wieder auf die Ethik Kants Bezug nimmt. So hat er sich wiederholt zu einem Personenbegriff bekannt, der verlangt, Menschen als Selbstzweck und niemals nur als Mittel zu betrachten<sup>60</sup>; dies entspricht Kants kategorischem Imperativ in seiner dritten Formulierung.<sup>61</sup> Auch hat Buchanan bereits früher darauf hingewiesen, er und Rawls teilten die gleiche „quasi-kantische“ Ausgangsnorm der Vertragstheorie.<sup>62</sup> Gelegentlich hat Buchanan sogar direkt auf die Universalisierbarkeitsvoraussetzung vertragstheoretischer Abreden Bezug genommen.<sup>63</sup> Beson-

<sup>58</sup> Vgl. auch Müller (1997), S. 247.

<sup>59</sup> Für eine „kantische Interpretation“ des Buchananschen Werks, die sich insbesondere auf das ohne einen Schleier der Unsicherheit auskommende Buch „Die Grenzen der Freiheit“ stützt, vgl. auch Petersen (1996).

<sup>60</sup> So schreibt Buchanan (1977c), S. 244: „In my view, democratic values must be founded on the basic Kantian notion that individual human beings are the ultimate ethical units, that persons are to be treated strictly as ends and never as means ...“ Vgl. auch Buchanan (1972), S. 124.

<sup>61</sup> „Handle so, daß du die Menschheit, sowohl in deiner Person, als in der Person eines jeden andern, jederzeit zugleich als Zweck, niemals bloß als Mittel brauchest.“ Kant (1983 [1785]), S. 61.

<sup>62</sup> Vgl. Buchanan (1977 [1976]), S. 209: „We (= Rawls and Buchanan; C. M.) share a set of quasi-Kantian, contractarian presuppositions as opposed to a Benthamite utilitarian conception.“

<sup>63</sup> „Any justification or legitimization of rules changes that fails the unanimity test must call upon non-contractarian criteria of evaluation, which must be *non-individualistic* in origin, or at least, non-individualistic in any *universalizable* sense.“ Buchanan (1986), S. 245 (zweite Hervorhebung von mir; C. M.).

ders deutlich wird die „kantische Interpretation“ des Schleiers der Unsicherheit schließlich in Buchanans jüngstem, gemeinsam mit Roger Congleton verfaßten Buch „Politics by Principle, Not Interest“, in dem die Autoren ständig zwischen einer Verwendung des Schleiers der Unsicherheit und einer Argumentation mit ihrem „Allgemeinheitsprinzip“ schwanken, das erklärtermaßen kantisch inspiriert ist.<sup>64</sup> Der Schleier der Unsicherheit dient hier also, ganz wie bei Rawls, als eine Prämisse, welche die fiktiven Entscheidungsträger zur Wahl von im Sinne der Universalisierungsnorm verallgemeinerbaren Regeln zwingen soll.

Wie bei Rawls, scheint daher auch bei Buchanan - wenngleich uneingestanden - die Entscheidung der ökonomischen Rationalentscheider unter konstitutioneller Positionsunsicherheit lediglich ein anderes Darstellungsmittel für eine Entscheidung zu sein, die kantisch-rationale moralische Personen (unter sicheren Erwartungen) treffen würden. Die Individuen sind damit nicht mehr wirklich frei, buchstäblich jede von ihnen gewünschte Entscheidung zu treffen. Der Schleier der Unsicherheit ist vielmehr so geschneidert, daß er die verallgemeinerbaren unter den Pareto-Optima zu *Optima optimorum* erhebt. Die ergebnisbezogene Auswahl gesellschaftlicher Endzustände trifft daher nicht jeder einzelne unter realer Unsicherheit handelnde Entscheidungsträger im Urzustand, sondern die ethische Norm der Verhaltensuniversalisierung. Strikt individualistisch ist eine solche Vertragstheorie nicht; die mit dem Schleier der Unsicherheit operierenden Teile der Buchananschen Verfassungsökonomik haben den Charakter einer inhaltlichen Gerechtigkeitstheorie.

---

<sup>64</sup> Vgl. *Buchanan/Congleton* (1998), S. xi.

## § 4 Die schwache Vertragstheorie (II.): Die Konklusion

In der ökonomischen Literatur wird der vertragstheoretische Kalkül der fiktiven Rationalakteure unter dem Schleier der Unkenntnis in seinen beiden Varianten häufig allein als eine Metapher für eine Entscheidung nach Maßgabe des Universalisierungsprinzips interpretiert. Um festzustellen, ob eine Entscheidung verallgemeinerbar ist, genügt es, sich hilfsweise in einen fiktiven Urzustand hineinzudenken und sich zu fragen, ob die betrachteten rationalen Individuen sich in der gleichen Weise entscheiden würden. Zwar ist die Universalisierungsnorm eine inhaltliche Gerechtigkeitsnorm, die selbst nicht individualistisch legitimiert ist. Dies wird aber zum einen deshalb als vergleichsweise unschädlich angesehen, weil das Verallgemeinerungsprinzip eine weithin geteilte moralische Norm darstellt, die im Zentrum des politischen Liberalismus steht, dem die Ökonomen traditionell zuneigen.

Zum anderen ist die Verallgemeinerungsnorm - obwohl endzustandsbezogen - weithin offen und beläßt den Individuen einen großen Spielraum, innerhalb dessen sie zwischen gesellschaftlichen Zuständen frei entscheiden können. Das mit der Prämisse eines Schleiers der Unkenntnis operierende Gedankenexperiment definiert - in seiner ersten Stufe - ebenso wie das Pareto-Prinzip eine kollektive Auswahlregel, welche die Eigenschaften einer *strikten partialen Ordnung* aufweist. Da dieser *Collective choice rule* die wesentliche Eigenschaft der Vollständigkeit fehlt, kann sie zwar einige, nicht jedoch alle gesellschaftlichen Regeln (Situationen) in eine kollektive Rangfolge bringen. Beispielsweise läßt sich plausibel argumentieren, daß sich Individuen in einer Anarchiesituation unter einem Schleier der Unkenntnis auf die Errichtung eines im weiteren Sinne demokratischen Staates einigen werden; *welche* konkrete Form der Demokratie die Individuen aber wählen werden - etwa eine direkte oder eine indirekte Repräsentationsform -, läßt sich mit Hilfe dieser kollektiven Auswahlregel nicht entscheiden. Der vertragstheoretische Universalisierungskalkül der ersten Stufe kann *zwischen* Regeln, die allesamt universalisierbar sind, nicht vergleichen. Verlangt man, eine Regel habe den vertragstheoretischen Universalisierbarkeitstest zu bestehen, so beläßt man den Individuen - im Rahmen der Verallgemeinerbarkeit - einen großen Teil ihrer Entscheidungsfreiheit. Die liberale Regelperspektive, deren hervorstechendster Charakter ihre Ergebnisoffenheit ist, wird damit also nicht (völlig) verlassen.

Wäre eine bildhafte Darstellung einer Entscheidung, die dem Verallgemeinerungspostulat entspricht, die einzige Funktion, die dem Gedankenexperiment mit dem Schleier der Unkenntnis zukommt, so könnte man sich jedoch die ganze vertragstheoretische Argumentation sparen: Statt sich in fiktive Rationalakteure unter fiktiven Entscheidungsbedingungen hineinzudenken, könnte man auch unmittelbar danach fragen, welche Regeln universalisierbar im Sinne des kategorischen Imperativs sind und welche nicht. Die Vertragstheorie wäre nichts als ein krypto-normativer Ansatz, dessen einzige Aufgabe es wäre, die alleinige Verwendung des Pareto-Prinzips zu behaupten, aber die Einführung der zusätzlichen inhaltlichen Gerechtigkeitsnorm zu „verschleiern“. Eine unmittelbare Verwendung der Universalisierungsnorm wäre in diesem Fall daher auch klarer, da das verwendete Werturteil offen zutage läge und auf diese Weise einer Kritik zugänglich wäre.

Die Vertragstheorie von Rawls und - trotz gegenteiliger Behauptungen ihres Verfassers - offenbar auch jene von Buchanan wollen jedoch mehr als das. Die Definition einer Situation, die Entscheidungen induziert, welche dem Verallgemeinerungsprinzip entsprechen, ist nur ein erster Schritt ihrer vertragstheoretischen Argumentation. Ihr Ziel ist es nicht nur zu zeigen, daß unter bestimmten idealisierenden Prämissen über die Informationsverteilung in der konstitutionellen Ausgangssituation selbst ökonomisch-rationale Individuen wie rein moralische Personen entscheiden würden; vielmehr versuchen sie - in einem zweiten Schritt - durch einen deduktiven Schluß zu beweisen, daß sich die rationalen Individuen in der jeweiligen Entscheidungssituation im Sinne einer ganz *konkreten* Gerechtigkeitsauffassung einigen würden.

In der Folgerung einer bestimmten „politischen Gerechtigkeitskonzeption“ aus den allgemein anerkannten Prämissen des Urzustands besteht bei Rawls die eigentliche „Begründungs-“Leistung der Vertragstheorie. Wer die Universalisierungsnorm teilt, so ließe sich sein Legitimationsanliegen verkürzen, muß auch die „Maximin-Lösung der sozialen Gerechtigkeit“ - gegossen in seine beiden berühmten Gerechtigkeitsgrundsätze - wollen, die das Wohl des jeweils am schlechtesten gestellten Individuums in der Gesellschaft befördern soll.

Buchanan hingegen spricht von einer beabsichtigten „Normbegründung“ durch seinen Vertragsansatz nirgendwo explizit. Wer auf der konstitutionellen Ebene unter einem Schleier der Unsicherheit entscheidet, so argumentierte er jedoch wiederholt, maximiert seinen Erwartungsnutzen unter der Annahme, sich mit gleicher Wahrscheinlichkeit in jeder postkonstitutionellen Gesellschaftsposition wiederzufinden; das Gedankenexperiment unter dem Schleier der Unsicherheit, so dieses Argument in Kurzform, impliziert ein übereinstimmendes Verhalten nach Maßgabe einer spezifischen (Laplace-)Wohlfahrtsfunktion.<sup>1</sup>

Die Konsequenzen dieses Deduktionsschritts sind weitreichend. Gelingt er, so wandelt sich die unvollständige strikte partiale Ordnung, welche der Universalisierungskalkül (erster Argumentationsschritt) konstituierte, in eine *vollständige*, mittels einer sozialen Wohlfahrtsfunktion repräsentierte *Ordnung* sozialer Regeln bzw. Zustände (zweiter Argumentationsschritt). Alle gesellschaftlichen Situationen - auch die universalisierbaren - werden nunmehr miteinander vergleichbar. Das vertragstheoretische Argument kann nun nicht mehr nur eine Menge von Regeln als legitim ausweisen; im Fall einer strikt konkaven Nutzenmöglichkeitsgrenze wird es nunmehr einen einzigen gesellschaftlichen Zustand - ein *Optimum optimorum* - auswählen. Die vertragstheoretische Rechtfertigungstheorie wird mit diesem Schritt völlig ergebnisbezogen. Welche Entscheidung legitim ist und welche nicht, liegt von vornherein exakt fest; ein Entscheidungsspielraum verbleibt den betroffenen Individuen nicht mehr.

Mit dem Deduktionsschritt von der unvollständigen kollektiven Auswahlregel zur sozialen Wohlfahrtsfunktion - von der ergebnisoffenen zur vollkommen endzustandsbezogenen Gerechtigkeitskonzeption - überschreitet der kontraktualistische Rechtfertigungsansatz die letzte Schwelle vom Individualismus zum Kollektivismus. An die Stelle der letzten Relikte des liberalen Freiheitsideals, nach dem die Menschen - gegeben die Freiheitsrechte aller anderen - selbst über die sie betreffenden Belange entscheiden sollen, tritt, ganz wie in der überwunden geglaubten modernen Wohlfahrtsökonomik à la Bergson und Samuelson, eine endzustandsbezogene Gerechtigkeitsutopie vom guten Leben in Form einer sozialen Wohlfahrtsfunktion, die sich nicht nur völlig unabhängig von ihren Realisierungsmöglichkeiten, sondern auch von den Wünschen der betroffenen Individuen bestimmt.<sup>2</sup>

Im vorliegenden § 4 werde ich allerdings bestreiten, daß man mit der Akzeptanz der legitimationsverleihenden Kraft des Urzustandsmodells diesen Weg in den Kollektivismus mitgehen muß. Ich werde sowohl für die Rawlssche als auch für die Buchanan'sche Theorie die Ansicht zurückweisen, daß der jeweilige Deduktionsschritt tatsächlich eine zwingende Folgerung aus den Prämissen der jeweiligen vertragstheoretischen Entscheidungssituation darstellt.

Die Aufgabe sicherzustellen, daß das einstimmig gewählte *Optimum optimorum* gerecht im Sinne der jeweils konkretisierten Gerechtigkeitsauffassung ist, fällt der zugrunde gelegten Theorie der Rationalentscheidung unter unsiche-

---

<sup>1</sup> Beide konkreten Gerechtigkeitsauffassungen werden im einzelnen in diesem Kapitel vorgestellt und ausführlich diskutiert.

<sup>2</sup> Für die moderne Wohlfahrtsökonomik hat bereits Bergson (1948) betont, daß die soziale Wohlfahrtsfunktion auf einer sozialistischen Gerechtigkeitsvorstellung beruht.

ren Erwartungen zu. Je nach entscheidungstheoretischer Interpretation der Ausgangssituation ergibt sich entsprechend bei Rawls eine Alleinberücksichtigung der Interessen des am schlechtesten gestellten Individuums, bei Buchanan eine gleichwertige Berücksichtigung der Interessen aller Positionsinhaber. In Abschnitt A. dieses Kapitels werde ich jedoch zu zeigen versuchen, daß man keines der beiden Interessengewichtungsprinzipien zu teilen gezwungen ist, auch wenn man den Schleier der Unkenntnis als Metapher für eine moralnorm-analoge Entscheidung akzeptiert. Die Akzeptanz der einen oder der anderen Gewichtungsvariante folgt nicht notwendig aus dem Kalkül unter dem Schleier der Unkenntnis, sondern allein aus der impliziten Einführung eines zusätzlichen, über die Universalisierungsnorm hinausgehenden externen Werturteils.

Darüber hinaus werde ich in Abschnitt B. argumentieren, daß nicht einmal dann, wenn die Individuen alle das gleiche Entscheidungskriterium unter unsicheren Erwartungen anwendeten, eine *Einigung* im Sinne der jeweils konkreten Gerechtigkeitsauffassung zu erwarten wäre. Ob sich die Individuen im Urzustand einigen oder nicht, hängt vom „Universalisierungsgrad“ ab, d.h. vom Ausmaß der Berücksichtigung fremder Interessen in der eigenen konstitutionellen Nutzenfunktion. Unter dem Rawlsschen Schleier des Nichtwissens ist dieser Universalisierungsgrad vollkommen, unter Buchanans Schleier der Unsicherheit nur partiell. Ich werde argumentieren, daß die Buchanansche Theorie die konstitutionelle Einstimmigkeit nicht, die Rawlssche Theorie nur formal garantieren kann.

### **A. Die Gerechtigkeitskonzeption: Das Entscheidungskriterium unter Unsicherheit**

Die Deduktion einer konkreten Fairneßvorstellung aus den Prämissen des Urzustandes hängt maßgeblich von der zugrunde gelegten Theorie der Entscheidung unter unsicheren Erwartungen ab. Rawls konstruiert den Urzustand in entscheidungslogischer Hinsicht als eine Situation der *Unsicherheit i.e.S.*<sup>3</sup>, in der niemand in der Lage ist, subjektive Wahrscheinlichkeiten dafür anzugeben,

---

<sup>3</sup> In der Entscheidungstheorie unterscheidet man zwei Arten von Entscheidungen unter *Unsicherheit*: Entscheidungen unter *Unsicherheit i.e.S.* und solche unter *Risiko*. Beide haben gemeinsam, daß sich der Entscheidungsträger mindestens zwei Umweltzuständen gegenüber sieht, von denen genau einer eintreten wird. Während aber in Unsicherheitssituationen i.e.S. der Entscheider lediglich in der Lage ist, anzugeben, welche Umweltzustände überhaupt eintreten werden, vermögen rationale Individuen nach der Bayesschen Entscheidungstheorie in Risikosituationen auch subjektive Wahrscheinlichkeitsurteile für das Eintreten der möglichen Umweltzustände zu bilden. Vgl. zu dieser Unterscheidung z. B. *Luce/Raiffa* (1957), S. 13; *Harsanyi* (1977), S. 22.

sich in der einen oder anderen postkonstitutionellen Situation wiederzufinden. Entsprechend dem Maximinprinzip der individuellen Rationalentscheidung unter Unsicherheit entwickelt er seine „Maximin-Lösung des Problems der sozialen Gerechtigkeit“ (Abschnitt I.). Buchanans Schleier der Unsicherheit definiert hingegen - seinem Namen zum Trotz - aus entscheidungslogischer Sicht eine Entscheidung unter *Risiko*, da die Individuen annahmegemäß in der Lage sind, jeder Möglichkeit, sich in einer gesellschaftlichen Position wiederzufinden, eine positive subjektive Wahrscheinlichkeit zuzuordnen. Buchanan hält es für rational, daß die Akteure auf der konstitutionellen Entscheidungsstufe gemäß dem Laplace-Prinzip vom unzureichenden Grund entscheiden und es folglich für gleich wahrscheinlich halten, die eine oder andere gesellschaftliche Position zu besetzen (Abschnitt II.).

### I. Maximin-Gerechtigkeit

Gegenstand der Entscheidung der Individuen im Rawlsschen Urzustand ist nicht unmittelbar die Wahl einer Regel bzw. eines Regelsystems in Form einer gesellschaftlichen „Grundstruktur“, sondern die Entscheidung für eine Gerechtigkeitskonzeption, welche eine optimale Grundstruktur für die Gesellschaft impliziert. Die Wahl dieser Gerechtigkeitskonzeption ist endgültig, d.h. für die ganze, potentiell unendliche Geschichte der betreffenden Gesellschaft verbindlich. Das zu wählende Gerechtigkeitskriterium soll allgemein verständlich sein und eindeutige, nachvollziehbare Bewertungen liefern; es soll Gegenstand eines dauerhaften, übergreifenden Konsenses sein können.<sup>4</sup>

Rawls legt den fiktiven Entscheidungsträgern im Urzustand eine Liste historisch einflußreicher Gerechtigkeitskonzeptionen vor, die für ihre Wahl in Frage kommen könnten.<sup>5</sup> Diese Liste, die nicht abschließend ist und beliebig erweitert werden kann, enthält als wichtigste Gegenspieler zu der von Rawls präferierten Gerechtigkeitskonzeption die Varianten des klassischen Utilitarismus, dessen Gemeinwohlbegriff in der Maximierung der ungewichteten Nutzensumme (des „Glücks der größten Zahl“) bzw. des Durchschnittsnutzens aller Gesellschaftsmitglieder besteht.<sup>6</sup>

---

<sup>4</sup> Vgl. Pogge (1994), S. 65.

<sup>5</sup> Vgl. Rawls (1994 [1971]), S. 146 f.

<sup>6</sup> Vgl. Bentham (1970 [1781]), S. 12: „The community is a fictitious *body*, composed of the individual persons who are considered as constituting as it were its *members*. The interest of the community then is, what? - the sum of the interests of the several members who compose it.“ (Hervorhebungen im Original)



Der Urzustand ist so gestaltet, daß die Individuen rationalerweise die Rawlsche Gerechtigkeitskonzeption wählen. Indem die Beteiligten unter dem Schleier des Nichtwissens aller Informationen über ihre persönlichen und sozialen Lebensumstände beraubt sind, entscheiden sie bei der Wahl von Gerechtigkeitskonzeptionen allein nach Maßgabe ihrer höchstrangigen Interessen, die jeder von ihnen als moralische Person an den Grundgütern hat. Daher wählen sie Gerechtigkeitsgrundsätze, die ihnen die notwendigen institutionellen und materiellen Voraussetzungen dafür gewähren, daß sie ihre beiden grundlegenden moralischen Vermögen entwickeln und ausüben können.

Der Schleier des Nichtwissens definiert nach Rawls eine Situation der *Unsicherheit i.e.S.*, da er den Rationalentscheidern auch die Möglichkeit nimmt, irgendwelche Wahrscheinlichkeiten dafür anzugeben, sich in der postkonstitutionellen Gesellschaft in der einen oder der anderen sozialen Position wiederzufinden. Folglich geht Rawls davon aus, daß die Individuen unter dem Schleier des Nichtwissens ein typisches Rationalentscheidungskriterium unter unsicheren Erwartungen - das Maximin-Kriterium der ökonomischen Entscheidungstheorie - verwenden. Die Maximin-Entscheidungsregel<sup>7</sup> orientiert sich am ungünstigsten Ergebnis und empfiehlt, jene Handlungsalternative zu wählen, bei der von allen Aktionen das schlechteste Ergebnis maximiert wird.

Wendet ein Entscheidungsträger die Maximin-Regel im Urzustand als Auswahlkriterium für eine Gerechtigkeitskonzeption an, so entscheidet er, als wüßte er, daß er unter jeder beliebigen Sozialordnung auf jeden Fall die schlechteste Position einnehmen wird, so als würde ihm in der Gesellschaft „sein Feind einen Platz zuweisen“<sup>8</sup>. Gewählt wird folglich jene Gerechtigkeitskonzeption, durch welche der jeweils Armseligste am besten gestellt ist. Unter dieser Gerechtigkeitskonzeption sind die Schlechtestgestellten besser als unter jedem anderen Gerechtigkeitskonzept in der Lage, ihre drei höherrangigen Interessen zu verfolgen.

Die Gerechtigkeitskonzeption, welche die Individuen nach Rawls in dieser Unsicherheitssituation wählen und welche die drei regulativen Interessen moralischer Individuen am besten zum Ausdruck bringen soll, besteht aus zwei Gerechtigkeitsprinzipien zur Verteilung von Grundgütern. Nach Rawls ist die Wahl der beiden Gerechtigkeitsgrundsätze „die einzige Lösung des Problems, das der Urzustand stellt.“<sup>9</sup> In einer neueren Formulierung lauten sie:

---

<sup>7</sup> Vgl. grundlegend Wald (1945) sowie Milnor (1964 [1951]); Luce/Raiffa (1957), S. 278-280; Krelle (1968), S. 185 f.

<sup>8</sup> Rawls (1994 [1971]), S. 178.

<sup>9</sup> Rawls (1994 [1971]), S. 141.

1. „Jede Person hat ein gleiches Recht auf ein völlig adäquates System gleicher Grundrechte und Grundfreiheiten, das mit dem gleichen System für alle anderen vereinbar ist.“

2. „Soziale und ökonomische Ungleichheiten müssen zwei Bedingungen erfüllen: *erstens* müssen sie mit Ämtern und Positionen verbunden sein, die allen unter Bedingungen fairer Chancengleichheit offenstehen, und *zweitens* müssen sie zum größten Vorteil der am wenigsten begünstigten Mitglieder der Gesellschaft sein.“<sup>10</sup>

Beide Grundsätze - die Rawls auch die „Maximin-Lösung des Problems der sozialen Gerechtigkeit“<sup>11</sup> nennt - sind auf unterschiedliche Kategorien von Grundgütern gerichtet. Der erste Gerechtigkeitsgrundsatz - der *Grundsatz gleicher Grundfreiheiten* -, der die Etablierung eines demokratischen und liberalen Verfassungsstaats fordert, regelt die Verteilung der ersten beiden Grundgüter - der politischen und bürgerlichen Grundfreiheiten. Das zweite Prinzip bezieht sich auf die letzten drei Grundgüter: auf soziale Vorrechte, Einkommen und Besitz.<sup>12</sup>

Zwischen beiden Gerechtigkeitsprinzipien herrschen Vorrangverhältnisse: Der Grundsatz gleicher Grundfreiheiten ist dem zweiten Prinzip, wie Rawls sagt, „lexikalisch“ (gemeint ist: „lexikographisch“<sup>13</sup>) vorgeordnet.<sup>14</sup> Dies ergibt sich aus der Gewichtung der Grundgüter durch die fiktiven Individuen im Urzustand: Wissen die Beteiligten, daß sie als moralische Personen zwar die drei höherrangigen Interessen haben, kennen sie aber unter dem Schleier des Nichtwissens deren konkreten Inhalt nicht, so werden sie nach Rawls den Grundrechten absoluten Vorrang geben vor allen übrigen Grundgütern. Vergleicht man zwei Grundgüterausstattungen miteinander, so hat man immer zunächst auf die Grundfreiheiten und -rechte zu achten; wer in bezug auf seine Freiheitsrechte besser ausgestattet ist als ein anderer, der möglicherweise mehr von nachrangigen Gütern besitzt, gilt dem anderen gegenüber als besser gestellt.<sup>15</sup>

In bezug auf die Freiheit ist Rawls' Theorie strikt egalitär.<sup>16</sup> Der „Vorrang der Freiheit“ grenzt die Rawlsianische scharf von der utilitaristischen Ethik ab, die Rechte nur relativ zum jeweiligen Gemeinwohlkalkül, nicht aber absolut legitimieren kann. Da er den einzelnen gegebenenfalls dem Glück aller opfert, ist der

<sup>10</sup> Rawls (1992 [1985]), S. 261 (Hervorhebungen von mir; C. M.). Die ausführlichste Fassung der beiden Gerechtigkeitsprinzipien fand sich ursprünglich in Rawls (1994 [1971]), S. 336 f.

<sup>11</sup> Rawls (1994 [1971]), S. 177.

<sup>12</sup> Vgl. Rawls (1993), S. 76; Hinsch (1997a), S. 10 f.

<sup>13</sup> Den Begriff „lexikographisch“ hält Rawls (1994 [1971]), S. 62, für „zu umständlich“.

<sup>14</sup> Vgl. Rawls (1994 [1971]), S. 82.

<sup>15</sup> Vgl. Pogge (1994), S. 73.

<sup>16</sup> Vgl. Nagel (1975 [1973]), S. 3.

Mensch im Utilitarismus nur ein Mittel, nicht aber ein Zweck an sich. Um sich selbst davor zu schützen, daß sie der Bedürfnisbefriedigung anderer Menschen zum Opfer fallen, werden nach Rawls die Individuen unter dem Schleier des Nichtwissens daher den allen Individuen gleichermaßen zukommenden Freiheiten - dem „Rechten“ - einen absoluten Primat gegenüber der Interessenverfolgung - dem „Guten“ - einzuräumen.<sup>17</sup> Der Vorrang der Freiheit ist also Rawls' Art, den moralischen Imperativ, Menschen nur als Selbstzweck zu betrachten, zu verwirklichen.<sup>18</sup>

Einen solchen lexikographischen Vorrang räumen die Entscheidungsträger im Urzustand auch dem ersten Teil des zweiten Prinzips - dem Grundsatz der fairen Chancengleichheit, kurz: dem „*Chancenprinzip*“<sup>19</sup> - ein, das dem zweiten Teil des zweiten Grundsatzes, dem sog. *Differenzprinzip*, vorgeordnet ist; das Differenzprinzip soll nur dann zum Zuge kommen, wenn die Bedingungen des Chancenprinzips erfüllt sind.<sup>20</sup> In Konfliktfällen müssen die armseligsten Gesellschaftsmitglieder also grundsätzlich vermeidbare Einschränkungen ihrer Einkommen und Vermögen hinnehmen, wenn dies die einzige Möglichkeit ist, das Chancenprinzip zu verwirklichen. Konkret verlangt dieser Grundsatz nicht nur formale Chancengleichheit in dem Sinne, daß in Politik und Gesellschaft geltende Regeln niemanden diskriminieren dürfen. Erforderlich ist vielmehr auch „faire Chancengleichheit“, nach der „Menschen mit gleichen Fähigkeiten und gleicher Bereitschaft, sie einzusetzen, gleiche Erfolgsaussichten haben sollen, unabhängig von ihrer anfänglichen gesellschaftlichen Stellung.“<sup>21</sup>

Bei der Behandlung der Rawlsschen Gerechtigkeitskonzeption werde ich mich allein auf das Differenzprinzip konzentrieren und sowohl den Grundsatz gleicher Grundfreiheiten als auch das Chancenprinzip vernachlässigen. Für diese Beschränkung spricht nicht nur, daß das Differenzprinzip, welches insbesondere eine Richtschnur für die Sozialpolitik sein soll, der für eine ökonomische Analyse relevanteste Rawlssche Grundsatz ist. Das Differenzprinzip ist darüber hinaus auch deshalb von Interesse, weil es eine soziale Wohlfahrtsfunktion impliziert und sich daher unmittelbar mit der Tradition der modernen Wohlfahrtsökonomik vereinbaren läßt. Der wichtigste Grund für die Beschränkung auf das Differenzprinzip ist indes, daß dieser Grundsatz die meisten Parallelen zum Buchanan'schen Ansatz aufweist, um den es in dieser Arbeit besonders geht.

<sup>17</sup> Vgl. Rawls (1994 [1971], S. 607.

<sup>18</sup> Vgl. Wolff (1977), S. 88; Arneson (1992), S. 213.

<sup>19</sup> Diesen Begriff verwendet Pogge (1994), 93 ff.

<sup>20</sup> Vgl. Rawls (1994 [1971]), S. 110 und S. 337.

<sup>21</sup> Rawls (1994 [1971]), S. 93. Vgl. zum Chancenprinzip im einzelnen Pogge (1994), S. 93-106.

Nach dem Differenzprinzip sind sozioökonomische Ungleichheiten in einer Gesellschaft nur dann gerechtfertigt, wenn die in der Gesellschaft am schlechtesten gestellten Individuen davon profitieren.<sup>22</sup> Grundlage des Differenzprinzips ist die Überzeugung, daß zufällige Merkmale und Eigenschaften von Menschen unverdient sind und diesen daher nicht zum Vorteil gereichen dürfen, es sei denn, die Schlechtestgestellten profitieren hiervon. Das Differenzprinzip zeichnet damit diejenige Grundstruktur einer Gesellschaft aus, welche die bestmögliche sozioökonomische Mindestposition hervorbringt. Gemessen wird diese Minimalposition anhand des Grundgüterindex eines repräsentativen Individuums aus der Gruppe der Schlechtestgestellten.<sup>23</sup>

Rawls hat die Gerechtigkeitsauffassung, für welche das Maximinprinzip steht, nicht erfunden. Als moralische Norm ist sie schon jahrtausendealt. So findet sich bereits in der Bibel die Aufforderung, auf das Wohl der Schlechtestgestellten zu achten: „Was ihr für einen meiner geringsten Brüder getan habt, das habt ihr mir getan.“<sup>24</sup> Rawls war nicht einmal der erste, der dieses Postulat systematisch auf die Verteilungspolitik angewendet sehen wollte. Im Jahr 1953 stellte Gerhard Weisser in einem Vorlesungszyklus „Grundsätze der Verteilungspolitik“ vor, die zentrale Aussagen der Rawlsschen Theorie antizipieren.<sup>25</sup> Wie bei Rawls, ist auch bei Weisser nicht allein die Regelung der Einkommens- und Vermögensverteilung Gegenstand der Redistributionspolitik. Verteilt werden vielmehr „Lebenslagen“, die starke Überschneidungen mit den Rawlsschen Grundgütern aufweisen; zu Weissers Lebenslagen gehören Güter wie die Arbeitsfreude, die Sicherheit der Lebenshaltung oder das „Selbstbewußtsein, das sich aus dem Gefühl, ein aktives Glied der Gesellschaft zu sein, ... ergibt.“<sup>26</sup> Und wie bei Rawls steht im Mittelpunkt von Weissers Verteilungstheorie ein Distributionsprinzip, welches das Los der in der Gesellschaft am schlechtesten gestellten Individuen verbessern will. Dieser Verteilungsgrundsatz, den Weisser seinerseits - ohne nähere Angaben - dem in der Nazizeit umgekommenen Philosophen Kurt Grelling zuschreibt, lautet:

„Unter mehreren zur Auswahl stehenden Wirtschaftsverfassungen soll diejenige bevorzugt werden, bei der die Lebenslage der wirtschaftlich schwächsten Schichten günstiger als in jeder anderen zur Auswahl stehenden Wirtschaftsverfassung ist.“<sup>27</sup>

<sup>22</sup> Vgl. Rawls (1994 [1971]), S. 60-65 und S. 336 f.

<sup>23</sup> Ein solcher Rückgriff auf ein „repräsentatives Individuum“ ist allerdings höchst problematisch, wie Kirman (1992) am Beispiel der Makroökonomik zeigt.

<sup>24</sup> Matthäus 25,40.

<sup>25</sup> Vgl. Weisser (1971 [1954]).

<sup>26</sup> Weisser (1971 [1954]), S. 111. Die Ähnlichkeit zum Rawlsschen Grundgut „soziale Grundlagen der Selbstachtung“ liegt hier auf der Hand.

<sup>27</sup> Weisser (1971 [1954]), S. 131 (im Original komplett kursiv). Zum Hinweis auf Grelling vgl. dort Anmerkung 3.

Originär von Rawls ist indes - neben einer detaillierten Ausleuchtung der Implikationen dieses Verteilungsprinzips, die bei Weisser offen bleiben - die vertragstheoretische Begründung des Differenzprinzips mit dem Gedankenexperiment des Urzustandes. Zu diesem Zweck verwendet Rawls die folgende Argumentationskette: (1) Normativer Ausgangspunkt der Gestaltung sozialer Institutionen ist die kantische Universalisierungsnorm, für die der Schleier des Nichtwissens lediglich ein bildhaftes Darstellungsmittel ist. (2) Unter diesem Schleier des Nichtwissens ist es rational, sich nach der Maximin-Regel der Entscheidungstheorie zu verhalten und das Differenzprinzip zu wählen. Losgelöst von der metaphorischen Sprache der Vertragstheorie, ist der Kern des Arguments von Rawls daher: Eine Akzeptanz des Universalisierungspostulats impliziert die Pflicht, die gesellschaftliche Grundstruktur nach Maßgabe des Differenzprinzips zu gestalten.

Diese Deduktion einer Maximin-Wohlfahrtsfunktion aus den Prämissen des Urzustandes unter dem Schleier des Nichtwissens ist der wohl umstrittenste Teil von Rawls' Gerechtigkeitstheorie.<sup>28</sup> Bereits in der ökonomischen Theorie rationaler Entscheidungen, aus der diese Regel ursprünglich stammt, ist das Maximin-Prinzip stark kritisiert worden und wird heute kaum mehr ernsthaft als Entscheidungsregel im Unsicherheitsfall diskutiert. Denn ungeachtet der Höhe des Gewinns, den eine andere Alternative möglicherweise bietet, fordert das Maximin-Prinzip dazu auf, stets jene Alternative zu wählen, die den geringsten Verlust zuläßt, und zwar selbst dann, wenn die Wahrscheinlichkeit, diesen

---

<sup>28</sup> Die Literatur zu Rawls' Maximin-Prinzip ist unüberschaubar, und es ist im Rahmen der vorliegenden Arbeit weder möglich noch nötig, sie zu wiederholen. Kritisch äußern sich u.a. *Arrow* (1973); *Barry* (1973), *ders.* (1989), v.a. S. 213 ff., 330 ff. *Gordon* (1973); *Pettit* (1974); *Musgrave* (1974); *Scanlon* (1975 [1973]), S. 197-205; *Mueller/Tollison/Willet* (1974), S. 345-350; *Nagel* (1975); *Rae* (1975); *Harsanyi* (1976 [1975]); *Samuelson* (1976); *Wolff* (1977), S. 55 f., 67-70; *Gibbard* (1979); *Kulenkampff* (1979); *Kaye* (1980); *Koller* (1983), *ders.* (1987), S. 93-102; *Mueller* (1989a), S. 414-423; *Kersting* (1993), S. 75-81; *Jakob* (1996), S. 192-203; *Beckmann* (1998), S. 127-137. (Partielle) Verteidigungen des Maximin-Prinzips finden sich bei *Rawls* (1974), *Phelps* (1976); *Beauchamp* (1980); *Corrado* (1980); *Binmore* (1994), S. 316 f. *Strasnick* (1976) versuchte sogar, mit einer an *Arrow* orientierten *Social-choice*-Axiomatisierung des Rawlsschen Urzustandes zu beweisen, daß bei Akzeptanz der Rawlsschen Ausgangsbedingungen man auch auf die Akzeptanz des Maximin-Prinzips festgelegt sei; dieser „Beweis“ beruht jedoch, wie *Wolff* (1976) gezeigt hat, nicht nur auf unplausiblen Annahmen, sondern ist auch in formaler Hinsicht falsch. Zum Teil wird zur Begründung des Maximinprinzips auch auf andere Rechtfertigungsmethoden als die der Entscheidung unter dem Schleier des Nichtwissens zurückgegriffen, so z. B. auf die Methode der Rechtfertigung in einer fiktiven Situation der Gleichheit (*Barry*, 1989, S. 226 ff.), das Verfahren der Wahl unter Bedingungen fiktiver Gleichheit an Handlungs- und Lebenschancen (*Wildt*, 1996, S. 272) und die Methode der öffentlichen Rechtfertigung (*Hinsch* 1997a,b).

Verlust zu erleiden, nahe null ist. Hat ein New Yorker Bürger etwa die Wahl zwischen der Alternative, ein Flugzeug - das mit einer geringen Wahrscheinlichkeit abstürzen könnte - nach Chicago zu besteigen, um dort einen sehr einträglichen Job anzunehmen, und der Alternative, seine alte, wesentlich weniger lukrative Stelle in New York zu behalten, so ist es nach dem Maximin-Prinzip rational, in New York zu bleiben; denn der schlimmste denkbare Umweltzustand - das Abstürzen mit einem Flugzeug - wird dadurch in jedem Fall vermieden.<sup>29</sup>

Schon das Bestehen der geringsten Möglichkeit also, daß bei Wahl einer Handlung eine sehr negative Konsequenz eintrifft, läßt es nach dem Maximin-Prinzip als rational erscheinen, den Eintritt dieses Verlustes - unabhängig davon, wie wahrscheinlich dieser Fall tatsächlich ist - mit einer Wahrscheinlichkeit von eins zu erwarten. Dem Maximin-Prinzip der Entscheidungstheorie ist daher vorgeworfen worden, es liege „an der äußersten Grenze der Rationalität“, indem es „einen geradezu pathologischen Pessimismus voraussetzt.“<sup>30</sup> Würde man das Maximin-Prinzip ernst nehmen, so spottete auch Harsanyi in seiner Kritik der Rawlsschen Theorie,

„... then you could not ever cross a street (after all, you might be hit by a car); you could never drive over a bridge (after all, it might collapse); you could never get married (after all, it might end in a disaster), etc. If anybody really acted this way he would soon end up in a mental institution.“<sup>31</sup>

Rawls bestreitet überhaupt nicht, daß „die Maximin-Regel im allgemeinen keine gute Regel für Entscheidungen unter Unsicherheit“<sup>32</sup> darstellt, sondern bestenfalls nur in ganz bestimmten Situationen nützlich ist. Die Annahme gleicher Wahrscheinlichkeiten, wie sie in der Buchanan-Variante des Schleiers der Unkenntnis verlangt wird, lehnt Rawls aber vor allem mit drei Begründungen ab:<sup>33</sup>

Als ersten Grund führt Rawls an, daß für die Individuen im Urzustand sehr viel auf dem Spiel steht. Es handele sich nicht nur um eine Wahl von Alternativen, sondern die zu wählenden Prinzipien im Urzustand beeinflussen das ganze Leben. In den fundamentalen Fragen ihres Lebens könne es somit rational sein, sich extrem risikoscheu zu verhalten. Dieses Argument vermag indes nur bedingt zu überzeugen, da auch in Harsanyis Beispiel des Flugzeugs eine Menge -

<sup>29</sup> Vgl. zu diesem Beispiel *Harsanyi* (1976 [1975]), S. 39.

<sup>30</sup> *Krelle* (1968), S. 185 (beide Zitate).

<sup>31</sup> *Harsanyi* (1976 [1975]), S. 40.

<sup>32</sup> *Rawls* (1994 [1971]), S. 179.

<sup>33</sup> Zur Begründung der Unmöglichkeit von Wahrscheinlichkeitsschätzungen im Urzustand vgl. *Rawls* (1994 [1971]), S. 179 ff.

nämlich das eigene Leben - auf dem Spiel steht und es nach landläufiger Meinung trotzdem rational sein kann, ein Flugzeug zu besteigen.<sup>34</sup>

Ein zweiter Grund für die Maximin-Entscheidungsregel ist nach Rawls, daß es dem einzelnen im Urzustand „so gut wie gleichgültig ist, ob er über das nach der Maximin-Regel sicher erreichbare Minimum hinaus noch etwas gewinnt. Es ist ihm nicht der Mühe wert, einen darüber hinausgehenden Vorteil zu suchen, vor allem, wenn die Gefahr besteht, daß er vieles ihm Wichtige verliert.“<sup>35</sup> Dieses Argument ist jedoch noch weniger überzeugend als das erste, weil *erstens* die Individuen unter dem Schleier des Nichtwissens den Entwicklungsstand ihrer Gesellschaft nicht kennen. Dann aber können sie auch nicht wissen, welches „sicher erreichbare Minimum“ durch die Maximin-Regel überhaupt gewährleistet wird und ob sie damit zufrieden sein können;<sup>36</sup> auch stellt sich die Frage, warum alle Individuen unabhängig voneinander das gleiche soziale Minimum für akzeptabel halten sollten.<sup>37</sup> *Zweitens* dürfte das Verlangen, über das Minimum hinaus noch etwas hinzuzugewinnen, wesentlich von der Risikoeinstellung der Beteiligten abhängen. Nur dann, wenn die Individuen den Grad ihrer Risikoaversion kennen, können sie wissen, wie hoch der Risikoabschlag ist, den sie im Vergleich zum Nutzen des Erwartungswerts einer Lotterie abzugeben bereit sind. Gerade diese Risikoneigung - „die Besonderheiten seiner Psyche wie seine Einstellung zum Risiko oder seine Neigung zu Optimismus oder Pessimismus“<sup>38</sup> - kennt der einzelne nach Rawls' Annahme aber nicht unter dem Schleier des Nichtwissens. Es ist daher „überhaupt nicht einzusehen, wieso es rational sein soll, sich so zu verhalten, *als ob* man Pessimist und risikoscheu wäre, wenn man nicht weiß, ob man es tatsächlich ist, und man der Möglichkeit, es zu sein, nicht einmal einen Wahrscheinlichkeitswert zuordnen kann.“<sup>39</sup>

Als drittes Argument zugunsten der Maximin-Regel führt Rawls an, daß die im Urzustand abgelehnten alternativen Gerechtigkeitsprinzipien „zu unannehmbaren Ergebnissen führen.“<sup>40</sup> Das utilitaristische Nutzenprinzip - das wohl wichtigste alternative Prinzip, das Rawls seinen fiktiven Urzustandsbewohnern zur Entscheidung vorlegt - hat nach Rawls etwa die unannehmbare Folge, daß der

„größere Vorteil anderer ... auch den Benachteiligten Grund genug zu sein [hat; C. M.], während ihres ganzen Lebens Verzicht zu leisten. ... Stellt man sich die Ge-

<sup>34</sup> Vgl. Hinsch (1997a), S. 77 f.

<sup>35</sup> Rawls (1994 [1971]), S. 179.

<sup>36</sup> Vgl. Nagel (1975 [1973]), S. 12; Kaye (1980), S. 38 f.

<sup>37</sup> Vgl. Rae (1975), S. 636 f.

<sup>38</sup> Rawls (1994 [1971]), S. 160.

<sup>39</sup> Kulenkampff (1979), S. 102.

<sup>40</sup> Rawls (1994 [1971]), S. 179.



sellschaft als ein System der Zusammenarbeit zur Förderung des Wohls seiner Mitglieder vor, so will es überhaupt nicht einleuchten, daß die politischen Grundsätze von einigen Bürgern verlangen könnten, um anderer willen noch geringere Lebenschancen in Kauf zu nehmen.“<sup>41</sup>

So zutreffend diese Kritik am Utilitarismus auch sein mag: zum Argument für das Differenzprinzip läßt sie sich nicht wenden, gilt sie doch nicht minder für die Rawlssche Maximin-Regel. Wie im Utilitarismus gibt es auch nach dem Maximin-Prinzip eine Gruppe von Personen, die um der Interessen anderer willen geringere Lebenschancen in Kauf zu nehmen hat. Nach dem Maximinprinzip sind dies alle Gesellschaftsmitglieder, die nicht die armseligste Position bekleiden. Die Begriffe „Nutzen-“ und „Differenzprinzip“ sind, worauf Thomas Nagel hingewiesen hat, in der zitierten Passage völlig austauschbar; der einzige Unterschied zwischen beiden Argumenten besteht darin, daß nach dem Maximin-Prinzip stets die Chancen der Bessergestellten denen der Benachteiligten geopfert werden. In der postkonstitutionellen Gesellschaft aber wird man Schwierigkeiten haben, die Maximin-Lösung, nach der allein die Interessen der Schlechtergestellten zählen, bessergestellten Personen gegenüber zu rechtfertigen.<sup>42</sup>

Ganz generell scheint das Argument, alternative Gerechtigkeitsprinzipien führten im Einzelfall zu unannehmbaren Ergebnissen, bestenfalls geeignet zu sein, ein soziales Minimum oder den Vorrang der Freiheit zu rechtfertigen; für eine darüber hinausgehende allgemeine Anwendung des Maximin-Prinzips wären demgegenüber stärker egalitäre Annahmen erforderlich.<sup>43</sup>

Von diesen (und anderen) Rawlsschen Argumenten abgesehen<sup>44</sup>, ist erst kürzlich für das Maximin-Prinzip ins Feld geführt worden, daß die Entscheidung, in ein Flugzeug zu steigen, und die Entscheidung über Gerechtigkeitsgrundsätze im Urzustand einen wesentlichen Unterschied aufwiesen: die Annahme, die Sache werde „schon gutgehen“, sei mit Blick auf die Verteilung gesellschaftlicher Positionen, notwendigerweise falsch. Denn auch wenn man hinter dem Schleier des Nichtwissens nicht sagen könne, wer *in persona* zur Gruppe der Benachteiligten gehören wird, so wüßten doch alle, daß einige Gesellschaftsmitglieder dieses Los in jedem Falle zu erdulden hätten. Der Kalkül, die Wahrscheinlichkeit dieser Benachteiligung sei denkbar gering, könne daher

<sup>41</sup> Rawls (1994 [1991]), S. 204.

<sup>42</sup> Vgl. Scanlon (1975[1973]), S. 198; Rae (1975), S. 637 f.; Mueller (1989a), S. 417 f.

<sup>43</sup> Vgl. Nagel (1975 [1973]), S. 12 f. - Daß auch das Maximin-Prinzip zu unannehmbaren Ergebnissen führen kann, wird in § 5 dieser Arbeit gezeigt.

<sup>44</sup> Weitere Rawlssche Argumente für das Maximin-Prinzip finden sich in Rawls (1974).

nicht aufgehen.<sup>45</sup> Doch auch dieses Argument für das Maximin-Prinzip sticht nur dann, wenn man ein zentrales Axiom der Rawlsschen Theorie aufgibt: die Annahme der gegenseitigen Desinteressiertheit der Individuen im Urzustand. Ist es den Rationalentscheidern unter dem Schleier des Nichtwissens jedoch gleichgültig, daß es - solange man nicht selbst betroffen ist - einigen von ihnen schlecht gehen wird, so scheint es aus einer rein eigeninteressierten Perspektive betrachtet keineswegs irrational zu sein darauf zu bauen, daß man selbst schon nicht soviel Pech haben werde, ausgerechnet die schlechteste Position in der Gesellschaft einzunehmen.

Als Zwischenergebnis läßt sich damit festhalten, daß selbst dann, wenn man die Universalisierungsnorm und den Schleier des Nichtwissens als ihre bildhafte Explikation akzeptiert, man nicht gezwungen ist, die Ansicht zu teilen, daß allein das Wohl des Schlechtestgestellten zu maximieren sei. Die Universalisierungsnorm verlangt nur, daß ein Entscheider sich in die Lage der anderen Betroffenen versetzen soll; über die Gewichtung der berücksichtigten Präferenzen sagt sie indes nichts aus. Die Festlegung einer Gewichtungsregel wie dem Maximin-Prinzip verlangt ein *zusätzliches*, stärkeres Gerechtigkeitsurteil, das nicht aus dem Universalisierungspostulat folgt. Die beiden Rawlsschen Gerechtigkeitsgrundsätze sind daher nicht, wie behauptet, „die einzige Lösung des Problems, das der Urzustand stellt.“<sup>46</sup> In § 5 dieser Arbeit werde ich darüber hinaus zeigen, daß unter bestimmten Umständen ein Verhalten nach dem Maximin-Prinzip mit den Implikationen des Universalisierungspostulats kollidiert.

## II. Laplace-Gerechtigkeit

In bezug auf die von seinen Annahmen implizierte konkrete Gerechtigkeitsauffassung bleiben Buchanans Ausführungen sehr knapp. Tatsächlich weist er sogar die Vermutung von sich, er habe eine „Normbegründung“ - die Deduktion einer speziellen Gerechtigkeitsnorm aus seinem vertragstheoretischen Argument - beabsichtigt. In einer retrospektiven Betrachtung des gemeinsam mit Tullock verfaßten „Calculus of Consent“ schreibt Buchanan beispielsweise:

„Our analysis differed from that of Rawls ... in the important respect that we made *no attempt to generate specific predictions* about what might emerge from the prospective agreement among the contractors who choose rules from behind the veil of uncertainty. ... our construction suggested that *no single decision rule* was likely to be chosen for general applicability over the whole range of political action. We used the construction to eliminate some sets of outcomes rather than to specify those sets that

<sup>45</sup> So Hirsch (1997a), S. 95 f.

<sup>46</sup> Rawls (1994 [1971]), S. 141.

would be selected. By contrast, Rawls was led (we think, misled) to attempt to use the veil of ignorance construction to make specific predictions. He suggested that his two principles of justice would uniquely emerge from the preconstitutional stage of contractual agreement."<sup>47</sup>

Der Kalkül unter dem Schleier der Unsicherheit, so erscheint es in dieser Passage, soll allein ein Test für die Universalisierbarkeit sozialer Regeln sein. Aus der Menge der wählbaren Alternativen sortiert er jene Regeln aus, die *nicht* als legitim anzusehen sind; eine Rangordnung innerhalb der verbleibenden legitimen Regeln wird indes nicht angestrebt. Das konstitutionelle Wohlfahrtskriterium Buchanans definiert nach dieser Textpassage also lediglich eine strikte partielle Ordnung, nicht aber eine vollständige Ordnung aller gesellschaftlichen Zustände durch eine soziale Wohlfahrtsfunktion.

Nicht nur einmal hat Buchanan wortgewaltig den klassischen Utilitarismus und die aus ihm hervorgegangene moderne Wohlfahrtsökonomik<sup>48</sup> für ihren Versuch attackiert, inhaltliche Gerechtigkeitsnormen in Form von sozialen Wohlfahrtsfunktionen zu formulieren, um aus „kollektiver Sicht“ zwischen gesellschaftlichen Endzuständen zu unterscheiden.<sup>49</sup> Vertreter solcher (und ähnlicher) Auffassungen verspottete er mit der Bemerkung, sie „spielten Gott“, indem sie ihre eigenen privaten Wertanschauungen verabsolutierten, anstatt sich an den Interessen der Betroffenen zu orientieren.<sup>50</sup> Auch Rawls warf er, wenngleich deutlich milder im Ton, wiederholt vor, mit der Herleitung seines Differenzprinzips die (starke) vertragstheoretische Argumentation zu verlassen und sich dem Verdacht auszusetzen, den kontraktualistischen Rahmen nur dazu zu verwenden, eine Norm zu empfehlen, die Rawls aus persönlichen Gründen für wünschenswert halte.<sup>51</sup>

Es gehört zu den Widersprüchen in Buchanans Werk, daß er die Maßstäbe, die er an andere Autoren anlegt, häufig selbst zu brechen pflegt. Immer dann nämlich, wenn Buchanan konkret wird in bezug auf die Beantwortung der Frage, wie er sich selbst den Kalkül unter dem Schleier der Unsicherheit vorstellt, greift er nicht nur auf das in der ökonomischen Theorie übliche Rationali-

<sup>47</sup> Buchanan (1991 [1988]), S. 48 (Hervorhebungen von mir; C. M.). Vgl. ähnlich Buchanan (1977a), S. 132.

<sup>48</sup> Daß die Wohlfahrtsökonomik eine Weiterentwicklung des klassischen Utilitarismus sei, behauptet bereits Robbins (1932), S. 141. Eine ausführliche Begründung erfährt diese Behauptung durch Bohnen (1964). Die gegenteilige Auffassung vertritt jedoch Schumann (1994), S. 219 f.

<sup>49</sup> Vgl. z. B. Buchanan (1977 [1962]), S. 215; ders. (1977 [1976]), S. 209; ders. (1977a), S. 126; ders. (1979), S. 22-24.

<sup>50</sup> Vgl. zu solchen und ähnlichen Formulierungen z. B. Buchanan (1977a), S. 134; ders. (1977b), S. 142.

<sup>51</sup> Vgl. Buchanan (1977a), S. 124 und 132.

tätsprinzip der Maximierung des Erwartungsnutzens zurück, sondern auch darauf, daß die Individuen erwarten, jede gesellschaftliche Position mit *gleicher* Wahrscheinlichkeit zu besetzen.<sup>52</sup> Diese beiden Prämissen zusammengenommen implizieren aber, bei Anwendung der gängigen ökonomischen Entscheidungslogik, genau eine solche endzustandsbezogene soziale Wohlfahrtsfunktion vom Bergson-Typ<sup>53</sup>, die gesellschaftliche Situationen in eine vollständige kollektive Präferenzordnung bringt; in diesem Sinne ist die Position Buchanans daher auch immer wieder interpretiert worden.<sup>54</sup> Ist diese Interpretation zutref-

<sup>52</sup> Zur Annahme der Maximierung des Erwartungsnutzens unter dem Schleier der Unsicherheit vgl. z. B. *Buchanan* (1967), S. 236; *Buchanan/Faith* (1980), S. 26. Zur Annahme *gleicher* Wahrscheinlichkeiten (bzw. gleicher Gewichtung) aller Positionen unter dem Schleier der Unkenntnis vgl. z. B. *Buchanan* (1967), S. 234; *ders.* (1977 [1976]), S. 200; *ders.* (1986), S. 244; *Buchanan/Faith* (1980), S. 25 f.

<sup>53</sup> Die *Bergsonsche* Wohlfahrtsfunktion  $W_i$  ist eine reellwertige, ordinale Funktion  $W$  eines Individuums  $i$  über die Menge  $X$  aller gesellschaftlichen Zustände  $x$ :  $W_i(x) = W_i(x_1, x_2, \dots, x_n)$ . Inhaltlich wird  $W_i$  üblicherweise durch das sog. Welfarismus-Postulat (vgl. z. B. *Sen* 1982 [1977], S. 227 und S. 248-251) konkretisiert, dem zufolge die Bewertung gesellschaftlicher Zustände allein den von ihnen betroffenen Individuen zu überlassen ist.  $W_i$  wird daher auch in Abhängigkeit von den Nutzenwerten  $U_j(x_j)$  der Gesellschaftsmitglieder  $j$  (mit  $j = 1, \dots, n$ ) geschrieben:  $W_i(x) = W_i(U_1, U_2, \dots, U_n)$ . Das sog. Paretianische Werturteil sichert, daß die soziale Wohlfahrt  $W_i(x) = W_i(U_1, U_2, \dots, U_n)$  *ceteris paribus* immer dann ansteigt, wenn sich das Nutzenniveau eines Individuums  $j$  erhöht, d.h.  $\partial W / \partial U_j > 0$ . Vgl. grundlegend *Bergson* (1938), *Samuelson* (1948), Kap. 8, sowie die Darstellungen bei *Schmidt* (1991) und *Giersch* (1993). Von der sozialen Wohlfahrtsfunktion im Sinne von *Arrow* (1963) unterscheidet sich die *Bergson*-Funktion u.a. dadurch, daß sie ein moralisches Werturteil eines Kollektivmitglieds über seine Gesellschaft ist, während die *Arrowsche* Funktion als eine kollektive Präferenz einer Gesellschaft als solcher definiert ist. In einem  $n$ -Personen-Kollektiv gibt es mithin  $n$  verschiedene (individuelle) *Bergson*-Wohlfahrtsfunktionen, aber nur eine einzige *Arrow*-Wohlfahrtsfunktion; vgl. *Little* (1952), S. 424. *Buchanans* Behauptung, die neoklassische soziale Wohlfahrtsfunktion repräsentiere stets die Bewertung eines externen, außerhalb des Kreises der Betroffenen stehenden Planers, ist daher, soweit sie die *Bergson*-Variante betrifft, eine Fehlinterpretation. Vielmehr ist die *Bergsonsche* Wohlfahrtsfunktion nichts anderes als die Zielfunktion eines „participant-observer“, wie ihn *Buchanan* (1977b), S. 143, als charakteristisch für die Vertragstheorie ansieht.

<sup>54</sup> Vgl. z. B. die *Buchanan*-Interpretationen von *Mueller* (1973), *ders.* (1996), S. 63, von *Buchanans* Schüler *Barbosa* (1974), *ders.* (1978) sowie in jüngerer Zeit von *Neumärker* (1995). Demgegenüber deuten Passagen in *Buchanan/Tullock* (1962), S. 77-79, darauf hin, daß der Schleier der Unsicherheit sich daraus ergibt, daß die Entscheidungsträger die Situationen, in denen sie sich im Zeitablauf wiederfinden werden, nicht präzise vorhersagen können. Die konstitutionelle Unsicherheit ist in dieser Version des Schleiers dann vollkommen, wenn jede Partei die Anwendung der Regel von jeder relativen Position aus mit der gleichen ex-ante-Wahrscheinlichkeit erwartet; vgl. die

fend, so besteht der Unterschied zwischen dem Ansatz von Buchanan und jenem von Rawls nicht in der angeblich fehlenden Ergebnisbezogenheit des Buchananschen Modells, sondern vor allem in dessen entscheidungstheoretischer Charakterisierung der konstitutionellen Entscheidungssituation als einen Zustand des *Risikos* statt der Unsicherheit i.e.S.

Buchanans Gleichwahrscheinlichkeitsannahme bezüglich der unter dem Schleier der Unkenntnis erwarteten gesellschaftlichen Positionen entspricht Laplaces Prinzip des unzureichenden Grundes, nach dem Individuen in Entscheidungssituationen, in denen sie die tatsächlichen Wahrscheinlichkeitsverteilungen über den Zustandsraum nicht kennen, sich so verhalten sollen, als ob alle Umweltzustände mit gleicher Wahrscheinlichkeit einträten. Die Anwendung dieser Entscheidungsregel auf der konstitutionellen Ebene läuft auf eine Maximierung des Erwartungsnutzens eines beliebigen Individuums über alle postkonstitutionellen Positionen hinaus.

Buchanans Wohlfahrtsfunktion weist damit in formaler Hinsicht eine bemerkenswerte Nähe zum modernen Nutzensummen-Utilitarismus Harsanyis auf, obwohl Buchanan seine Vertragstheorie doch gerade vom Utilitarismus stets deutlich abzugrenzen suchte. Harsanyi postuliert in seinem sog. *Gleichwahrscheinlichkeitsmodell moralischer Werturteile*<sup>55</sup>, daß ein Entscheidungsträger, um moralische Werturteile zu treffen, so entscheiden solle, als wisse er nicht, in welcher Position er sich selbst dabei befinde. Handeln die Entscheidungsträger dabei Bayes-rational (agieren sie also, als ob sie eine kardinale von-Neumann-Morgenstern-Nutzenfunktion maximierten) und unterstellen sie gemäß dem Laplace-Prinzip, daß sie jede Betroffenenposition mit gleicher Wahrscheinlichkeit selbst einnehmen könnten, so maximiert jeder von ihnen durch sein individuell rationales Verhalten eine utilitaristische (Benthamsche) Nutzensummen-

---

Formalisierungen dieser Interpretation bei Schweizer (1988), ders. (1990) und Eichberger/Pethig (1993), dies. (1994). Bereits Buchanan (1967), S. 234, hat jedoch darauf hingewiesen, daß die Gleichwahrscheinlichkeits- und die Zeitablaufvariante des Schleiers im Falle vollkommener konstitutioneller Unsicherheit identisch sind, was später von Eichberger/Pethig (1994) formal bewiesen wurde. Die zweite Interpretation des Schleiers der Unsicherheit wird daher hier und im folgenden vernachlässigt.

<sup>55</sup> Vgl. grundlegend Harsanyi (1953) sowie z. B. Harsanyi (1982 [1977]), ders. (1994). Gesamtdarstellungen des Gleichwahrscheinlichkeitsmodells finden sich bei Schmidt (1991), S. 152-169, Jakob (1996), Kliemt/Lahno (1996). Harsanyis Gleichwahrscheinlichkeitsmodell sollte nicht, wie in der Literatur häufig anzutreffen, mit seiner axiomatischen Reformulierung des Utilitarismus verwechselt werden. Mit seinem axiomatischen Ansatz versuchte Harsanyi nachzuweisen, daß die gleichzeitige Akzeptanz des Pareto-Prinzips und der Bayesianischen Entscheidungstheorie zwangsläufig die utilitaristische Moraltheorie impliziert; vgl. dazu grundlegend Harsanyi (1955) sowie z. B. die Darstellung in Harsanyi (1979).

Wohlfahrtsfunktion auf der Basis des von-Neumann-Morgenstern-Nutzenkonzepts.

Trotz der offenbaren Ähnlichkeit zwischen dem Buchananschen Ansatz und dem Gleichwahrscheinlichkeitsmodell Harsanyis führt die Interpretation Dennis C. Muellers zu weit, Buchanans (und Tullocks) Argument unter Verwendung des Schleiers der Unsicherheit impliziere in formaler Hinsicht eine utilitaristische (Benthamsche) Nutzensummen-Wohlfahrtsfunktion.<sup>56</sup> Denn die Buchanansche unterscheidet sich von der utilitaristischen Wohlfahrtsfunktion nicht allein, wie Mueller meint, dadurch, daß ihre Existenz bei Buchanan als empirisches Phänomen behauptet wird, während sie im Utilitarismus den Charakter einer Norm hat. Ein weiterer, von Mueller übersehener Unterschied der Buchananschen im Vergleich zur utilitaristischen Wohlfahrtsfunktion besteht nämlich darin, daß bei Buchanan sich die Rationalentscheider nur mit ihrer *eigenen* Nutzenfunktion in die Position der jeweils anderen Betroffenen versetzen. Der „Universalisierungsgrad“ ist hier - im Gegensatz zum Modell von Harsanyi - nur *partiell*<sup>57</sup>; eines intersubjektiven Nutzenvergleichs bedarf es daher, im Gegensatz zum Utilitarismus, nicht. Nur im Extremfall einer *vollkommenen* Universalisierung - nach der sich alle Individuen nicht nur in die Position, sondern auch in die Nutzenfunktion der einzelnen Positionsinhaber versetzen - fallen Buchanans und Harsanyis Ansatz zusammen und es resultiert die utilitaristische Maximierung des Durchschnittsnutzens aller Gesellschaftsmitglieder.

In einem zentralen Punkt stimmen Buchanans Modell und der utilitaristische Ansatz allerdings immer überein: Anders als die Rawlssche „Theorie der Gerechtigkeit“, kennen weder der Utilitarismus noch Buchanans Vertragstheorie auf der Basis des Schleiers der Unsicherheit einen „Vorrang der Freiheit“. Da der Schleier der Unsicherheit bei Buchanan ein empirisches Phänomen sein soll, das alle konstitutionellen Entscheidungen kennzeichnet, sind Rechte niemals absolut - unabhängig von individuellen Interessen - definiert, sondern stets, wie im klassischen Utilitarismus - nur relativ zum Nutzenkalkül der konstitutionellen Entscheidungsträger. Der größte Teil der Utilitarismus-Kritik, die Rawls in seinem Buch übt, trifft somit gleichermaßen für diesen Teil von Buchanans Vertragstheorie zu.

Gegen das in Buchanans (wie Harsanyis) Wohlfahrtsfunktion vorausgesetzte Laplace-Prinzip ist gelegentlich eingewandt worden, daß das Fehlen objektiver Wahrscheinlichkeiten kein Grund sei für die Annahme, die Ereignisse des Zu-

---

<sup>56</sup> Vgl. Mueller (1998b), S. 183; ähnlich ders. (1996), S. 64. Mueller (1998a) entwickelt auf der Basis des so interpretierten *Buchanan/Tullock*-Schleiers eine utilitaristische Verfassungstheorie.

<sup>57</sup> Zur Unterscheidung eines partiellen und eines vollkommenen „Universalisierungsgrads“ siehe ausführlich unten Abschnitt B.



standsraumes träten mit gleicher Wahrscheinlichkeit ein; jede andere Wahrscheinlichkeitsschätzung wäre gleichermaßen möglich. Dies ist grundsätzlich richtig. Jede andere Annahme einer Wahrscheinlichkeitsverteilung wäre ebenso denkbar. Aber im Vergleich zur Rawlsschen Maximin-Regel weist sie einen deutlichen Vorteil auf: Während die Gleichwahrscheinlichkeitsannahme willkürlich und damit *möglicherweise* falsch ist, fordert das Maximin-Prinzip dazu auf, so zu entscheiden, als ob man mit einer Wahrscheinlichkeit von eins zu erwarten hätte, der „Pechvogel der Nation“ zu sein.<sup>58</sup> Diese Annahme jedoch ist *in jedem Fall* falsch, weil nicht alle Bürger gleichzeitig die Position des am meisten Benachteiligten einnehmen können.<sup>59</sup>

Aus subjektivistischer Sicht ist es nicht notwendig anzunehmen, daß die Wahrscheinlichkeiten, welche das Individuum unter dem Schleier der Unsicherheit kalkuliert, für jede erwartete Gesellschaftsposition gleich sein werden.<sup>60</sup> Anders als in Harsanyis Gleichwahrscheinlichkeitsmodell, führt Buchanan die Gleichwahrscheinlichkeitsannahme nicht als moralisches Gleichbehandlungspostulat ein, sondern - wie die gesamte Konstruktion seines Schleiers der Unsicherheit - als angeblich empirisches Phänomen. Wo Buchanan die Laplace-Annahme überhaupt begründet, beschränkt er sich auf Plausibilitätsüberlegungen. Gleichwohl könnte auch hier ein verstecktes inhaltliches Gerechtigkeitsurteil der eigentliche Grund für die Buchanansche Modellprämisse gleicher Wahrscheinlichkeiten aller Gesellschaftspositionen sein. Nicht zufällig dürfte diese Annahme auch als Metapher für die liberalistische Gleichbehandlungsnorm interpretierbar sein, die, in der Fassung Brennans und Buchanans, verlangt, daß „alle Personen als moralisch gleichwertig angesehen werden, als Subjekte, die alle gleichermaßen in der Lage sind, die jeweiligen Optionen bewerten zu können.“<sup>61</sup>

Als Ergebnis der Prüfung zentraler Argumente für eine Anwendung des Maximin- und des Laplace-Prinzips im Urzustand läßt sich somit festhalten, daß es eine Sache ist, das Gedankenexperiment unter dem Schleier der Unkenntnis zu

---

<sup>58</sup> Vgl. Harsanyi (1976 [1975], S.41 und 47.

<sup>59</sup> Vgl. Pettit (1974), S. 316. - Auch ein anderes Argument gegen das Laplacesche Prinzip - die fehlende Invarianz gegen Spaltenverdoppelungen (vgl. Milnor 1964 [1951], S. 132 f.; Ellsworth 1978, S. 33) - läuft im Zusammenhang mit dem Schleier der Unkenntnis ins Leere, da mit den post-konstitutionellen Positionen die Menge der „Spalten“ (Umweltzustände) feststeht, so daß das Argument auf die Situation im Urzustand nicht anwendbar ist; vgl. dazu auch Hinsch (1997a), S. 78-80. Überdies genügt auch nur Rawls' Differenzprinzip dem Axiom der Spaltenverdoppelung, nicht aber sein (hier nicht behandeltes) „lexikographisches Differenzprinzip“; vgl. dazu Ellsworth (1978), S. 34 f.

<sup>60</sup> Vgl. Binmore (1994), S. 328.

<sup>61</sup> Brennan/Buchanan (1993 [1985]), S. 28.



akzeptieren; eine andere Sache ist es jedoch, eine konkrete inhaltliche Gerechtigkeitsposition zu vertreten. Aus der Akzeptanz der hypothetischen Entscheidungssituation unter einem Schleier der Unkenntnis als legitimitätsstiftende Ausgangssituation folgt nur, daß man sich in die Lage der übrigen Betroffenen versetzt; eine konkrete, als einzige rational erscheinende Gewichtungsgesamtregel für diese Präferenzen folgt aus der Akzeptanz dieser Situation nicht.

Die Annahme gleicher Wahrscheinlichkeiten mag, wie die vorangegangene Diskussion zeigt, unter den Bedingungen unsicherer Positionserwartungen bei der konstitutionellen Entscheidung plausibler sein als jene der Maximinposition; zwingend sind indes beide nicht. Im Modell der konstitutionellen Unkenntnis ist die Deduktion einer bestimmten Gerechtigkeitsauffassung von der nicht objektivierbaren Auffassung abhängig, welches Entscheidungsprinzip unter unsicheren Erwartungen als „rational“ anzusehen ist. Ob man sich einer der beiden diskutierten Rationalitätsauffassungen anschließt und, wenn ja, welcher von ihnen, ist unter Umständen nicht unabhängig vom jeweils gewünschten Rechtfertigungsergebnis. Die Maximin- bzw. die Laplace-Wohlfahrtsfunktion folgen nur nach Einführung zusätzlicher, über das Universalisierungsprinzip hinausreichender Gerechtigkeitswerturteile aus der Konjunktion mit den Prämissen der vertragstheoretischen Entscheidungssituation.

Durch Einführung anderer „Annahmen“ über das Rationalverhalten der Individuen unter den Unsicherheitsbedingungen des Urzustandes ließen sich andere Wohlfahrtsfunktionen gewinnen. Formal erscheint es nicht möglich, irgendeinen der unendlich vielen denkbaren Verläufe der Bergsonschen Wohlfahrtsfunktion *a priori* als Ergebnis des konstitutionellen Kalküls unter dem Schleier der Unkenntnis auszuschließen. Um einen bestimmten Verlauf der Wohlfahrtsfunktion zu generieren, wären lediglich die Annahmen über die Wahrscheinlichkeitsschätzungen der Beteiligten in der konstitutionellen Entscheidungssituation entsprechend zu variieren. Das einzige Problem bestünde hier darin, die jeweiligen Wahrscheinlichkeitsannahmen, die nötig sind, um ein bestimmtes Rechtfertigungsergebnis herzuleiten, im Kontext des Urzustandsmodells plausibel zu machen.<sup>62</sup>

---

<sup>62</sup> Reale Individuen verhielten sich in Experimenten, welche die Bedingungen eines Schleiers der Unkenntnis approximierten, weder nach Maßgabe des Maximin- noch des Laplace-Prinzips. Vielmehr wählten sie ein „intuitionistisches“ Prinzip bzw. eine „gemischte Gerechtigkeitskonzeption“ als eine Art Kompromiß zwischen beiden Rationalitätsregeln, das nicht nur die Situation des armseligsten Individuums, sondern auch die potentiellen erwarteten Gewinne für den Rest der Gesellschaft berücksichtigt. Vgl. *Frohlich/Opppenheimer/Eavey* (1987a,b); *Frohlich/Opppenheimer* (1992), S. 30 und v.a. S. 58-66.

## B. Konsensinduzierung: Der Universalisierungsgrad

Aber nehmen wir einmal - der vorangegangenen Analyse zum Trotz - an, das Gedankenexperiment mit dem Schleier der Unkenntnis impliziere notwendig die Befolgung eines einzigen Rationalitätsprinzips - des Maximin- oder des Laplace-Prinzips - durch alle Beteiligten im Urzustand. Das vertragstheoretische Argument hätte in diesem Fall immer noch nachzuweisen, daß alle konstitutionellen Entscheidungsträger die *gleiche* Maximin- oder die *gleiche* Laplace-Wohlfahrtsfunktion maximieren. Nur in diesem Fall wäre garantiert, daß alle Individuen sich auf den gleichen Gesellschaftsvertrag stillschweigend einigen.

*Einstimmigkeit* aller Modellakteure muß garantiert werden, da nur bei Einhaltung dieser Bedingung das vertragstheoretische Legitimitätskriterium der hypothetischen Zustimmung aller Beteiligten erfüllt ist. Gelänge es nicht, Einstimmigkeit herzustellen, weil jedes Individuum eine andere Wohlfahrtsfunktion maximiert, so wäre die Ergebnisoffenheit der strikt paretianischen Vertragstheorie nicht überwunden. Bestünde nämlich auch unter dem Schleier der Unkenntnis die Möglichkeit, daß die Individuen unterschiedliche Situationen als *Optima optimorum* ansehen, so verfehlte auch die schwache Vertragstheorie das Ziel, konkrete Politikempfehlungen zu geben.

*Stillschweigend* - d.h. ohne jede Verhandlung - muß der Konsens zustande kommen, da die genannten Theorien andernfalls festzulegen hätten, welche Verhandlungsergebnisse unter welchen Bedingungen - Gruppengrößen, Verhandlungsstärken, Verhandlungsgeschick der Beteiligten etc. - zu erwarten wären. Dazu aber bedürfte es einer Verhandlungstheorie, die keiner der beiden genannten Ansätze aufbietet. Der Schleier der Unkenntnis hat damit zu garantieren, daß beide Theorien *Choice versions* der Vertragstheorie sind, d.h. solche Kontraktansätze, in denen die Individuen ohne Verhandlungen identische Wahlentscheidungen treffen. Im Mittelpunkt dieser *Choice versions* steht nicht eine Verhandlungssituation, sondern die Wahlhandlung eines einzigen beliebigen Entscheidungsträgers, der als repräsentativ für alle anderen angesehen wird.<sup>63</sup> Binmore<sup>64</sup> hat zudem darauf hingewiesen, daß konstitutionelle Zwei-Personen-Modelle, mit denen Rawls und Buchanan mit Vorliebe arbeiten, nur

---

<sup>63</sup> Solche Kontraktansätze, die Verhandlungen der Entscheider im Urzustand unterstellen, werden demgegenüber *Bargaining versions* der Vertragstheorie genannt; vgl. in bezug auf Rawls' Theorie Alexander (1974), S. 604; Hampton (1980). Kukathas/Pettit (1990), S. 33 f., sprechen bei gleicher Bedeutung von „non-interactive“ bzw. „interactive contracts“.

<sup>64</sup> Vgl. Binmore (1994), S. 79.

dann ein Modell für moderne Massengesellschaften sein könnten, wenn gezeigt werden könne, daß sich die Beteiligten stillschweigend einigen werden.

Die Grundannahme der mit dem Schleier der Unkenntnis arbeitenden Vertragstheorie steht daher nicht nur im Gegensatz zur Habermasschen Diskursethik mit ihrem Rekurs auf einen „diskursiven Konsens“ der Entscheidungsträger, sondern auch zu neueren ökonomischen Arbeiten, die den Einfluß der Kommunikation auf das Kooperationsverhalten von Individuen in sozialen Konfliktsituationen untersuchen.<sup>65</sup>

Das Mittel zur Förderung eines konstitutionellen Konsenses ist in beiden Varianten der Vertragstheorie das der Erhöhung des Universalisierungsgrades gegenüber der Situation sicherer Erwartungen. Versetzen sich die Individuen in die Rolle des jeweils anderen und berücksichtigen sie in ihrem Kalkül, welche Interessen sie selbst in der Position des Gegenübers hätten, wird - so die Grundidee - ein Konsens leichter erzielbar sein. Die Theorievarianten von Rawls und Buchanan unterscheiden sich gleichwohl im Ausmaß jenes Rollentausches, für den der Schleier der Unkenntnis steht - kurz: in ihrem „Universalisierungsgrad“. Je stärker das betrachtete Individuum fremde Interessen in der eigenen Nutzenkalkulation berücksichtigt, desto höher ist der Universalisierungsgrad - und desto „dicker“ ist der Schleier der Unkenntnis. Als *partiell* werde ich im folgenden den Universalisierungsgrad bezeichnen, wenn sich die Individuen mit ihrer *eigenen* Nutzenfunktion in die postkonstitutionellen Gesellschaftspositionen versetzen; von einem *vollkommenen* Universalisierungsgrad soll demgegenüber dann die Rede sein, wenn die Parteien im Urzustand gedanklich nicht nur in die einzelnen Positionen der postkonstitutionellen Gesellschaft schlüpfen, sondern auch in die (*fremde*) Nutzenfunktion des jeweiligen Positionsinhabers.<sup>66</sup>

### I. Partielle Universalisierung: Der Tausch von Positionen

Eine Regel ist universalisierbar, wenn alle Betroffenen auch dann noch zu ihr stehen können, wenn sie sich hypothetisch in die Lage aller mitbetroffenen Individuen versetzen. Partielle Universalisierung verlangt von einem Entscheider, sich mit seinen eigenen Vorlieben, Wertvorstellungen und Idealen - kurz: mit seiner *eigenen* Nutzenfunktion - in die Position eines jeden anderen Betroffenen zu versetzen und zu ermitteln, wie man *sich selbst* in dieser Situation

<sup>65</sup> Vgl. insbesondere Bohnet (1997) mit weiteren Nachweisen.

<sup>66</sup> Die hier diskutierte partielle und vollkommene Universalisierung entspricht der Universalisierung zweiter und dritter Stufe bei Mackie (1981 [1977]), Kap. 4 und S. 192-196.

fühlen würde. Vereinfachend formuliert, bedeutet die partielle Universalisierung ein Verhalten nach der in der Bibel „*Goldene Regel*“ genannten Norm, die im deutschen Sprachraum besser bekannt ist als Reimspruch: „Was du nicht willst, das man dir tu“, das füg’ auch keinem andern zu.“<sup>67</sup> Die Goldene Regel ist auch die Kurzfassung, auf die Hobbes die „Gesetze der Natur“ in seinem Naturzustand bringt.<sup>68</sup>

Eine Lösung, diesen Grundsatz zu operationalisieren, ist der Buchanansche Schleier der Unsicherheit, auch wenn diese Prämisse - wie geschildert - nicht-normativ begründet wird. Nutzen ist nach Buchanan eine allein subjektiv quantifizierbare Größe; Wohlstandsniveaus sind folglich nur für den einzelnen Entscheidungsträger (ordinal oder kardinal) meßbar.<sup>69</sup> Eine vollkommene Identifikation mit der Rolle des anderen und damit ein Hineinfühlen in dessen Präferenzen, wie es in der Bergsonschen Wohlfahrtsökonomik und, wie wir noch sehen werden, unter dem Rawlsschen Schleier des Nichtwissens unterstellt wird, wäre demgegenüber eine Unmöglichkeit, setzte sie doch die Allwissenheit des Entscheiders voraus. Die in der konstitutionellen Vertragssituation bestehende Unsicherheit der Entscheidungsträger über ihre eigene postkonstitutionelle Situation veranlaßt nach Buchanan die Individuen, sich allein mit ihrer *eigenen* Nutzenfunktion in die Situation einer jeden einzelnen postkonstitutionellen Position zu versetzen und nach Maßgabe der eigenen Präferenzen einen Nutzenwert für die Position einer Putzfrau, eines Selbständigen oder eines Generaldirektors zu ermitteln.

Diese Modellierung hat einen ganz entscheidenden Vorteil: Einer mühsamen und aufwendigen, wahrscheinlich sogar praktisch unmöglichen Erhebung der tatsächlichen Nutzenfunktionen der postkonstitutionellen Positionsinhaber bedarf es nicht, um aus der Menge der wählbaren Pareto-Optima ein *Optimum optimorum* zu bestimmen. Die aus der Wohlfahrtsökonomik (und der Rawlschen Theorie) bekannte Allwissenheitsannahme der Entscheidungsträger<sup>70</sup> wird deutlich abgeschwächt: Um den Rollentausch unter dem Schleier der Unsicherheit vorzunehmen, müssen die Individuen nicht mehr vollkommen über die Nutzenfunktionen aller Mitglieder der postkonstitutionellen Gesellschaft infor-

---

<sup>67</sup> Die biblischen Stellen sind z. B. Matthäus 7,12 („Alles, was ihr von anderen erwartet, das tut auch ihnen!“) oder Lukas 6,31 („Was ihr von anderen erwartet, das tut ebenso auch ihnen.“). Zur Goldenen Regel allgemein vgl. *Singer* (1963); *Reiner* (1974); *Mackie* (1981 [1977]), Kap. 4; *Brülisauer* (1980); *Wimmer* (1980), S. 254-295.

<sup>68</sup> Vgl. *Hobbes* (1976 [1651]), Kap. 15, S. 120 f.

<sup>69</sup> Vgl. *Buchanan* (1960 [1959]), S. 108.

<sup>70</sup> Schon *Buchanan* (1960 [1959]), S. 107 f., wendet sich gegen die in der Wohlfahrtsökonomik übliche implizite Annahme der Allwissenheit des gesellschaftlichen Entscheiders, der alle Nutzenfunktionen der Individuen kennt und daher in der Sozialen Wohlfahrtsfunktion berücksichtigen kann.

miert sein, sondern es genügt - was immer noch eine sehr weitreichende (und unrealistische) Annahme ist - allein die vollkommene Information über die Art und Beschaffenheit der postkonstitutionellen Positionen. In jede einzelne dieser Gesellschaftspositionen versetzt sich jeder Entscheidungsträger unter dem Schleier der Unsicherheit nämlich mit seiner *eigenen* Nutzenfunktion; und nur diese *eigenen* (kardinalen) Nutzenwerte werden in der sozialen Wohlfahrtsfunktion des betreffenden Individuums aufsummiert. Ein intersubjektiver Nutzenvergleich ist folglich nicht erforderlich; es bedarf lediglich eines *intrapersonellen* Nutzenkalküls.<sup>71</sup>

Eine nähere Betrachtung zeigt indes, daß diese Modellierung nur scheinbar von Vorteil ist. Im Falle einer lediglich partiellen Universalisierung wissen die betrachteten Entscheidungsträger nämlich trotz ihres hypothetischen Positionsaustausches stets genau, wer sie sind und welche Nutzenfunktion sie haben.<sup>72</sup> Der Grad des konstitutionellen Nichtwissens ist hier also weniger total als bei Rawls; er beschränkt sich nur auf einen Teilaspekt des Rawlsschen Nichtwissensschleiers: die mangelnde Voraussicht. Anders als bei Rawls kennen die Individuen stets ihr Geschlecht, sie wissen, ob sie groß oder klein, schwarz oder weiß sind oder auch welcher Generation sie angehören. „Diese subjektiven Elemente“, so betonen Brennan und Buchanan, „sind in einem mit der Kategorie der Unsicherheit arbeitenden Kalkül immer präsent.“<sup>73</sup> Ein wesentlicher Teil des Wissens über eigene Partialinteressen verbleibt den Individuen daher. Sie werden also auch unter dem Unsicherheitsschleier nicht völlig identische Interessen haben und folglich im Regelfall unterschiedliche konstitutionelle Wohlfahrtsfunktionen maximieren.

---

<sup>71</sup> „We are *not* concerned with *interpersonal comparability of utilities* among separate members of the community at the post-constitutional stage ... The relevant comparison for our purposes is that which takes place, within the calculus of the single person at the moment of constitutional choice among alternative redistribution schemes. In this choice, the individual must somehow assign utilities or utility weights to himself in each one of the set of possible positions in the postconstitutional distribution ... This assignment will presumably involve some predictions as to the utilities that will be perceived post-constitutionally, but that which the individual maximizes in making the selection among institutions is *his own utility* at the moment of choice.“ *Buchanan* (1986), S. 161 (Hervorhebungen von mir; C. M.). Vgl. auch *Buchanan/Tullock* (1962), S. 95.

<sup>72</sup> „In the Rawlsian limit, the person does not know who he will be in the settings where the chosen rule is to be operative. In the somewhat less rarified Buchanan-Tullock idealization, the *person* may *himself be identified*, but there is such uncertainty about the effects of rules on separate individual positions that particular interests cannot be related to particular rules.“ *Buchanan* (1986), S. 243 (Hervorhebungen von mir; C. M.).

<sup>73</sup> *Brennan/Buchanan* (1993 [1985]), S. 41. Wie sich die Kenntnis dieser „subjektiven Elemente“ auf die Entscheidung im Rawlsschen Urzustand auswirken kann, untersuchen *Buchanan/Faith* (1980).

Dies hat zum einen den Nachteil, daß verbleibende Interessenunterschiede der Erreichung eines Konsenses im Wege stehen können. Ist der Schleier der Unsicherheit, wie von Buchanan behauptet, ein empirisches Phänomen, so werden - von altruistischen Präferenzen abgesehen - Weiße nicht die Interessen von Schwarzen, Männer nicht die Interessen von Frauen, Europäer nicht die Interessen von Afrikanern berücksichtigen. Aber selbst dann, wenn alle Beteiligten die Interessen aller anderen gleichwertig in ihren konstitutionellen Wohlfahrtsfunktionen mit berücksichtigen, können sie sich in den Nutzenwerten, die sie in den einzelnen Positionen beziehen würden, unterscheiden. Das jeweils individuell angestrebte Pareto-Optimum kann sich daher von jenen der anderen Individuen unterscheiden, so daß es im Extremfall so viele *Optima optimorum* gibt wie Gesellschaftsmitglieder. Diese unter dem Schleier der Unsicherheit individuell erstrebten Pareto-Optima mögen zwar näher bei einander liegen als die ursprünglichen Partialinteressen unter sicheren Erwartungen; zu einem Konsens müssen sie jedoch nicht führen. In § 5 dieser Arbeit wird darüber hinaus gezeigt, daß die Entscheider unter einem Buchananschen Schleier der Unsicherheit eine Einigung bereits dann verfehlen können, wenn sie zwar identische ordinale (Gefangenendilemma-)Präferenzen haben, sich aber in ihren relativen Präferenzintensitäten bei der Bewertung der jeweiligen Ergebnisse, welche die zur Wahl stehenden Regeln erzielen, unterscheiden. Nur in den unrealistischen Fällen, daß alle Individuen entweder identische individuelle Nutzenfunktionen oder zumindest - bei unterschiedlichen individuellen Interessen - identische konstitutionelle Präferenzen haben, ist ein Konsens gewährleistet.

Das Problem der rein paretianischen Vertragstheorie, im Regelfall kein Legitimationsergebnis ausweisen zu können, wird in Buchanans Theorie daher nicht behoben, sondern nur verschoben: Eine ohne den Schleier der Unkenntnis operierende starke Vertragstheorie determiniert im Regelfall kein universales *Optimum optimorum*, weil die Individuen ihre eigenen Partialinteressen genau kennen und daher nicht voraussehen ist, ob und, wenn ja, auf welches Pareto-Optimum sie sich einigen werden. Wird demgegenüber in die paretianische Vertragstheorie mit dem Schleier der Unsicherheit eine (wie auch immer begründete) Situationsannahme eingeführt, welche die Beteiligten zu einer partiellen Universalisierung ihres Verhaltens veranlaßt, so kann jedes Individuum nach Maßgabe seiner konstitutionellen Präferenzen zwar ein *Optimum optimorum* ermitteln, das - gemessen in seinen eigenen Nutzeneinheiten - die Interessen der anderen Kollektivmitglieder mitberücksichtigt; die konstitutionellen Interessen der Individuen - d.h. ihre Wohlfahrtsfunktionen - können sich aber ihrerseits unterscheiden, so daß man einer Einigung der Individuen nicht notwendig näher gekommen ist. Das Problem, eine Einigung vorherzusagen, deren Gegenstand herbeizuführen legitimes Ziel der praktischen Politik sein soll, hat sich lediglich von der Ebene der Partial- auf jene der konstitutionellen Interes-



sen verschoben. Auf der Suche nach dem aus Sicht *aller* Individuen akzeptablen Gesellschaftsvertrag ist man damit nicht weiter gekommen.

Ein zweiter wesentlicher Nachteil einer lediglich partiellen Universalisierung ist, daß auch diskriminierende Gesellschaftsverträge denkbar sind. Eine Gesellschaft, die in bezug auf eines ihrer Merkmale homogen ist, könnte sich etwa auf Regeln einigen, die alle andersartigen Individuen diskriminieren, die - aus welchen Gründen auch immer - an der Abstimmung nicht teilnehmen. Ein realistisches Beispiel für diesen Fall wäre ein Gesellschaftsvertrag zu Lasten zukünftiger Generationen. Brennan und Buchanan<sup>74</sup> unterstellen in ihrer Besteuerungstheorie, daß sich rationale Individuen unter dem Schleier der Unsicherheit zur Beschränkung des fiskalischen Hungers eines steueraufkommensmaximierenden „Leviathan“-Staates durch die Einführung eines Systems spezieller Verbrauchssteuern auf Güter mit hoher Preiselastizität schützen würden, da man dem hierdurch ausgeübten staatlichen Zwang durch Nichtkonsumtion der besteuerten Güter ausweichen könne. Diese „Finanzverfassung“ für den Leviathan ist indes nur dann optimal, wenn von Persönlichkeitsunterschieden oder zumindest dem Wissen hierüber auf der konstitutionellen Stufe abstrahiert wird.

Nehmen wir an, die Parteien im Urzustand kennten alle ihr Alter und wüßten, daß in dreißig Jahren keiner von ihnen mehr erwerbstätig sein wird.<sup>75</sup> Eine aus der Sicht strikt eigennütziger Entscheider optimale Finanzverfassung, welche die betrachteten Individuen weit besser stellen würde als die von Brennan und Buchanan beschriebene Lösung, könnte dann aus den folgenden beiden Bestimmungen bestehen: Erstens, die gesamte Staatslast wird mit Krediten finanziert, die in frühestens dreißig Jahren zurückzuzahlen sind (es sei unterstellt, daß solche Kredite tatsächlich auf dem Kapitalmarkt angeboten werden); und zweitens wird die Besteuerung von Rentenbezügen verfassungsmäßig verboten. Die gesamten Lasten des staatlichen Gegenwartskonsums würden hiermit auf spätere Generationen verschoben, was aus der Sicht der gegenwärtigen Erwerbstätigengeneration - und nur auf diese kommt es im Falle partieller Universalisierung an, da keine moralische Verpflichtung besteht, auch die Interessen späterer Generationen zu wahren - eindeutig als paretoverbessernd zu werten wäre.<sup>76</sup>

<sup>74</sup> Vgl. Brennan/Buchanan (1988 [1980]).

<sup>75</sup> Diese Annahme hebt die zentrale Modellprämisse des *Ricardo-Barro*-Äquivalenztheorems (vgl. Barro 1974) auf, nach der die Wirtschaftssubjekte einen unendlichen Zeithorizont haben. Im hier entwickelten Beispiel ist eine Lastenverschiebung auf zukünftige Generationen also möglich.

<sup>76</sup> Buchanan/Congleton (1998), S. 99, räumen ein, daß selbst die Verwendung der Einstimmigkeitsregel bei der Entscheidung zwischen Besteuerung und Verschuldung eine Diskriminierung späterer Generationen nicht verhindern kann. Genau ein solches Ergebnis versucht Rawls jedoch mit seinem Schleier des Nichtwissens, unter dem nie-



Daß derartige, die persönlichen Unterschiede von Menschen ausbeutenden Gesellschaftsverträge auch in Buchanan's Ansatz nicht vorkommen, ist nur damit zu erklären, daß er in seiner Theorie des hypothetischen Gesellschaftsvertrags notwendig fiktive, eigenschaftslose Individuen mit einem unendlichen Zeithorizont als Entscheidungsträger unterstellen muß, anstatt von realen Menschen aus Fleisch und Blut auszugehen.<sup>77</sup> Da die unterstellten *Homines oeconomici* einander gleichen wie ein Haar dem anderen, ist in der Analyse per definitionem ein wesentlicher Teil jener individuellen Unterschiede ausgeblendet, die realiter einer fairen und konsensualen Entscheidungsfindung im Wege stehen könnten. Den oben genannten Gesellschaftsvertrag zu Lasten späterer Generationen werden die Beteiligten allein schon deshalb nicht abschließen, weil die üblichen *Homo-oeconomicus*-Annahmen nicht spezifizieren, welcher Generation die Entscheider angehören. Die fehlende Spezifizierung von Eigenschaften der betrachteten ökonomischen Rationalentscheider, die im Beispiel auf die implizite Annahme eines unendlichen Planungshorizonts der Parteien hinausläuft, ist gleichbedeutend mit der Annahme eines dickeren Schleiers als behauptet: er umfaßt im Ergebnis nicht nur die Unsicherheit der Parteien darüber, wie sich die zur Diskussion stehende Regel auf ihren eigenen Nettowohlstand auswirken wird; vielmehr handeln die Individuen *de facto* so, als ob sie - wie unter dem Rawlsschen Schleier des Nichtwissens - auch ihre persönlichen Unterschiede nicht kennen bzw. alle die gleiche Nutzenfunktion hätten. Trotz der unter dem Schleier der Unsicherheit prinzipiell möglichen Kenntnis der jeweils eigenen Partialinteressen haben die Individuen insoweit keinen Anlaß zur Diskriminierung.<sup>78</sup>

---

mand weiß, welcher Generation er angehört, auszuschließen. Wenn die Menschen wüßten, „daß sie Zeitgenossen sind (wenn man die Gegenwart als Eintrittszeit nimmt), können sie ihre Generation bevorzugen, indem sie sich weigern, irgend etwas für ihre Nachfahren aufzusparen; sie stellen einfach zusammen den Grundsatz auf, daß niemand die Pflicht hat, für die Nachwelt etwas zu tun. ... Um also mit der Frage der Gerechtigkeit zwischen den Generationen zurechtzukommen, ändere ich die Motivationsannahme ab und füge eine weitere Einschränkung hinzu ... Damit kann keine Generation mehr Grundsätze formulieren, die besonders auf ihren eigenen Vorteil gerichtet sind ... Was auch die Stellung eines Menschen in der Zeit sein mag, er ist stets gezwungen, für alle Menschen zu entscheiden.“ (*Rawls* 1994 [1971], S. 163)

<sup>77</sup> Daß das *Buchanan'sche* Gedankenexperiment unter dem Schleier der Unsicherheit nur dann funktioniert, wenn es, wie das *Rawlssche*, von hypothetischen Rationalentscheidern statt von tatsächlichen Menschen ausgeht, konstatiert, mit anderer Begründung, auch *Goldberg* (1974), S. 571.

<sup>78</sup> Es ist bezeichnend, daß *Buchanan/Congleton* (1998), S. 74 und Kap. 9, in ihrer Behandlung des Problems der intergenerationellen Gerechtigkeit, statt auf die *Buchanan'sche* Annahme, auf den *Rawlsschen* Schleier des *Nichtwissens* bzw. direkt auf die Norm intergenerationaler Gleichbehandlung zurückgreifen, da sich mit dem Schleier der *Unsicherheit* eine faire Behandlung späterer Generationen nicht begründen ließe.

Für Buchanans Ansatz, dessen hypothetischer Vertragskalkül nicht selbst legitimitätsstiftend sein, sondern nur Vorschläge für eine spätere tatsächliche verfassungsgebende Versammlung vorselektieren soll, ist diese Einschränkung weitreichend: Der Kalkül der eigenschaftslosen *Homines oeconomici* unter dem Schleier der Unsicherheit sondert eine Vielzahl von zwar diskriminierenden Vorschlägen aus, auf die sich die Teilnehmer einer realen verfassungsgebenden Versammlung aber nichtdestoweniger einstimmig einigen könnten. Das Spektrum der für die realen Entscheidungsträger verfügbaren Alternativen wird hierdurch über Gebühr eingeschränkt.

## II. Vollkommene Universalisierung: Der Tausch von Präferenzen

Ein häufig geäußelter Einwand gegen das in der Goldenen Regel zum Ausdruck kommende Postulat nach partieller Universalisierung ist, mit welchem Recht es dazu auffordert, andere auf dieselbe Weise zu behandeln, wie man selbst behandelt werden möchte. Denn unproblematisch wäre ein solches Prinzip nur dann, wenn man bei allen Menschen die gleichen Wünsche, Neigungen und Bedürfnisse voraussetzen könnte wie bei sich selbst. Ist dies indes nicht der Fall, kann der Positionstausch zu absurden Konsequenzen führen, indem er Verhaltensweisen zu fordern scheint, die den Präferenzen der anderen widersprechen. Ein Gastgeber etwa, der am liebsten Bier trinkt, müßte danach selbst Weinliebhabern ausschließlich Bier servieren; ein Masochist müßte sich sittlich verpflichtet fühlen, andere zu quälen; oder jemand, der sich aus Stolz niemals helfen lassen würde, dürfte selbst auch anderen nicht beistehen.<sup>79</sup>

Wegen derart unsinniger Konsequenzen der Goldenen Regel in ihrer überkommenen Formulierung findet sich in der Ethik häufig die Forderung, nur solchen Normen das Prädikat „moralisch“ zu reservieren, die einen weitergehenden Universalisierungstest bestehen, den George Bernard Shaw pointiert auf den Nenner brachte: „Don't do unto others as you would have them do unto you; their tastes might be different.“<sup>80</sup> In positiver Formulierung verlangt die so präzisierte Goldene Regel etwa: „Wie du willst, daß man deine Bedürfnisse und Interessen berücksichtigt, so berücksichtige auch du die Bedürfnisse und Interessen der anderen.“<sup>81</sup> Verlangt wird also von einem moralisch Handelnden die Einfühlung in die Position des anderen nicht nur nach Maßgabe der eigenen Präferenzen, sondern nach Maßgabe der persönlichen Vorlieben, Wertvorstel-

<sup>79</sup> Vgl. zu diesen Beispielen Jakob (1996), S. 98; Wimmer (1980), S. 259; Höffe (1981), S. 66.

<sup>80</sup> Zitiert nach Wimmer (1980), S. 393.

<sup>81</sup> Höffe (1981), S. 66.

lungen und Ideale des anderen selbst. Der gedanklich vollzogene Rollentausch ist hier perfekt, da man sich nicht nur in die gesellschaftliche Position des anderen versetzt, sondern sogar „in dessen Seele schlüpft“ und dessen Präferenzfunktion als die eigene annimmt.

Ein Argument, das von einer solchen vollkommenen Verhaltensuniversalisierung ausgeht, ist die Rawlssche „Theorie der Gerechtigkeit“.<sup>82</sup> Unter dem Unwissenheitsschleier kennen zwar alle Entscheidungsträger die möglichen Gesellschaftspositionen und (mittels Grundgüterindizes gemessenen) Präferenzen, die mit diesen Positionen verbunden sind; die Individuen können jedoch nicht zuordnen, welche dieser Positionen und Präferenzen ihre eigenen oder jene der anderen Individuen sind. Unter dem Schleier des Nichtwissens kennt kein Entscheider „seine Vorstellung vom Guten, die Einzelheiten seiner Psyche wie seine Einstellung zum Risiko oder seine Neigung zu Optimismus oder Pessimismus“<sup>83</sup> - kurz: niemand kennt seine eigenen Handlungsinteressen. Mangels Kenntnis der eigenen partialen Präferenzen, kann ein empathisches Einfühlen in die Position eines beliebigen Mitglieds der postkonstitutionellen Gesellschaft nur darin bestehen, dessen Präferenzen anzunehmen. Die absurden Konsequenzen der Goldenen Regel sind damit ausgeschlossen. Nach dem Maximinprinzip entscheiden alle Beteiligten nur noch nach Maßgabe der *tatsächlichen* (und allen bekannten) Präferenzen des Armseligsten in der Gesellschaft. Eine Entscheidung aber, bei der alle gemäß der gleichen Präferenzfunktion entscheiden, kann nicht anders als einstimmig ausfallen. Die vollkommene Universalisierung führt daher im Regelfall in einen monologischen Konsens.<sup>84</sup>

Einstimmigkeit auf der konstitutionellen Ebene ist jedoch nur dann garantiert, wenn erstens alle Individuen vollkommen über alle Präferenzen der Inhaber postkonstitutioneller Gesellschaftspositionen informiert sind, wenngleich die Beteiligten auch nicht wissen, wie sich diese Präferenzen auf sie selbst und die übrigen Abstimmungsteilnehmer verteilen. Zweitens müssen alle Individuen die gleichen intersubjektiven Vergleiche individueller Wohlfahrtsniveaus unternehmen. Rawls versucht mit seiner Darstellung des Urzustands, diese beiden

---

<sup>82</sup> Zur Einordnung der Rawlsschen Theorie als Rechtfertigungsmodell für vollkommen universalisierbare Regeln vgl. Mackie (1981 [1977]), S. 120. - Unabhängig vom vertragstheoretischen Kontext arbeitet etwa auch Harsanyis Gleichwahrscheinlichkeitsmodell mit vollkommen universalisierbaren Regeln. Vgl. z. B. Harsanyi (1958), S. 312, ders. (1977), S. 51, ders. (1982 [1977]), S. 50.

<sup>83</sup> Rawls (1994 [1971]), S. 160.

<sup>84</sup> Das ist nur dann nicht der Fall, wenn das Spiel, das die Beteiligten unter dem Schleier der Unkenntnis spielen, den Charakter eines „Indifferenzspiels“ hat, in dem eine Koordination auf ein einziges Gleichgewicht scheitert; vgl. dazu unten § 5 E. III.

zentralen Probleme des Utilitarismus und der modernen Wohlfahrtstheorie durch zwei Kunstgriffe zu umgehen<sup>85</sup>:

Zum einen verwirft er den herkömmlichen Nutzenbegriff wegen dessen zwangsläufiger Subjektivität als Bezugspunkt einer ethischen Theorie. Rawls' Ansatz will vielmehr „ressourcenorientiert“<sup>86</sup> sein; die Individuen im Urzustand maximieren alle den gleichen, objektiv gegebenen Index von Grundgütern, von denen jedes rationale Individuum auch unter dem Schleier des Nichtwissens - d.h. in Unkenntnis der eigenen Grundgüterausstattung - weiß, daß es mehr statt weniger von ihnen besitzen möchte. Die Grundidee hinter dem Begriff der Grundgüter ist, daß selbst Individuen, die sich fundamental in ihren Werten unterscheiden, gleiche grundlegende Bedürfnisse nach bestimmten Gütern haben, die daher als kleinster gemeinsamer Nenner für interpersonelle Vergleiche individueller Wohlfahrt dienen können.<sup>87</sup> Das aus der klassischen und der modernen Wohlfahrtstheorie bekannte Problem der „Allwissenheit“ des moralischen Beurteilers soll damit entschärft werden. Die Bedürftigkeit eines Individuums ist nach Rawls ablesbar an dem Bestand an Grundgütern, der eine objektiv feststellbare und keine bloß introspektive Größe sein soll wie das konventionelle Nutzenkonzept. Eine vollkommene Information über die individuelle Wohlfahrt aller Betroffenen erscheint somit selbst in großen Massengesellschaften leichter herstellbar.

Zum anderen folgt nach Rawls aus dem Gedankenexperiment im Urzustand - als Substitut für die herkömmliche Bergsonsche Wohlfahrtsfunktion - das Differenzprinzip, zu dessen Erfüllung es ausschließlich *ordinaler* intersubjektiver Vergleiche von Grundgüterausstattungen bedarf, d.h. einer „Niveauevergleichbarkeit“ (*level comparability*)<sup>88</sup> von Wohlstandsmaßen. Um die am wenigsten begünstigte repräsentative Person zu ermitteln, benötigt man danach nur Rangordnungsurteile über das Wohlbefinden der Gesellschaftsmitglieder. Auf eine Feststellung, um wieviel die schlechtestgestellte repräsentative Person, von deren Position allein das Gesellschaftssystem zu beurteilen ist, schlechter steht, kommt es danach nicht an. Die Schwierigkeiten einer objektivierbaren kardinalen Messung individueller Wohlfahrt treten daher nicht auf.

„Rawls' Konzeption der Grundgüter“, rühmt Wolfgang Kersting, „ist ein großartiger philosophischer Einfall“, da sie „dem leidigen Problem der interindividuellen Nutzenvergleiche [entgeht; C. M.], das die Wohlfahrtsökonomie und die utilitaristische Moralphilosophie bei ihrer Bestimmung des zu beför-

<sup>85</sup> Vgl. Rawls (1994 [1971]), S. 112.

<sup>86</sup> Zur Unterscheidung zwischen „ressourcen-“ und „wohlfahrtsorientierten“ Ansätzen vgl. Dworkin (1981).

<sup>87</sup> Vgl. Arneson (1992), S. 217.

<sup>88</sup> Zu diesem Begriff vgl. Sen (1982 [1979]), S. 273.

dernden größeren gesellschaftlichen Nutzens drückt.“<sup>89</sup> Diese Einschätzung ist gleichwohl falsch. Auch die Rawlssche Konzeption der Grundgüter vermag keines der beiden zentralen Probleme - die Messung und den interpersonellen Vergleich individueller Wohlfahrtsniveaus - zu umgehen, geschweige denn zu lösen.

a) Für den *interpersonellen Vergleich* intersubjektiver Wohlfahrtsniveaus ist nichts dadurch gewonnen, daß man sich auf die angeblich objektivierbare Größe „Grundgüterindex“ stützt, anstatt auf eine so introspektive Größe wie „Nutzen“ zurückzugreifen. Die Angabe der Höhe der Güterausstattung eines bestimmten Individuums liefert noch keine Information darüber, welche Wertschätzung diese Person den ihr verfügbaren Gütern zumißt. Auf diese individuellen Wohlfahrtsniveaus kommt es aber letztlich an, um das am schlechtesten gestellte Gesellschaftsmitglied zu ermitteln. Güter werden nicht um ihrer selbst willen erstrebt, sondern wegen des Maßes der Zufriedenheit, das sie stiften.

Verschiedene Individuen, so ist zu erwarten, werden die einzelnen Grundgüter in unterschiedlicher Weise schätzen.<sup>90</sup> Stützt man den Vergleich von Wohlfahrtsniveaus von Menschen allein auf die ihnen verfügbaren Indexeinheiten von Grundgütern, so muß man annehmen, daß die individuelle Wohlfahrt aus einer solchen Indexeinheit für alle Individuen gleich ist. Nur dann, wenn alle Menschen alternative Grundgüterausstattungen gleich bewerten, kann man schließen, daß, wenn der Güterindex eines Individuums *A* niedriger ist als jener einer anderen Person *B*, *A* schlechter gestellt ist als *B*. Ganz in diesem Sinne legt Rawls, der ansonsten über das Bewertungsproblem schweigt, fest, daß alle Individuen den *gleichen* Grundgüterindex maximieren, also alle Individuen die *gleichen* Kombinationen von Grundgütern erstreben. Je höher, mit anderen Worten, der Grundgüterindex ist, desto höher ist das Befriedigungsniveau im für alle Individuen gleichen Ausmaß.

Mit dieser Gleichheitsannahme aber löst Rawls das Problem der intersubjektiven Vergleiche auf eine ganz konventionelle Weise. Ganz wie Rawls verfuhr nämlich bereits der klassische Utilitarismus, indem er das Problem der intersubjektiven Nutzenvergleiche dadurch zu umgehen suchte, daß er die Gleichheit aller individuellen Nutzenfunktionen behauptete.<sup>91</sup> Ähnlich verfährt heute auch Harsanyis auf der Basis des von-Neumann-Morgenstern-Nutzenkonzepts reformulierter Nutzensummen-Utilitarismus, der - als ein nicht-empirisches *A-priori*-Postulat - ein „Ähnlichkeitsaxiom“ enthält, welches fordert, daß Menschen mit völlig identischen Lebenserfahrungen und genetischen Anlagen so anzusehen

<sup>89</sup> Kersting (1993), S. 71.

<sup>90</sup> Vgl. Nagel (1975 [1973]), S. 9; Schwartz (1973), S. 301 ff.; Arneson (1990).

<sup>91</sup> Vgl. Arrow (1973), S. 253 f.

seien, als hätten sie den gleichen (von-Neumann-Morgenstern-)Nutzen.<sup>92</sup> Selbst Rawls' Begründung für die postulierte Gleichheit aller individuellen Präferenzordnungen ähnelt derjenigen Harsanyis: Während letzterer sein Ähnlichkeitsaxiom damit verteidigt, daß dieses Postulat aus der Anerkennung anderer Menschen als mit Bewußtsein und Gefühlen begabt resultiere<sup>93</sup>, leitet Rawls die intersubjektive Gleichheit der Grundgüterindizes aus seiner Annahme ab, daß sich die Individuen gegenseitig als freie und gleiche moralische Personen respektieren.

Der wesentliche (modelltechnische) Vorteil solcher Gleichheitsaxiome ist, daß sie in den individuellen Wohlstandsmaßen einen interpersonell identischen Nullpunkt definieren und sich dadurch *interpersonelle* zu lediglich *intrapersonellen* Vergleichen individueller Wohlstandsniveaus verkürzen.<sup>94</sup> Wenn alle Individuen den gleichen Index von Grundgütern (als Maß für ihre individuellen

---

<sup>92</sup> Vgl. Harsanyi (1977), S. 58: „... if Peter had Paul's biological makeup, had Paul's life history behind him, and were currently subject to Paul's environmental influences, then he would presumably have the *same* personal preferences as Paul has now and would ascribe the *same* utility as Paul does now to each particular situation“ (Hervorhebungen im Original). An anderer Stelle schreibt Harsanyi (1979), S. 301: „By this [= the similarity axiom; C. M.] I mean the principle that, given the basic similarity in human nature (i.e., in the fundamental psychological laws governing human behavior and human attitudes), it is reasonable to assume that different people will show very *similar* psychological reactions to any given objective situation, and will derive much the *same* utility or disutility from it - *once proper allowances have been made for any empirically observed differences* in their biological make-ups, in their social positions, in their educational and cultural backgrounds and, more generally, in their past life histories. In other words, in the absence of clear evidence to the contrary, the presumption must always be that people's behavior and psychological reactions will be similar in similar situations.“ (Hervorhebungen im Original) Zum Ähnlichkeitsaxiom vgl. bereits auch Harsanyi (1977), S. 58; ders. (1982 [1977]), S. 49-52. - De facto ist Harsanyis „Ähnlichkeits-“ ein *Identitätsaxiom* in Form einer Tautologie des folgenden Typs: „Wenn Peter Paul wäre, dann hätte Peter Pauls Nutzen.“ - In der Literatur ist das Ähnlichkeitsaxiom auch unter dem Begriff der „Harsanyi-Doktrin“ bekannt; vgl. Aumann (1987), S. 1 und S. 7; Binmore (1989), S. 90-92; ders. (1994), S. 61 und S. 210 ff.

<sup>93</sup> Vgl. Harsanyi (1979), S. 302: „The similarity postulate is the logical basis, not only for interpersonal comparisons of utility, but also for *assigning other people conscious experiences at all*. From a strictly empirical point of view, a world in which I were the only person with conscious experiences and in which all other people were unfeeling robots would be indistinguishable from our actual world, where, as common sense tells us, *all* normal humans are full self-conscious human beings. If those people who rejected the similarity postulate, and rejected interpersonal utility comparisons based on this postulate, were really consistent then they would have to conclude that they were the only persons with a conscious awareness, and that all other people were mindless automata.“ (Hervorhebungen im Original)

<sup>94</sup> Vgl. Harsanyi (1977), S. 59, in bezug auf sein Ähnlichkeitsaxiom.



Zufriedenheitsniveaus) maximieren, braucht ein Entscheider, der sich in die Seele eines anderen Individuums versetzen möchte, nur zu fragen, wie hoch *sein eigener* Grundgüterindex wäre, bekleidete er die Position des anderen. Wenn also die Individuen unter dem Schleier des Nichtwissens die Welt aus dem Blickwinkel des Schlechtestgestellten zu bewerten versuchen, legen sie alle ihrer Entscheidung dessen *tatsächlichen* Grundgüterindex zugrunde. Unter dieser Bedingung können die Beteiligten im Urzustand einen Konsens, der - von konstitutionellen Koordinationsproblemen abgesehen - nur an einer unterschiedlichen Grundgüter- bzw. Präferenzeinschätzung scheitern könnte, nicht verfehlen.

Wären aber alle individuellen Bewertungsfunktionen identisch, so wäre auch eine *partielle* Universalisierung, wie sie aus dem Buchananschen Schleier der Unsicherheit folgt, kein Problem; gegeben eine bestimmte gesellschaftliche Position, erführen alle Individuen die gleiche Wohlfahrt. Legten alle Betroffenen überdies den gleichen Rationalitätsgrundsatz - das Maximin-, das Laplace- oder irgend ein anderes Entscheidungsprinzip unter unsicheren Erwartungen - zugrunde, so wäre auch bei partieller Universalisierung ein monologischer Konsens garantiert.

b) Die Voraussetzung für die Möglichkeit eines intersubjektiven Vergleichs individueller Wohlfahrtsniveaus auf der Basis eines Gleichheitsaxioms ist es jedoch, daß das zu vergleichende Zufriedenheitsmaß für alle Individuen in objektiv überprüfbarer Weise ermittelbar ist. Hierin lag das zentrale Problem des utilitaristischen Maximanden „Nutzen“, der sich aufgrund seiner Introspektivität einer *objektivierbaren Messung* entzog. Wenn Rawls sich also schon nicht in der Technik des intersubjektiven Vergleichs individueller Wohlfahrtsniveaus - d.h. in der Verwendung von Gleichheitspostulaten bezüglich der individuellen Bewertungsfunktionen - vom Utilitarismus unterscheidet, so muß sein Konzept der Grundgüter, um traditionellen Ansätzen überlegen zu sein, wenigstens das Meßproblem auf zufriedenstellende Weise lösen. Aber auch dies ist nicht der Fall.

Das Problem der Messung individueller Wohlfahrt verlagert sich bei Rawls nur vom Problem der Quantifizierung von Nutzen auf jenes der Messung von Grundgüterindizes. Wäre nur ein einziges Grundgut - etwa Einkommen - zu quantifizieren, so mag man in der Gütermengenmessung noch wenig Probleme sehen, läßt man die aus der Besteuerungstheorie bekannten Abgrenzungsprobleme der Größe Einkommen einmal beiseite. Doch sind die Niveaus nicht aller Grundgüter so leicht zu messen wie das des Einkommens. Wären etwa die „sozialen Grundlagen der Selbstachtung“ das einzige Grundgut, so erschiene das Meßproblem keineswegs mehr so leicht wie jenes von Einkommen. Die intersubjektiv objektivierbare Messung unterschiedlicher Niveaus des Grundguts „soziale Grundlagen der Selbstachtung“ dürfte sich als ein nicht weniger kom-



plexes Problem erweisen als das der Messung der introspektiven Größe Nutzen, deren Verwendung Rawls umgehen möchte.

Bereitet aber die Messung des Niveaus von Grundgütern schon dann Schwierigkeiten, wenn nur ein einziges solcher Güter als Wohlfahrtsmaßstab herangezogen wird, so muß die Aggregation der von einem ganzen Bündel von Grundgütern gestifteten individuellen Wohlfahrt in einem einzigen Grundgüterindex um so problematischer erscheinen. In diesem Fall ergibt sich zusätzlich das nicht-triviale Problem einer Gewichtung der Grundgüter in einer individuellen Zielfunktion. Wieviel mehr muß man etwa einem Individuum von dem Grundgut „die mit öffentlichen Ämtern verbundenen Vorteile“ geben, um es für einen Verlust des Grundgutes „soziale Grundlagen der Selbstachtung“ in einer bestimmten Höhe zu entschädigen?

Fragen wie diese beantwortet Rawls nicht. Die objektive Meßbarkeit aller Grundgüter und ihre Gewichtung in einem einzigen Güterindex wird von ihm nur behauptet, nicht aber belegt. Die herkömmlichen Gerechtigkeitstheorien stießen auf das Problem, daß der Nutzen - da introspektiv - unbekannt war. Ganz analog sind bei Rawls die Grundgüterindizes - wegen der Introspektivität der individuellen Gütergewichtungen - nicht meßbar. Aus diesem Grund bleibt Rawls jenen „objektiven“ Grundgüterindex schuldig, der für alle Menschen qua Zugehörigkeit zur Gruppe der moralischen Personen gleichermaßen gelten soll; vermutlich wird er sich gar nicht ermitteln lassen.<sup>95</sup>

Der Rawlsche Ansatz der Grundgüter ist daher ebenso unoriginell wie untauglich: Unoriginell ist die Grundgüterkonzeption, indem sie das Problem der interpersonellen Vergleichbarkeit letztlich auf die gleiche Weise - nämlich durch ein Gleichheitsaxiom - zu lösen versucht wie der von ihm geschmähte Utilitarismus schon viele Jahrzehnte früher; untauglich ist die Theorie der Grundgüter, da sie das Problem der fehlenden Meßbarkeit des Zufriedenheitsmaßes der Individuen ebensowenig lösen kann wie der Utilitarismus oder die Wohlfahrtsökonomik. Mit seinem Konzept der Grundgüter füllt Rawls lediglich alten Wein in neue Schläuche.

Hebt man indes die unhaltbare und offensichtlich *ad hoc* - d.h. nicht unabhängig vom gewünschten Theorieergebnis - eingeführte Behauptung der Meßbarkeit aller Grundgüterausstattungen auf, so ist ein Konsens im Urzustand genauso wenig garantiert wie unter Buchanans Schleier der Unsicherheit. Die Parteien könnten selbst dann, wenn sie das Maximinprinzip gewissenhaft anzuwenden trachteten, zu unterschiedlichen Nutzenschätzungen gelangen und folglich unterschiedliche Personen für den Armseligsten in ihrer Gesellschaft halten.

---

<sup>95</sup> Vgl. Arrow (1973), S. 254; Wolff (1977), S. 134; Arneson (1990), S. 445, ders. (1992), S. 215; Pogge (1994), S. 134.

Im Extremfall könnte es dann, aus Sicht der Beteiligten im Urzustand, so viele „ärmste“ Individuen geben wie Abstimmungsteilnehmer. Wiederum wäre man dem Ziel, einen von allen Parteien befürworteten Gesellschaftsvertrag zu ermitteln, nicht näher gekommen.

Indem sie die Lösung der Probleme der objektivierbaren Messung und interpersonellen Vergleichbarkeit von Grundgüterausstattungen nur behauptet, aber nicht löst, stößt die Rawlssche Theorie auf erhebliche praktische Anwendungsschwierigkeiten. Wie will man etwa - im Falle der Zugrundelegung unterschiedlicher sozialer Wohlfahrtsfunktionen - das Differenzprinzip gegenüber jenen Personen verteidigen, die es zwar nicht schaffen, von allen Gesellschaftsmitgliedern als die am schlechtesten gestellten Individuen angesehen zu werden, sich aber als solche fühlen? Glaubt jemand tatsächlich, daß das Los, welches er selbst trägt, das schlechtestmögliche ist, wird es schwierig sein, ihn unter Hinweis auf das Gedankenexperiment im Urzustand zu der Einsicht zu bringen, daß sein Schicksal in der gesellschaftlichen Wohlfahrtsfunktion keinerlei Berücksichtigung finden soll, weil andere Individuen aufgrund einer Grundgüterschätzung, die er selbst nicht teilt, statt seiner als die Armseligsten der Gesellschaft angesehen werden.<sup>96</sup>

Rechtfertigungsprobleme dieser Art lassen große Zweifel daran aufkommen, ob das Differenzprinzip tatsächlich ein Gerechtigkeitsgrundatz ist, der das Leben in einer „wohlgeordneten Gesellschaft“ regulieren könnte, die von Rawls ja als eine Gesellschaft definiert wird, „in der alle dieselben obersten Grundsätze von Recht und Gerechtigkeit anerkennen und jeder weiß, daß alle anderen sie ebenso anerkennen.“<sup>97</sup>

Ohne die unhaltbaren Annahmen, daß alle Grundgüterausstattungen objektiv meßbar sind, tauchen mithin alle Probleme, die sich dem Utilitarismus, der modernen Wohlfahrtsökonomik oder der Buchananschen Theorie wegen der Introspektivität ihrer Basisgröße Nutzen in den Weg stellten, in der Rawlsschen Konzeption in analoger Form wieder auf. Es ist nicht zu sehen, daß das Problem der Messung und des intersubjektiven Vergleichs von Grundgütern in irgendeiner Weise einfacher sein sollte als das der Messung und des intersubjektiven Vergleichs der traditionellen Maßgröße Nutzen, die Rawls so elegant zu umschiffen suchte. Sind folglich die Meß- und Vergleichsprobleme von Grundgütern lösbar, so müßte auch das Problem der Quantifizierung und utilitaristischen Saldierung interpersoneller Nutzengrößen lösbar sein, und die „Theorie der Gerechtigkeit“ besäße insoweit gegenüber dem Utilitarismus keine Vorteile. Rawls' Konzept ist daher im Vergleich zum Utilitarismus nur insofern weniger

---

<sup>96</sup> Vgl. *Mueller* (1989a), S. 418.

<sup>97</sup> *Rawls* (1992 [1980]), S. 88.

voraussetzungsreich, als das Differenzprinzip lediglich eine interpersonelle Vergleichbarkeit *ordinaler* Grundgüterindizes verlangt.<sup>98</sup>

### C. Ergebnis

Die in § 4 angestellten Überlegungen zeigen, daß die schwache Vertragstheorie weder in der Version von Rawls noch in jener von Buchanan eine befriedigende Lösung des Optimumauswahlproblems leisten kann. Aus dem Gedankenexperiment unter dem Schleier der Unkenntnis folgt weder eine konkrete Gerechtigkeitskonzeption, noch ist stets eine Einigung in diesem Sinne zu erwarten.

Erstens folgt aus dem Gedankenexperiment unter dem Schleier der Unkenntnis in beiden Varianten keine konkrete *Gerechtigkeitskonzeption*. Selbst dann, wenn man unterstellt, daß der von Rawls beschriebene Urzustand die ökonomischen Rationalentscheider stets zu Entscheidungen im Sinne der kantischen Universalisierungsnorm zwingt, ist man nicht gezwungen anzunehmen, daß diese *Original position* eine Situation der Unsicherheit i.e.S. darstellt, d.h., daß die Individuen nicht in der Lage sind, subjektive Wahrscheinlichkeiten dafür zu bilden, sich in der einen oder anderen gesellschaftlichen Situation wiederzufinden. Das Maximinprinzip ist daher nur eine unter einer Vielzahl denkbarer Gerechtigkeitskonzeptionen, die sich aus dem Gedankenexperiment mit dem Schleier des Nichtwissens herleiten lassen. Es folgt nicht aus dem Universalisierungspostulat, sondern ist die Folge der impliziten Einführung eines zusätzlichen und weitergehenden externen Werturteils.

Geht man hingegen, wie Buchanan, davon aus, daß die konstitutionellen Ausgangsbedingungen eine Entscheidung unter Risiko definieren und Wahrscheinlichkeitsschätzungen also möglich sind, so ist man keineswegs auf die Verwendung des Laplaceschen Prinzips des unzureichenden Grundes festgelegt. Auch die Buchanansche Gleichwahrscheinlichkeitsannahme folgt nicht aus seinem Modell der Entscheidung unter dem Schleier der Unsicherheit, sondern nur bei zusätzlicher und uneingestandener Einführung eines Gleichwertigkeitspostulats, das genau ein solches externes Werturteil darstellt, das nach Buchanans eigenen Annahmen in der Konstitutionenökonomik eigentlich nichts zu suchen hat.

---

<sup>98</sup> Vgl. in diesem Sinne Arrow (1973), S. 254. Offensichtlich falsch ist hingegen die Behauptung von Wolff, Rawls' Argument für das Maximin-Prinzip setze interpersonell vergleichbare *kardinale* Nutzenfunktionen voraus; vgl. Wolff (1976), S. 855; ders. (1977), S. 28, S. 87 und S. 133 ff.

Darüber hinaus wurde gezeigt, daß selbst dann, wenn alle Individuen im Urzustand das gleiche Entscheidungsprinzip unter unsicheren Erwartungen anwenden, ein *monologischer Konsens*, der es möglich macht, den Entscheidungskalkül eines beliebigen Individuums für repräsentativ für alle zu halten, nur unter zusätzlichen Bedingungen folgt. Buchanans Schleier der Unsicherheit kann eine solche stillschweigende Einigung überhaupt nicht garantieren, da er - als angeblich empirisches Phänomen - nur auf einer partiellen Universalisierung beruhen kann. Das aus der starken Vertragstheorie bekannte Problem der Indeterminiertheit des Legitimationsergebnisses wird hierdurch nur von der Ebene der individuellen auf jene der konstitutionellen Präferenzen verschoben. Überdies bezahlt man den Verzicht auf intersubjektive Nutzenvergleiche mit der Möglichkeit (zusätzlicher) diskriminierender Vertragsergebnisse wie dem diskutierten Gesellschaftsvertrag zu Lasten künftiger Generationen.

Die Rawlssche Konzeption des Schleiers des Nichtwissens, die eine vollkommene Universalisierung unterstellt, erreicht demgegenüber zwar formal einen monologischen Konsens, weil sich alle Parteien in die „wahren“ Präferenzen des jeweiligen schlechtestgestellten Positionsinhabers versetzen. Dieser beruht aber auf den unrealistischen Annahmen, daß zum einen das Meß- und Indizierungsproblem der individuellen Grundgüterausstattungen gelöst ist; zum anderen müssen die so ermittelten Grundgüterindizes intersubjektiv vergleichbar und die Ergebnisse dieses Vergleichs gemeinsames Wissen aller Beteiligten sein, damit es keinen Dissens darüber gibt, wer von den postkonstitutionellen Positionsinhabern die am schlechtesten gestellte Person ist. § 5 wird diese Ergebnisse mit Hilfe eines einfachen spieltheoretischen Modells untermauern.

## **§ 5 Diskriminierung und Dissens unter dem Schleier der Unkenntnis**

Im vorangegangenen § 4 wurde gezeigt, daß weder die Rawlsschen noch die Buchananischen Annahmen über die fiktive Entscheidungssituation des Urzustandes hinreichen, um die Maximierung der jeweils gleichen Maximin- bzw. Laplace-Wohlfahrtsfunktion durch alle Entscheidungsträger behaupten zu können. Beide Vertragstheorien formulieren nur partielle (fiktive) Erklärungs- bzw. Prognoseargumente; die Annahmen in ihren Explanantia sind zu unvollständig, um den Explanandum-Satz aus ihnen zu deduzieren.

In diesem § 5 wird - in umgekehrter Argumentationsrichtung - dargelegt, daß die beiden Wohlfahrtsfunktionen Implikationen haben, die jenen der Universalisierungsnorm und der sie operationalisierenden Prämissen des Gedankenexperiments widersprechen; in beiden Fällen führt der fiktive Verfassungskalkül der Individuen nicht ausschließlich zu fairen, sondern unter bestimmten, näher zu spezifizierenden Umständen auch zu (höchst) unfairen und nicht-einstimmigen Entscheidungen. Die Behauptung, unter einem Schleier der Unkenntnis werde es stets zu fairen und einstimmigen Lösungen kommen, ist darum logisch falsch.

### **A. Der Schleier der Unkenntnis und das Hobbessche Ordnungsproblem**

Der Schleier der Unkenntnis - gleichgültig, ob in Rawls' oder in Buchanans Variante - kann, so werde ich argumentieren, die an ihn geknüpfte Erwartung, daß in seiner Gegenwart Entscheidungen stets fair (im Sinne der Universalisierungsnorm) und einstimmig getroffen werden, nicht erfüllen. Mit Hilfe eines einfachen spieltheoretischen Modells werde ich darlegen, daß diese Erwartung schon dann nicht mehr in allen Fällen erfüllt wird, wenn angenommen wird, daß die auf der konstitutionellen Ebene verabredeten Gesellschaftsverträge der Anreizstruktur eines *Gefangenendilemmas* (GD) unterliegen. Insbesondere diese Annahme hat in der Vertragstheorie eine lange Tradition und Verbreitung. Schon die Anreizsituation im Hobbesschen Anarchiezustand wird im allgemeinen dahingehend interpretiert, daß der Autor das im Sinn hatte, was heute als Gefangenendilemma bezeichnet wird.<sup>1</sup> Das Gefangenendilemma, so glaubt man, faßt „lediglich das Problem von Hobbes in neue Worte“<sup>2</sup>.

Das Gefangenendilemma-Spiel modelliert in stark vereinfachter Form einen Konflikt zwischen individueller und „kollektiver“ Rationalität, der zwischen egoistischen *Homines oeconomici* entstehen kann.<sup>3</sup> Eine Gruppe von (im Standardfall: zwei) Spielern erstrebt einstimmig die Bereitstellung eines Kollektivguts, welche die Beitragsleistung aller Beteiligten erfordert. Der Realisierung dieser möglichen Pareto-Verbesserung, die durch das Handeln aller Individuen herbeigeführt werden könnte, steht jedoch entgegen, daß jeder Spieler einen individuellen Anreiz hat, darauf zu setzen, daß nur der oder die anderen zur Bereitstellung des Gutes beitragen, er selbst aber nicht. Da im Gefangenendilemma alle Individuen diesem Kalkül folgen, wird das von allen gewünschte Gut nicht bereitgestellt, und das entstehende stabile Gleichgewicht ist paretoinferior. Die individuell rationale Verfolgung des eigenen Vorteils führt also zur kollektiven Selbstschädigung.

In der Terminologie der modernen Theorie der Kollektivgüter liegt dem Gefangenendilemma ein *Allmendeproblem*<sup>4</sup> zugrunde, das allgemein dadurch gekennzeichnet ist, daß eine grundsätzlich knappe Ressource mit der Eigenschaft der *Rivalität* im Konsum auch von jenen Individuen genutzt werden kann, die zu ihrer Erstellung oder Erhaltung nichts beigetragen haben (*Nicht-Ausschluß*). So könnten sich im Gefangenendilemma alle Spieler individuell besser stellen, wenn sie durch Wahl ihrer kooperativen Strategie einer allgemein akzeptierten Regel folgen. Niemand kann jedoch von den Nutzen des so produzierten Kollektivguts ausgeschlossen werden; jeder Spieler hat daher einen Anreiz, sich durch Wahl der defektiven Strategie seiner eigenen Beitragsleistung zu entziehen. Da das Gut rival ist, zwingt hierdurch jeder jeden anderen zum Konsumverzicht; zugleich trägt er - in Form von entgangenem Nutzen - die Kosten des

---

<sup>1</sup> Zum Gefangenendilemma allgemein vgl. *Rapoport/Chammah* (1965); *Rapoport* (1987); *Taylor* (1976), *ders.* (1987); *Brembs* (1996); *Binmore* (1996); *Kuhn* (1997). Eine Interpretation der Hobbesschen Theorie mittels des Gefangenendilemmas findet sich z. B. bei *Gauthier* (1969), S. 76-98; *ders.* (1986), S. 79-82; *Ullmann-Margalit* (1977), S. 62-73; *Mackie* (1981 [1977]), S. 144-151; *Hoerster* (1977), S. 128-139; *Kliemt* (1977), S. 92-100; *McLean* (1981), *ders.* (1985); *Taylor* (1976), S. 99-118, *ders.* (1985), *ders.* (1987), S. 126-150; *Kavka* (1986), S. 111-113.

<sup>2</sup> *McLean* (1985), S. 33.

<sup>3</sup> „Kollektive“ Rationalität steht hier nicht für die Rationalität eines Kollektivs als solchem, sondern für ein Verhalten, das aus Sicht eines jeden einzelnen Beteiligten unter Selbstbindung individuell rational ist. Der üblichen Unterscheidung zwischen individueller und „kollektiver“ (d.h. bei allen gleichermaßen gegebener individueller) Rationalität entspricht bei *Hobbes* die Trennung von *in foro interno* und *in foro externo* verpflichtenden individuellen Klugheitsregeln; vgl. *Kliemt* (1988), S. 154.

<sup>4</sup> Vgl. im einzelnen *Müller/Tietzel* (1998).

Konsums des anderen. Die hierdurch ausgelöste „Tragik der Allmende“<sup>5</sup> besteht in einer Übernutzung des gemeinschaftlich genutzten Guts und der Aufzehrung möglicher ökonomischer Renten.

Die Art der Nutzung von Gütern und Ressourcen im „Hobbesschen Dschungel“ hat genau diese Anreizstruktur.<sup>6</sup> Auch hier erfolgt *kein Ausschluß* nicht-kooperierender Dritter von der Nutzung von Gütern; vielmehr besitzt, in Abwesenheit einer formalen, rechtserzwingenden Institution, jeder Mensch ein faktisches „Recht auf alles“<sup>7</sup>, das ihn „berechtigt“, sich anzueignen, was immer er wünscht und soviel, wie er bekommen kann. Wenn es ihm beliebt, wird er daher Güter nicht selbst produzieren, sondern die Produkte anderer (gewaltsam) entwenden. Doch die Kehrseite eines jeden „Rechts“ ist die „Pflicht“ eines anderen Menschen, die „geschuldete“ Leistung zu erbringen bzw. deren Entzug zu erdulden. Glaubt jeder, ein „Recht auf alles“ zu haben, so geraten die Ansprüche zwangsläufig miteinander in *Rivalität*. Sein „Recht“ durchsetzen kann nur derjenige, der es sich als erster nimmt. Der auf diese Weise initiierte Aneignungswettlauf kulminiert in einem „Krieg eines jeden gegen jeden“, der gekennzeichnet ist durch „beständige Furcht und Gefahr eines gewaltsamen Todes - das menschliche Leben ist einsam, armselig, ekelhaft, tierisch und kurz.“<sup>8</sup>

Die meisten Vertragstheoretiker nach Hobbes übernahmen die Hobbessche Allmende-problematik eines anarchischen, bürgerkriegsähnlichen Urzustands als Ausgangssituation ihrer staatsphilosophischen Rechtfertigungsversuche: Strukturelle Ähnlichkeit mit dem „Hobbesschen Dschungel“ hat etwa die Ausgangssituation in der Vertragstheorie von John Locke und, auf dieser aufbauend, die von Robert Nozick<sup>9</sup>; und auch den vertragstheoretischen Urzuständen von Rousseau und Kant liegt die gleiche „Tragik der Allmende“ zugrunde.<sup>10</sup> Am deutlichsten unter den „neuen Vertragstheoretikern“ folgt diesem Vorbild der Buchanan'sche Staatsgründungsansatz, der sich erklärtermaßen als eine Fortent-

<sup>5</sup> Hardin (1968).

<sup>6</sup> Zur Interpretation des Hobbesschen Naturzustands als „Tragik der Allmende“ vgl. *Opahls* (1973).

<sup>7</sup> *Hobbes* (1966 [1647]), S. 83.

<sup>8</sup> *Hobbes* (1976 [1651]), S. 96.

<sup>9</sup> Zur strukturellen Ähnlichkeit der Naturzustände von *Locke* (1992 [1689]) und *Nozick* (1976 [1974]) mit dem Hobbesschen Dschungel bzw. dem Gefangenendilemma vgl. *Kliemt* (1977), S. 59 f. und Anhang S. 14, Fußnote 4; *Gauthier* (1993), S. 26.

<sup>10</sup> Interpretationen der Vertragstheorie von *Rousseau* (1977 [1762]) auf der Basis eines Gefangenendilemmas finden sich bei *Runciman/Sen* (1965) und *Gauthier* (1993), S. 26; zu Unterschieden zum durch das Gefangenendilemma beschriebenen Hobbesschen Naturzustand vgl. aber *Kersting* (1994), S. 157 f. Die strukturelle Ähnlichkeit von Kants Naturzustand mit dem Gefangenendilemma wird besonders deutlich in Kant (1983 [1797]), § 42, S. 424 f.



wicklung der Hobbesschen Vertragstheorie versteht.<sup>11</sup> Buchanan rekurriert explizit auf das Modell des Gefangenendilemmas, um den Zustand der Regellosigkeit zu modellieren und die Wohlstandsgewinne aufzuzeigen, die durch staatliches Zwangshandeln - insbesondere durch die Definition und Durchsetzung allgemein anerkannter Eigentumsrechte - realisiert werden können.<sup>12</sup> In allen diesen Fällen legen und legen die Vertragstheoretiker den Schluß nahe, der Staat - in Hobbesscher Diktion der „Leviathan“ - möge die Individuen durch Anwendung seines Zwangsinstrumentariums dazu bringen, die kooperative Strategie zu wählen, um ihre Dilemmasituation zu verlassen.

Trotz der jahrhundertealten Tradition, fiktive Anarchiesituationen zum Ausgangspunkt vertragstheoretischer Legitimationsbemühungen zu machen, bleibt der an Hobbes und dem Gefangenendilemma orientierte kontraktualistische Kalkül keineswegs auf Rechtfertigungen von Staatstätigkeit insgesamt beschränkt, sondern er läßt sich auf alle Allmendesituationen anwenden, in denen Ausschlußrechte an rivalen Gütern entweder nicht definiert sind oder zuge-schriebene Rechte nicht durchgesetzt werden. Ein prominentes Beispiel ist das Problem der Übernutzung unserer natürlichen Umwelt<sup>13</sup>, die heute in weiten Bereichen als freies Gut alloziert wird: Jedes Unternehmen kann natürliche Ressourcen wie Luft oder Wasser durch Abgasemissionen oder Abwässer verunreinigen, ohne hierfür bezahlen zu müssen (Nicht-Ausschluß). Die Rivalität dieses „Rechts“ mit den entgegengesetzten Ansprüchen anderer Bürger auf eine saubere Umwelt liegt auf der Hand.

Von der „Tragik der Allmende“ ist aber auch die Wirtschaftsordnungspolitik der deutschen Ordoliberalen bedroht<sup>14</sup>: Zwar hat jeder Marktteilnehmer ein „konstitutionelles“ Interesse an der Einführung einer fairen Wettbewerbsordnung; da er aber die Nutzen hieraus auch dann erhält, wenn er sich selbst nicht wettbewerbskonform verhält, wird er versuchen, für die eigene Situation eine Ausnahme hiervon zu erwirken. „*Rent seeking*“<sup>15</sup> - das Streben nach nicht-marktlichen Renten - ist lediglich ein anderer Name für den Positionswettlauf um die steuerbaren Einkommen der anonymen Masse der Steuerzahler, die interventionsbereite Politiker den Interessengruppen wie ein Allmendegut feil-

<sup>11</sup> Vgl. *Buchanan* (1984 [1975]), S. 76 ff.

<sup>12</sup> Vgl. *Buchanan* (1977 [1972]), S. 88; *ders.* (1984 [1975]), 39; *Brennan/Buchanan* (1993 [1985]), S. 3 ff. - Andere Modelle zur Formalisierung der natürlichen Verteilung im Sinne von *Hobbes* und *Buchanan* finden sich bei *Bush/Mayer* (1974); *Grossman/Kim* (1995).

<sup>13</sup> Vgl. z. B. *Dasgupta/Heal*, 1979, S. 18-21; *Taylor* (1987), S. 142.

<sup>14</sup> Vgl. ähnlich *Kirsch* (1981). Zum Konzept der Ordnungspolitik allgemein vgl. z. B. *Cassell/Ramb/Thieme* (Hrsg.) (1988).

<sup>15</sup> *Krueger* (1974); vgl. zur Theorie des *Rent seeking* den Überblick bei *Mueller* (1989a), Kap. 13.

bieten. In dem Maße, in dem Regierungen den Wünschen der Marktteilnehmer nach dem Aufbau oder Schutz von Privilegien im Wettbewerb nachgeben, erodiert die Wettbewerbsordnung, und im Extremfall stehen alle Marktteilnehmer so, als gäbe es überhaupt keinen institutionellen Rahmen ihres marktlichen Handelns. In allen diesen Fällen schließt die Vertragstheorie, der Staat solle eine Ordnung errichten und schützen, um den Individuen mittels wohlfahrtsfördernden Zwangs - durch Einsatz von Pigou-Steuern oder Durchsetzung von Eigentumsrechten - aus ihrem sozialen Dilemma herauszuhelfen.<sup>16</sup>

*Allmendeprobleme* (Ordnungsprobleme) sind indes nicht der einzige Anwendungsfall der Theorie des hypothetischen Gesellschaftsvertrags. In jüngerer Zeit kamen durch den Einfluß der „neuen Vertragstheoretiker“ zwei weitere Problembereiche der Kontraktansätze hinzu: Zum einen wird insbesondere seit den Arbeiten von Rawls die klassische Theorie des Gesellschaftsvertrags zunehmend zur normativen Lösung von *Verteilungsproblemen* verwendet, insbesondere zur Rechtfertigung einer interpersonellen Distribution von „Grundgütern“ wie Einkommen und Vermögen nach Maßgabe eines bestimmten Begriffs der „sozialen Gerechtigkeit“. Zum anderen nutzten Buchanan und Tullock in „The Calculus of Consent“ einen vertragstheoretischen Ansatz zur hypothetischen Legitimierung von Entscheidungsregeln zur Überwindung gesellschaftlicher *Abstimmungsprobleme*.<sup>17</sup>

Vergleicht man einerseits die Vertragstheorien, die sich mit der Überwindung von Verteilungs- und Abstimmungsproblemen beschäftigen, mit jenen Theorien andererseits, die auf einem Allmendeproblem mit Gefangenendilemmastruktur beruhen, so fällt auf, daß erstere mit der Prämisse konstitutioneller Unkenntnis arbeiten, letztere jedoch nicht. Weder im Hobbesschen „Leviathan“ noch in Buchanans „Die Grenzen der Freiheit“ findet sich ein dem Schleier der Unkenntnis auch nur vergleichbarer Mechanismus, worauf Buchanan ausdrücklich hinweist:

„Der hier (in *Die Grenzen der Freiheit*; C. M.) gemachte Ansatz unterscheidet sich gerade durch die Bedeutung der Erwartungen vom Ansatz in *The Calculus of Consent*. Jenes Buch ging von der Annahme aus, daß sich die Individuen bei Anwendung der Entscheidungsregeln über ihre Positionen so weit im ungewissen befinden, daß sie zumindest in dieser Hinsicht als Gleiche in das Verhandlungsstadium eintreten.“<sup>18</sup>

<sup>16</sup> Daneben belegen neuere empirische Forschungen, daß sanktionsbewehrtes Staatshandeln nicht als einziger Selbstbindungsmechanismus in Frage kommt, um Allmendedilemmata zu überwinden; unter Umständen kann dies auch den Nutzern einer gemeinschaftlich genutzten Ressource im Selbstmanagement gelingen. Eine Übersicht über die empirische und theoretische Literatur hierzu geben Müller/Tietzel (1998).

<sup>17</sup> Vgl. oben § 3 B.

<sup>18</sup> Buchanan (1984 [1975]), S. 78, Fußnote 3 (Hervorhebung im Original).

Dennoch findet sich in der Literatur häufig die Erwartung, daß die Existenz eines Schleiers der Unkenntnis die Überwindung einer regellosen Situation, die durch das Gefangenendilemma beschrieben wird, erleichtere. So diskutiert etwa Dennis C. Mueller drei alternative Situationen, in denen Individuen erfolgreich sein könnten, ein bestehendes Gefangenendilemma zu überwinden. Neben einer Situation, in der die Kosten einer Nichteinigung prohibitiv sein könnten, nennt er zum einen eine durch einen Buchananschein (empirischen) Schleier der Unsicherheit gekennzeichnete Entscheidungssituation und zum anderen die Bereitschaft der Individuen, sich bei ihrer Entscheidung gedankenexperimentell unter einen Schleier der Unkenntnis à la Rawls (bzw. Gauthier) zu versetzen:

„The second reason to believe that consensus can be reached stems from the long-term nature of constitutional choice and the uncertainty over the future this brings. ... individuals and groups, even if they think only of their own self-interest narrowly defined, may agree to provisions that protect and advance the interests of all individuals and groups out of a genuine uncertainty over what the future will bring. Finally, individuals may impose upon themselves uncertainty over future positions of the type that Rawls (1971) and, more recently, Gauthier (1986) have discussed ... If individuals imagine that they might occupy all future positions in the polity, the likelihood of obtaining a consensus on the articles of the constitution would be greatly enhanced.“<sup>19</sup>

Auch Gregory S. Kavka formuliert für das von ihm als Gefangenendilemma beschriebene Hobbessche Ordnungsproblem:

„Agreement would clearly be easier to obtain among parties deprived of information about their individual features and positions, for known individual differences can lead to irreconcilable disagreements about the content of the social contract“<sup>20</sup>.

Die Verbindung des Schleiers der Unkenntnis mit dem Gefangenendilemma ist keineswegs abwegig. Mit den beiden hier betrachteten Vertragstheorien von Rawls und Buchanan erscheint sie vielmehr vollkommen kompatibel. Buchanan unterstellt wiederholt, daß das Problem, das die Entscheidungsträger auf der konstitutionellen Ebene zu lösen versuchen, in der Überwindung eines Gefangenendilemmas besteht. Auch behauptet er die Existenz des Schleiers der Unsicherheit als ein empirisches Phänomen, das allgemein bei Verfassungsentscheidungen zu beobachten sei. Entsprechend müßten, auch wenn Buchanan diesen Fall (inkonsequenterweise) nicht berücksichtigt, die betrachteten Entschei-

<sup>19</sup> Mueller (1996), S. 320.

<sup>20</sup> Kavka (1986), S. 193. Ähnliche Äußerungen finden sich zuhauf. So beschreibt Vanberg (1997), S. 20 f., das „Rent seeking“ um Privilegien als Gefangenendilemma, das durch den Schleier des Nichtwissens bzw. der Ungewißheit überwunden werden könne. Leschke (1993), S. 88 und S. 90, fordert, den Schleier der Ungewißheit auch in Buchanans Anarchiesituation, welche die GD-Eigenschaften aufweist, zu übertragen. Frohlich/Oppenhimer (1996a) konstruieren einen Schleier der Unkenntnis, um in einer Experimentalsituation ein 5-Personen-Gefangenendilemma zu überwinden.

dungsträger auch dann unter unsicheren Erwartungen über ihre eigene postkonstitutionelle Position zu entscheiden haben, wenn sie eine Regel für ein Gefangenendilemma-Spiel wählen.<sup>21</sup>

Rawls behauptet nicht, daß das Problem, welches die betrachteten Rationalakteure unter dem Schleier des Nichtwissens lösen, ein Gefangenendilemma sei. Gleichwohl bleibt das konstitutionell zu lösende Entscheidungsproblem in seiner „Theorie der Gerechtigkeit“ vollkommen offen, da Rawls selbst unterschiedliche Verläufe der „OP-Kurve“ diskutiert, welche - analog zur Nutzenmöglichkeitsgrenze der traditionellen Wohlfahrtsökonomik - die maximal möglichen Grundgüterverteilungen auf die betroffenen Gesellschaftsmitglieder angibt.<sup>22</sup> Die Klasse der zulässigen Verläufe der OP-Kurve, die je ein anderes Entscheidungsproblem determinieren, schränkt Rawls nicht ein. Nichts spricht folglich gegen die Unterstellung eines Gefangenendilemmas als zu lösendes Problem; die (diskrete) OP-Kurve dieses Spiels wird im Zwei-Personen-Nutzen- (bzw. Grundgüter-)Diagramm durch die drei nord-östlichen Strategienkombinationen des Gefangenendilemmas beschrieben.<sup>23</sup>

In diesem und im nächsten Kapitel soll untersucht werden, ob und inwieweit Individuen unter einem Rawlsschen und einem Buchananschen Schleier der Unkenntnis in der Lage sind, auf faire und einstimmige Weise eine durch ein Gefangenendilemma beschriebene Allmendeproblematik zu überwinden. Es wird also, mit anderen Worten, geprüft, ob sich unter einem Schleier der Unkenntnis das Hobbessche Ordnungsproblem überhaupt noch stellt. Mit Hilfe eines einfachen spieltheoretischen Modells wird die Regelwahl unter einem Rawlsschen Schleier des Nichtwissens und einem Buchananschen Schleier der Unsicherheit für ein einfaches Zwei-Personen-Gefangenendilemma ( $G_2$ ) und für ein  $n$ -Personen-Gefangenendilemma ( $G_n$ ) analysiert. Es wird gezeigt, daß, wenn ein auf der konstitutionellen Ebene geschlossener Gesellschaftsvertrag durch ein Gefangenendilemma bedroht wird, auch ein „dicker“ Schleier der Unkenntnis - gleichgültig, in welcher Variante - nicht garantieren kann, daß eine faire (universalisierbare) und nicht eine diskriminierende Regel gewählt wird. Buchanans Schleier der Unsicherheit kann darüber hinaus nicht einmal garantieren, daß überhaupt eine Einigung erzielt wird.

<sup>21</sup> An einer Stelle (vgl. *Buchanan* 1977 [1976]) analysiert *Buchanan* allerdings Konsequenzen der Anwendung von *Rawls'* Schleier des Nichtwissens auf Kollektiventscheidungen in einer als „*Hobbessche Anarchie*“ bezeichneten Situation. Entgegen seinen üblichen Annahmen hat diese Situation jedoch nicht die Eigenschaften eines Gefangenendilemmas.

<sup>22</sup> Vgl. *Rawls* (1994 [1971]), S. 96 ff.

<sup>23</sup> Für eine graphische Darstellung des Zwei-Personen-Gefangenendilemmas vgl. unten § 6 Abbildung 6.2.

Die Wohlfahrtsfunktionen von Rawls und Buchanan haben damit Implikationen, die jenen der Universalisierungsnorm bzw. den Prämissen des Urzustands widersprechen. In § 6 werde ich darüber hinaus darlegen, daß auch die These falsch ist, daß die Existenz eines Schleiers der Unkenntnis die Fairneß und Einstimmigkeit der getroffenen Kollektiventscheidung in jedem Fall wahrscheinlicher mache. Unter *sicheren* Erwartungen, so werde ich zeigen, bestehen im Falle eines einmal gespielten Zwei-Personen-Gefangenendilemmas keinerlei Probleme, Einstimmigkeit und Fairneß zu erzielen, im Falle eines  $n$ -Personen-Gefangenendilemmas nur bei Erfüllung einer Zusatzbedingung. Die Akzeptanz der fairen Regel allseitiger Kooperation ist dann im Spiel, das die Entscheidungsträger auf der konstitutionellen Ebene spielen, jeweils die Gleichgewichtslösung in sukzessiver Dominanz.

## B. Verfassungswahl unter dem Schleier der Unkenntnis

Das Spiel, das die Entscheidungsträger im Urzustand bzw. auf der konstitutionellen Entscheidungsebene spielen, heißt *konstitutionelles Spiel*. Jeder Spieler  $i$  maximiert eine Zielfunktion, die im Rawls-Fall als ordinaler Grundgüterindex, im Falle von Buchanan als kardinale von-Neumann-Morgenstern-Nutzenfunktion interpretiert wird.<sup>24</sup> Das im Zeitpunkt der Verfassungsfindung  $t_c$  zu lösende Problem der postkonstitutionellen Gesellschaft wird durch ein Gefangenendilemma - kurz: das *postkonstitutionelle Spiel* - beschrieben, das als Zwei- und als Multi-Personen-Spiel (mit  $n > 2$ ) modelliert wird. Das konstitutionelle Entscheidungsproblem besteht darin, einstimmig eine Regel  $r$  für das postkonstitutionelle Spiel zu wählen, die das Gefangenendilemma überwindet. Einigen sich die Individuen nicht auf eine Regel, so wird in der postkonstitutionellen Periode  $t_p$  das Gefangenendilemma (in der jeweiligen Modellvariante) gespielt.

---

<sup>24</sup> Es wird dabei unterstellt, daß die Wahl gemäß den eigenen Präferenzen für jeden Beteiligten die dominante Strategie ist. Diese Annahme schließt die Möglichkeit einer „ineffizienten Einstimmigkeit“ im Sinne von *Brennan/Lomasky* (1984) aus, so daß Einstimmigkeit und Pareto-Effizienz stets einander entsprechen; vgl. oben S. 31, Fußnote 24.

## I. Das postkonstitutionelle Spiel

### 1. Das Zwei-Personen-Gefangenendilemma

Zunächst sei unterstellt, daß auf der konstitutionellen Ebene zwei Spieler  $A$  und  $B$  über eine Regel für ein einmal gespieltes Zwei-Personen-Gefangenendilemma  $G_2$  entscheiden; die Menge  $I$  der Individuen ist  $I_2 = \{A, B\}$ . Obwohl die einfache Zwei-Personen-Konstellation lediglich ein Spezialfall der anschließenden  $n$ -Personen-Modellierung ist, wird sie im folgenden stets gesondert mitbehandelt.

Der Grund hierfür ist zum einen, daß das einfache Zwei-Personen-Gefangenendilemma den Standardfall dieses Spiels darstellt, der bereits dem namensgebenden Problem der beiden Untersuchungshäftlinge<sup>25</sup> zugrunde lag; das  $n$ -Personen-GD stellt demgegenüber eine Erweiterung des Basisproblems dar, die zum Teil auf zusätzlichen Annahmen beruht, die in der Literatur nicht immer einheitlich gehandhabt werden. Zum anderen hat die Reduzierung des Entscheidungsproblems auf der konstitutionellen Ebene auf den Zwei-Personen-Fall den Vorzug der hohen Anschaulichkeit, der sich vor allem aus der Anwendbarkeit der einfachen Bimatrixrepräsentation des betrachteten Spiels ergibt.

Die Betrachtung nur zweier Spieler entspricht überdies der Vorgehensweise von Rawls und Buchanan, die in ihren Beispielen vorrangig auf Zwei-Personen-Konstellationen zurückgreifen, Buchanan explizit sogar auf  $G_2$ . Darüber hinaus spielt, wie sich noch zeigen wird, die Zwei-Personen-Konstellation auch eine Sonderrolle im Rahmen der konstitutionellen Entscheidungsfindung unter *sicheren* Erwartungen (§ 6).

		Spieler B:		
		$C_B$	$D_B$	
Spieler A:	$C_A$	$f_A[2], f_B[2]$	$f_A[1], g_B[1]$	
	$D_A$	$g_A[1], f_B[1]$	$g_A[0], g_B[0]$	mit $g_i[1] > f_i[2] > g_i[0] > f_i[1]$ (GD-Bedingung)

Abbildung 5.1: Zwei-Personen- Gefangenendilemma  $G_2$

Das einfache Zwei-Personen-Gefangenendilemma ist in Abbildung 5.1 gegeben, in dem die Strategienmenge  $\Phi_i = \{C_i, D_i\}$  eines Spielers  $i$  (mit  $i \in I_2 = \{A, B\}$ ) die beiden idealtypischen Handlungsalternativen „reine Kooperation  $C_i$ “ (= partieller Verzicht auf das eigene „Recht auf alles“) und „reine Defektion

<sup>25</sup> Vgl. Luce/Raiffa (1957), S. 94 ff.

$D_i$ “ (= keine Beschränkung) umfaßt.<sup>26</sup> Die erste Stelle einer jeden Zelle  $r \in M^2$  der Spielmatrix mit  $M^2 = \Phi_A \times \Phi_B$  gibt die jeweilige Auszahlung von Spieler  $A$  an, die zweite Stelle den Payoff von Spieler  $B$ . Die Funktion  $f_i[v(r)]$  beschreibt  $i$ 's Payoff, wenn er kooperiert; Spieler  $i$  erhält die Auszahlung  $g_i[v(r)]$ , wenn er defektiert.  $v(r)$  gibt hier und im folgenden an, wieviele Spieler insgesamt in einer Strategienkombination  $r$  kooperieren.

Situation  $(D_A, D_B)$  markiert den paretoinferioren Ausgangszustand, Situation  $(C_A, C_B)$  den kollektiv gewünschten Zustand einer allgemeinen Verhaltensbeschränkung. In der Anreizsituation eines Gefangenendilemmas wird jeder Spieler  $i$ , gleichgültig wie sich der andere Spieler verhält, rationalerweise seine individuellen Eigeninteressen in den Vordergrund stellen: Denn zum einen hat jeder wegen  $g_i[1] > f_i[2]$  durch Ausspielen seiner „Trittbrettfahrer-Strategie“  $D_i$  die Chance, die von ihm am meisten präferierte Situation  $(D_i, C_i)$  zu erreichen; zum anderen minimiert jeder einzelne Spieler durch Nichtkooperation seine zu erwartenden Verluste, da hierdurch die aus seiner Sicht schlechteste Konstellation vermieden wird, in der er selbst kooperiert, während der jeweils andere durch Nichtkooperation seinen besten Payoff erzielt. Während also das (schwache) Pareto-Prinzip wegen  $f_i[2] > g_i[0]$  für beide  $i$  zur Kooperation mahnt, drängt das Dominanzprinzip wegen  $g_i[1] > f_i[2]$  und  $g_i[0] > f_i[1]$  jeden Spieler  $i$  dazu, gegen das gemeinsame Interesse zu handeln. Im Ergebnis finden sich beide Individuen in der kollektiv schlechtesten Situation - dem paretoinferioren Nash-Gleichgewicht  $(D_A, D_B)$  - wieder.<sup>27</sup>

<sup>26</sup> In der Literatur wird, in Erweiterung des hier betrachteten Standardfalls, neuerdings auch der Fall einer kontinuierlichen Menge von Strategien im Gefangenendilemma betrachtet, welche die Unterscheidung verschiedener Grade der Kooperation bzw. Defektion erlaubt; vgl. *Frohlich/Opppenheimer* (1996b). Von dieser Möglichkeit wird nachfolgend wie üblich abgesehen.

<sup>27</sup> Ein Nash-Gleichgewicht ist eine Strategienkombination, bei der - gegeben die Strategien des jeweils anderen Spielers - kein Spieler einen Anreiz hat, von seiner gewählten Strategie abzuweichen. Formal (vgl. z. B. Gibbons 1992, S. 8) ist in einem in Normalform gegebenen  $n$ -Personen-Spiel  $G = \{S_1, \dots, S_n; u_1, \dots, u_n\}$  die Strategienkombination  $(s_1^*, \dots, s_n^*)$  ein Nash-Gleichgewicht, wenn für jeden Spieler  $i$  die Strategie  $s_i^*$  die beste Antwort auf die Strategien der  $n-1$  anderen Spieler  $(s_1^*, \dots, s_{i-1}^*, s_{i+1}^*, \dots, s_n^*)$  ist, d.h.  $u_i(s_1^*, \dots, s_{i-1}^*, s_i^*, s_{i+1}^*, \dots, s_n^*) \geq u_i(s_1^*, \dots, s_{i-1}^*, s_i, s_{i+1}^*, \dots, s_n^*)$  für jede verfügbare Strategie  $s_i$  in  $S_i$ .



## 2. Das $n$ -Personen-Gefangenendilemma

Rawls und Buchanan betrachten, soweit sie ihre Ausführungen formalisieren, nirgendwo explizit ein  $n$ -Personen-Modell. Aus dem Kontext ergibt sich dabei jedoch zumeist, daß die von ihnen untersuchten Zwei-Personen-Konstellationen zur Illustration von Entscheidungsproblemen großer Gesellschaften gemeint sind. Daher soll im folgenden explizit die Menge  $I$  der Individuen auf  $I_n = \{A, B, \dots, n\}$  erweitert werden. Auf der konstitutionellen Ebene wählen die Individuen eine Regel für ein postkonstitutionelles  $n$ -Personen-Gefangenendilemma  $G_n$ . Bei sicheren Erwartungen - d.h. *ohne* Schleier der Unkenntnis im Zeitpunkt der konstitutionellen Entscheidung - lautet  $G_n$  in Normalform:<sup>28</sup>

$$(5.1) \quad G_n = (\Phi_A, \Phi_B, \dots, \Phi_n; \pi_A, \pi_B, \dots, \pi_n)$$

$$\Phi_i = \{C_i, D_i\}$$

$$\pi_i(r) = \begin{cases} f_i[v(r)], & \text{falls } i \text{ kooperiert} \\ g_i[v(r)], & \text{falls } i \text{ defektiert} \end{cases}$$

$$\text{mit} \quad g_i[v(r)] > f_i[v(r)+1] \text{ und}$$

$$f_i[n] > g_i[0] \quad \text{und}$$

$$f_i[v(r)] \quad \text{streng monoton steigend für alle } 1 \leq v(r) \leq n \quad \text{und}$$

$$g_i[v(r)] \quad \text{streng monoton steigend für alle } 0 \leq v(r) \leq n-1.$$

Wie in  $G_2$ , enthält die Strategiemenge  $\Phi_i$  auch hier für jeden Spieler  $i$  die beiden reinen Strategien „Kooperation  $C_i$ “ und „Defektion  $D_i$ “:  $\Phi_i = \{C_i, D_i\}$ .  $r = (s_i, s_{-i})$  bezeichnet eine beliebige Strategienkombination des  $n$ -Personen-Gefangenendilemmas mit  $r \in M^n$ , und  $M^n = \Phi_A \times \Phi_B \times \dots \times \Phi_n$ ; dabei bezeichnet  $s_i$  die Strategie, die Spieler  $i$  wählt, und  $s_{-i}$  den Strategienvektor der  $n-1$  übrigen Mitspieler ohne  $i$ .

Die Auszahlungen  $\pi_i(r)$  eines jeden Spielers  $i$  hängen von zwei Faktoren ab: zum einen von der Wahl seiner eigenen Strategien  $C_i$  oder  $D_i$  und zum anderen von der Anzahl  $v$  der Individuen, welche insgesamt die Strategie  $C_i$  wählen. Letztere Bedingung unterstellt, daß es hierbei nicht darauf ankommt, *welcher* der anderen Beteiligten kooperiert, sondern lediglich auf die Gesamtzahl der Kooperierenden. Die Funktion  $f_i[v(r)]$  gibt die Auszahlung für Spieler  $i$  an, wenn er selbst einer von insgesamt  $v$  kooperierenden Individuen ist;  $g_i[v(r)]$

<sup>28</sup> In der Literatur kursieren mehrere Versionen des  $n$ -Personen-Gefangenendilemmas, die sich allerdings nur in Details unterscheiden. Die wichtigsten von ihnen finden sich bei Hardin (1971); Hamburger (1973), ders. (1979), Kap. 7; Schelling (1973); Taylor (1976), ders. (1987), S. 82 ff. Vgl. zu Taylors Modellierung auch Kliemt/Schauenberg (1982).

bezeichnet demgegenüber  $i$ 's Payoff, wenn er selbst defektiert, während  $v$  Spieler kooperieren. Die Auszahlungsfunktionen  $\pi_i(r)$  können sich intersubjektiv unterscheiden; das betrachtete Gefangenendilemma muß also nicht symmetrisch sein.

Die beiden Bedingungen  $g_i[v(r)] > f_i[v(r)+1]$  und  $f_i[n] > g_i[0]$  beschreiben den aus dem Zwei-Personen-Fall des Gefangenendilemmas bekannten Konflikt zwischen dem Dominanz- und dem Pareto-Prinzip: Während ersteres verlangt, unabhängig von den Handlungen aller anderen Spieler zu defektieren, sagt letzteres, daß die Wahl der kooperativen Strategien im gemeinsamen Interesse aller Individuen liegt. Die erste Bedingung besagt, daß gleichgültig, welche Strategien die anderen Spieler wählen, die defektive Strategie  $D_i$  einen höheren Payoff erzielt als die kooperative Strategie  $C_i$ . Defektion ist also wie in  $G_2$  für alle Spieler die dominante Strategie, was impliziert, daß  $r_q = (D_A, D_B, \dots D_n)$  das einzige Nash-Gleichgewicht des Spiels ist.

Die zweite Bedingung verlangt, daß jeder Spieler  $i$  die Situation, in der alle Individuen kooperieren, jener Strategienkonstellation vorzieht, in der alle defektieren. Die Situation fairer Kooperation aller Spieler  $r_f$  pareto-dominiert somit für alle Payoff-Konstellationen das Nash-Gleichgewicht  $r_q$ , d.h. für alle  $i = A, B, \dots, n$ :

$$\pi_i(r_f) = \pi_i(C_A, C_B, \dots C_n) > \pi_i(r_q) = \pi_i(D_A, D_B, \dots D_n)$$

Wie man leicht sieht, enthält das hier definierte  $n$ -Personen-Gefangenendilemma für den Spezialfall  $n = 2$  das übliche Zwei-Personen-Gefangenendilemma aus Abbildung 5.1. Nach (5.1) ergibt sich für jeden Spieler  $i$  (mit  $i = A, B$ ) die für das „normale“ Gefangenendilemma  $G_2$  übliche Auszahlungsrelation (GD-Bedingung):

$$g_i[1] > f_i[2] > g_i[0] > f_i[1].$$

## II. Das konstitutionelle Spiel

Betrachten wir nun das *konstitutionelle Spiel*  $K_n$  (bzw. im Zwei-Personen-Fall  $K_2$ ), das die Entscheidungsträger im Zeitpunkt  $t_c$  spielen. Die Spieler im Urzustand wissen, daß sie, wenn sie sich in der konstitutionellen Periode  $t_c$  nicht auf eine Regel einigen, sie in der postkonstitutionellen Periode  $t_p$  Spieler in einem Gefangenendilemma  $G_n$  sein werden. Aufgrund eines „dicken“ Schleiers der Unkenntnis weiß jedoch kein Spieler  $i$ , in welcher Position  $j$  (mit  $j = 1, \dots, n$ ) des Gefangenendilemmas er sich dabei befinden wird.<sup>29</sup> Solange die Identität

<sup>29</sup> Für eine ähnliche Idee den Zwei-Personen-Fall betreffend vgl. *Frohlich* (1992).

der Spieler in der postkonstitutionellen Periode  $t_p$  nicht feststeht, werde ich im folgenden von den Positionen bzw. Positionsinhabern  $j = 1, 2, \dots, n$  des postkonstitutionellen Spiels sprechen im Unterschied zu den Spielern  $i = A, B, \dots, n$  auf der konstitutionellen Ebene, deren Identität stets bekannt ist.

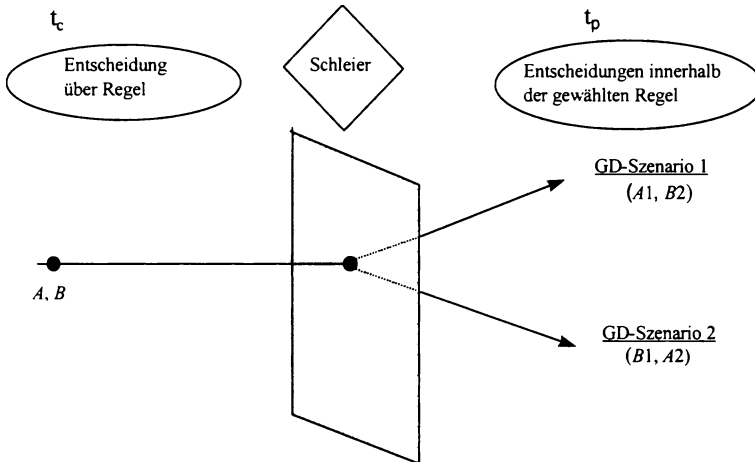


Abbildung 5.2: Die Struktur des konstitutionellen Spiels  $K_2$  unter dem Schleier der Unkenntnis (Zwei-Personen-Fall)

Da es in einem konstitutionellen Spiel mit  $n$  Personen  $n!$  verschiedene Permutationen von Spielerkonstellationen geben kann, unterstellt jeder Entscheidungsträger  $i$  im Urzustand  $n!$  verschiedene „Gefangenendilemma-Szenarien“. Die von den Spielern auf der konstitutionellen Ebene für die postkonstitutionelle Situation erwarteten Gefangenendilemma-Szenarien seien gemeinsames Wissen aller Spieler. Abbildung 5.2 stellt die Struktur dieses Modells für den einfachsten Fall - das konstitutionelle Spiel auf der Basis eines Zwei-Personen-Gefangenendilemmas  $G_2$  der Spieler  $A$  und  $B$  - dar, in dem es folglich lediglich zwei verschiedene GD-Szenarien zu berücksichtigen gilt: für Szenario 1 erwarten die Spieler, daß  $A$  in Position 1 („ $A1$ “) und  $B$  in Position 2 („ $B2$ “) sein wird; für Szenario 2 sind diese Überzeugungen genau entgegengesetzt („ $B1$ “ und „ $A2$ “).

Auf der konstitutionellen Ebene sollen die Spieler über eine vollkommene Fähigkeit zur Selbstbindung verfügen. Unter dieser Bedingung wählt jedes Individuum  $i$  nicht die üblichen GD-Strategien  $C_i$  oder  $D_i$ , sondern eine *Strategienkombination (Regel)*  $r \in M^r$  aus der Menge aller möglichen Spielergebnisse des durch (5.1) beschriebenen postkonstitutionellen Gefangenendilemmas  $G_n$ . Im Zwei-Personen-Gefangenendilemma  $G_2$  (mit  $n = 2$ ) haben die beiden

Spieler also vier Strategienkombinationen (Regeln) zur Auswahl, nämlich die vier Zellen  $r \in M^2$  mit  $M^2 = \{(D_1, C_2), (C_1, C_2), (D_1, D_2), (C_1, D_2)\}$  des postkonstitutionellen Zwei-Personen-Gefangenendilemmas.

Das  $n$ -Personen-Gefangenendilemma  $G_n$  verfügt über  $2^n$  mögliche Strategienkombinationen (Regeln)  $r \in M^n$ , die auf der konstitutionellen Ebene gewählt werden können:

- Zum einen können sich die Spieler dafür entscheiden, dem Gefangenendilemma zu entkommen, indem sie die „faire Regel“  $r_f = (C_1, C_2, \dots, C_n)$  wählen, die eine dritte Partei - etwa den Staat - dazu ermächtigt, die kooperative Lösung des Gefangenendilemmas  $G$  mit Zwang durchzusetzen. Als „fair“ wird die Regel allgemeiner Kooperation bezeichnet, weil unter ihr kein Spieler übergangen wird. Denn unter  $r_f$  wird *allen* Spielern von  $G_n$  gleichermaßen eine Nutzenverbesserung gegenüber dem Nash-Gleichgewicht des Gefangenendilemmas zugestanden. Jeder Positionsinhaber  $j$  erzielt in dieser Situation also die Auszahlung  $f_j[n]$  der Situation allgemeiner Kooperation.
- Zweitens könnten die Individuen auch entscheiden, das nicht-kooperative Nash-Gleichgewicht von  $G_n$  zwangsweise durchzusetzen: Die *Status-quo-Regel*  $r_q = (D_1, D_2, \dots, D_n)$  zwingt alle Positionsinhaber zur Defektion, und zwar selbst dann, wenn einige von ihnen beabsichtigen sollten, die dominierte Strategie zu wählen. Jeder Positionsinhaber erzielt unter  $r_q$  den Payoff  $g_j[0]$ .
- Alle übrigen  $2^n - 2$  wählbaren Strategienkombinationen des postkonstitutionellen Gefangenendilemmas sind *diskriminierende Regeln*  $r_d \in M^n \setminus \{r_f, r_q\}$ , unter welchen mindestens einer der  $n$  Spieler zur Kooperation gezwungen werden, während alle übrigen Positionsinhaber defektieren müssen. Die unter  $r_d$  Defektierenden werden durch die Staatsintervention also in die Lage versetzt, „Trittbrettfahrer“ zu sein und auf Kosten jener Positionsinhaber, welche die kooperative Strategie wählen müssen, einen Defektionspayoff in Höhe von  $g_j[v]$  zu erzielen, der stets größer ist als die Auszahlung  $g_j[0]$  im Status quo. Die ganze Last der Regelbefolgung wird hier also den zwangsweise kooperierenden Positionsinhabern aufgebürdet, welche die Auszahlung  $f_j[v]$  erhalten, die niedriger sein kann als die Auszahlung  $g_j[0]$  des Nash-Gleichgewichts. Diskriminierende Regeln bedeuten mithin die staatlich institutionalisierte Freigabe einiger Spieler zur Ausbeutung durch andere.

Nach den Theorien von Rawls und Buchanan ist zu erwarten, daß sich die Individuen unter dem Schleier der Unkenntnis stets auf die Regel  $r_f$  einigen.<sup>30</sup> In

<sup>30</sup> Genau genommen wählen die Individuen nach Rawls nicht eine konkrete Verhaltensregel, sondern eine Gerechtigkeitskonzeption, die in die Form von Gerechtigkeitsprinzipien gegossen ist. Unterstellt man jedoch mit Rawls, daß die jeweils auf der konstitutionellen Ebene gewählte Gerechtigkeitskonzeption das Leben in der postkon-

der Rawlsschen „Theorie der Gerechtigkeit“ ergibt sich dies aus dem ethischen Werturteil, das durch die Prämissen des Urzustandes unter dem Schleier des Nichtwissens illustriert werden soll. Wer die Universalisierungsnorm teilt, so versucht Rawls über den Umweg seiner komplizierten vertragstheoretischen Argumentation zu zeigen, muß auch das Differenzprinzip teilen. Das aber bedeutet, anders gesagt, daß auch alle Implikationen dieses Prinzips mit der Universalisierungsnorm in Einklang zu bringen sein müssen. In Gefangenendilemma-Situationen entspricht allein die Wahl der kooperativen Strategie dem ethischen Verallgemeinerungsprinzip; der kategorische Imperativ verlangt hier, die strategische Gefangenendilemma-Struktur zu ignorieren und unbedingt zu kooperieren.<sup>31</sup> Unter dem Schleier des Nichtwissens, der nach Rawls nur eine Metapher für das kantische Moralprinzip ist, müßten sich die Individuen daher immer auf die Durchsetzung der kooperativen Lösung des Gefangenendilemmas einigen.

Eine diskriminierende Regel  $r_d$ , die mindestens eines der Individuen zur Kooperation zwingt und damit zum Ausbeutungsobjekt der Defektierenden macht, wäre demgegenüber nicht mit der Universalisierungsnorm zu vereinbaren. Die Ausbeutung anderer Individuen ist gerade nicht eine verallgemeinerbare Maxime, da es nicht möglich ist, daß alle zugleich Ausbeuter anderer Individuen sind. Vielmehr versucht die Universalisierungsnorm gerade Situationen wie jene zu verhindern, welche die diskriminierenden Regeln herbeiführen.

Nach Hobbes ist auch die Status-quo-Regel  $r_q$  nicht universalisierbar; erst das Verlassen des bürgerkriegsähnlichen Anarchiezustands verwirkliche die biblische Goldene Regel.<sup>32</sup> Dennoch könnte man sich auf den Standpunkt stellen, daß das Verhalten, welches  $r_q$  verlangt, durchaus verallgemeinerbar sei, weil hier alle Individuen *gleichermaßen* defektieren und daher niemand einen Vorteil gegenüber einem anderen Beteiligten erziele. Selbst in diesem Fall aber wäre die Wahl von  $r_q$  unter dem Schleier des Nichtwissens nicht zu erwarten, da

---

stitutionellen - „wohlgeordneten“ - Gesellschaft vollständig (d.h. ohne Durchsetzungskosten) reguliert, so wählen die Individuen im Urzustand mit einer Gerechtigkeitsvorstellung zugleich auch eine konkrete Verhaltensregel für ein vorhergesehenes postkonstitutionelles Problem, etwa für das unterstellte Gefangenendilemma. Die hier gewählte Vorgehensweise, das Entscheidungsproblem unter dem Schleier des Nichtwissens als Wahl konkreter Verhaltensregeln zu modellieren, stellt daher keinen Widerspruch zu Rawls' Theorie dar.

<sup>31</sup> Daß die ethische Universalisierungsnorm im Gefangenendilemma die Wahl der kooperativen Strategie verlangt, behaupten für Kants kategorischen Imperativ Sen (1982 [1972]), S. 76, Hegselmann (1988), S. 13, ders. (1989), S. 23, und Binmore (1994), S. 301. Für die goldene Regel behaupten dies Axelrod (1984), S. 136 f., Lottenbach (1988), S. 34, und Dixit/Nalebuff (1991), S. 106. Bei Buchanan/Congleton (1998), S. 66, findet sich eine ähnliche Behauptung für ihr Allgemeinheitsprinzip.

<sup>32</sup> Vgl. Hobbes (1976 [1651]), S. 100.

eine Entscheidung für die Durchsetzung des Status quo nicht rational wäre: Alle Individuen könnten besser stehen, wenn sie ihre Selbstbindungsmacht dazu verwendeten, die gegenüber  $r_q$  paretoverbessernde Regel  $r_f$  einzuführen.

Auch unter dem nicht-normativ begründeten Buchananschen Schleier der Unsicherheit ist zu erwarten, daß sich die Individuen darauf einigen werden, ein antizipiertes postkonstitutionelles Gefangenendilemma durch die Durchsetzung der Situation allgemeiner Kooperation - der „fairen Regel“ - zu überwinden. Daß im Gefangenendilemma als „fair“ (im Sinne der „Allgemeinheitsprinzip“ genannten Universalisierungsnorm) allein die Situation allgemeiner Kooperation anzusehen ist, steht in Buchanans neuestem, gemeinsam mit Roger Congleton verfaßten Buch außer Frage. Der ganze erste Teil von „Politics by Principle, Not Interest“ beschäftigt sich damit, durch welche konstitutionellen Entscheidungen man in Spielen, welche formal eine Gefangenendilemma-Struktur haben, die Situationen „jenseits der Diagonalen“ (die „*off-diagonals*“) - in der hier verwendeten Terminologie: diskriminierende Regeln - eliminieren könne, in welchen jeweils eine Seite auf Kosten der anderen bevorzugt wird.

Auf der konstitutionellen Ebene  $t_c$  wird nur diejenige Regel gewählt, die einstimmig angenommen wird. Das konstitutionelle Spiel  $K_n$  ist daher ein *Einstimmigkeitsspiel*<sup>33</sup>, in dem es nur dann zu einer für alle Spieler vorteilhaften Einigung kommt, wenn sie sich auf dieselbe Regelalternative  $r \in M^n$  einigen können, d.h. wenn alle Spieler die gleiche konstitutionelle Strategie wählen. Verfehlen die Spieler eine einstimmige Entscheidung, so wird überhaupt keine Regel gewählt. Dann wird eines der  $n!$  erwarteten Gefangenendilemmata gespielt, und das paretoinferiore Nash-Gleichgewicht  $(D_1, D_2, \dots, D_n)$  dieses Gefangenendilemmas stellt sich ein.

In Normalform lautet das konstitutionelle Spiel  $K$ :

$$(5.2) \quad K_n = (M_A, M_B, \dots, M_n; W_A(\rho), W_B(\rho), \dots, W_n(\rho)).$$

Die Strategienmenge  $M_i = \{r_{if}, r_{iq}, r_{id1}, r_{id2}, \dots, r_{id(2^n-2)}\}$  eines jeden Spielers  $i$  enthält sovielle reine Strategien  $r_i$ , wie es im postkonstitutionellen Gefangenendilemma Strategienkombinationen (Regeln) gibt, also  $2^n$ : Mit  $r_{if}$  wählt  $i$  die faire Regel allseitiger Kooperation  $r_f$ , mit  $r_{iq}$  die Status-quo-Regel allgemeiner Defektion  $r_q$  und mit  $r_{idk}$  eine diskriminierende Regel  $r_{idk} \in M^n \setminus \{r_f, r_q\}$ . Als eine weitere,  $(2^n + 1)$ -te Strategie eines jeden Spielers  $i$  wäre vorstellbar, überhaupt keine Regel zu wählen und damit für den Verbleib im Status quo zu votieren. Dann würde das postkonstitutionelle Gefangenendilemma gespielt, und das paretoinferiore Nash-Gleichgewicht stellte sich ein. Weil diese Strategie jedoch auszahlungsäquivalent mit  $r_{iq}$  wäre, wird sie nicht explizit modelliert.

<sup>33</sup> Zum Begriff der Einstimmigkeitsspiele vgl. *Harsanyi/Selten* (1988), S. 213-216.

$\rho \in \mathbf{M}^n$  bezeichnet eine Strategienkombination des konstitutionellen Spiels (mit  $\mathbf{M}^n = M_A \times M_B \times \dots \times M_n$ ). Die Menge  $\mathbf{M}^n$  umfaßt zum einen die *Einigungspunkte (Regeln)*  $\bar{r} \in \bar{\mathbf{M}}^n \subset \mathbf{M}^n$  des konstitutionellen Spiels, d.h. solchen Strategienkombinationen, in denen alle Spieler die gleiche konstitutionelle Strategie wählen. Jedem dieser Einigungspunkte  $\bar{r}$  entspricht eine Strategienkombination  $r \in M^n$  des (unter dem Schleier der Unkenntnis entpersonalisierten) Gefangenendilemmas  $G_n$ , die mittels einer Strategie  $r_i \in M_i$  des konstitutionellen Spiels gewählt wird. Aus Vereinfachungsgründen werde ich sprachlich daher nicht zwischen den Strategienkombinationen  $r \in M^n$  im postkonstitutionellen Spiel und den ihnen entsprechenden Strategienkombinationen  $\bar{r} \in \bar{\mathbf{M}}^n \subset \mathbf{M}^n$  im konstitutionellen Spiel unterscheiden; in beiden Fällen werde ich einfach von „Regel“ sprechen. Ist beispielsweise von der fairen Regel die Rede, so ist stets zugleich die Strategienkombination  $r_f = (C_1, C_2, \dots, C_n)$  des postkonstitutionellen Spiels wie die Strategienkombination  $\bar{r}_f = (r_{Af}, r_{Bf}, \dots, r_{nf})$  des konstitutionellen Spiels gemeint.

Zum anderen enthält die Menge  $\mathbf{M}^n$  neben den Regeln  $\bar{r}$  auch die Menge aller konstitutionellen Strategienkombinationen  $\rho \in \mathbf{M}^n \setminus \bar{\mathbf{M}}^n$ , in denen nicht alle Entscheidungsträger die gleiche Strategie wählen; diese konstitutionellen Strategienvektoren heißen *Nichteinigungspunkte*.

Die konkrete Gestalt des konstitutionellen Einstimmigkeitsspiels hängt von der jeweiligen sozialen Wohlfahrtsfunktion ab, mit der die konstitutionelle Unkenntnis modelliert wird, und von der aus dieser Funktion resultierenden Auszahlungsfunktion  $W_i(\rho)$ . In den verbleibenden Abschnitten von § 5 wird gezeigt, daß der Rawlssche Schleier des Nichtwissens dann, wenn das postkonstitutionelle Problem, das es im Urzustand zu lösen gilt,  $G_2$  oder  $G_n$  ist, eine Ursache der *Unfairneß* des Abstimmungsergebnisses sein kann (Abschnitt C.). In Abschnitt D. wird dargelegt, daß ein ähnliches Ergebnis sich auch für den Buchananschen Schleier der Unsicherheit herleiten läßt; darüber hinaus wird argumentiert, daß der Buchanansche Schleier der Unsicherheit nicht einmal garantieren kann, daß überhaupt ein einstimmiges Abstimmungsergebnis zustande kommt.

### C. „Gerechtigkeit als Unfairneß“:

#### Das konstitutionelle Spiel unter Rawls' Schleier des Nichtwissens

Rawls unterstellt, daß jeder Spieler  $i$  - in Unkenntnis seiner eigenen postkonstitutionellen Position und unfähig, subjektive Wahrscheinlichkeiten hierfür anzugeben - eine Maximin-Wohlfahrtsfunktion maximiert, die sich in Abhängigkeit von den als allgemein bekannt und interpersonell vergleichbar ange-



nommenen Grundgüterpayoffs ergibt. Da jeder Spieler  $i$  jeder Position  $j$  genau den „wahren“, objektiv ermittelbaren Grundgüteranteil des jeweiligen Positionsinhabers zuweist, sind die von allen Individuen vorgenommenen Bewertungen der jeweiligen Positionen identisch. Alle Individuen maximieren dann per definitionem die gleiche soziale Wohlfahrtsfunktion, so daß - vom Fall konstitutioneller Indifferenzspiele abgesehen<sup>34</sup> - eine Einigung unter dem Schleier des Nichtwissens annahmegemäß garantiert ist. Auf die Problematik des hinter dieser Annahme stehenden Grundgüterkonzepts wurde in § 4 bereits aufmerksam gemacht.

Für das konstitutionelle  $n$ -Personen-Spiel unter Rawls' Schleier des Nichtwissens,  $K_n^R(\rho)$  ergibt sich die folgende für alle Spieler identische Auszahlungsfunktion  $W^R(\rho)$  (mit Index „ $R$ “ für „Rawls“):

$$(5.3) \quad W^R(\rho) = \begin{cases} \min \{ \pi_1(r), \pi_2(r), \dots, \pi_n(r) \}, & \text{falls } \rho = \bar{r} \in \bar{M}^n \\ \min \{ g_1(0), \dots, g_n(0) \} & \text{sonst} \end{cases}$$

Das bedeutet, daß jeder Spieler  $i$  auf der konstitutionellen Ebene sich für jede Regel  $r$  in alle postkonstitutionellen Positionen  $j$  (mit  $j = 1, 2, \dots, n$ ) versetzt und dieser, je nachdem, ob der jeweilige Positionsinhaber kooperiert oder defektiert, den Grundgüteranteil  $f_j[v(r)]$  bzw.  $g_j[v(r)]$  zuweist und die Regel  $r$  mit dem geringsten aller Positionspayoffs bewertet; dies ist der konstitutionelle Payoff, wenn alle die gleiche Regel wählen.

Wählen die Individuen auf der konstitutionellen Ebene hingegen nicht alle die gleiche Strategie  $r_i$ , so wird eines der  $n!$  erwarteten postkonstitutionellen Gefangenendilemma-Szenarien gespielt, und das Nash-Gleichgewicht stellt sich ein. Alle Nichteinigungspunkte  $\rho \in M^n \setminus \bar{M}^n$ , des konstitutionellen Spiels können somit mit dem minimalen Payoff der Status-quo-Regel  $\bar{r}_q$  bewertet werden:  $W^R(\bar{r}_q) = \min \{ g_1[0], \dots, g_n[0] \}$ .

Das konstitutionelle Spiel unter dem Rawlsschen Schleier des Nichtwissens,  $K_n^R$ , läßt sich anschaulich darstellen für den Fall  $n = 2$ . Gemäß (5.4) versetzt sich dann jeder Spieler  $i$  (mit  $i = A, B$ ) in die Schuhe eines jeden Positionsinhabers  $j$  (mit  $j = 1, 2$ ) und bewertet jede gesellschaftliche Position mit dessen Präferenzen. Ist Spieler  $i$  also in Position 1, so bewertet er das (als bekannt unterstellte) Gefangenendilemma mit dem Grundgüterindex des Spielers in Position 1; ist er hingegen in der zweiten Position, so wird er Bewertungen mit dem Grundgüterindex des zweiten Spielers vornehmen. Die unter einem Rawlschen Schleier des Nichtwissens erwarteten GD-Szenarien unterscheiden sich

<sup>34</sup> Siehe dazu näher unten § 5 E. III.

somit nur durch die Bezeichnung der Spieler in der jeweiligen Position (Abbildung 5.3).

		B2:				A2:	
		$C_{B2}$	$D_{B2}$			$C_{A2}$	$D_{A2}$
A1:	$C_{A1}$	$f_1[2], f_2[2]$	$f_1[1], g_2[1]$	B1:	$C_{B1}$	$f_1[2], f_2[2]$	$f_1[1], g_2[1]$
	$D_{A1}$	$g_1[1], f_2[1]$	$g_1[0], g_2[0]$		$D_{B1}$	$g_1[1], f_2[1]$	$g_1[0], g_2[0]$

mit  $g_j[1] > f_j[2] > g_j[0] > f_j[1]$  und  $j = 1, 2$  (GD-Bedingung)

(a) (b)

Abbildung 5.3:  $G_2$ -Szenarien 1 und 2

Gemäß (5.2) und (5.3) ergibt sich das in Abbildung 5.4 gegebene konstitutionelle Zwei-Personen-Spiel  $K_2^R$ , in dem die Zellen  $r \in M^2$  des zugrunde liegenden Gefangenendilemmas als Strategien gewählt werden. Auf der Diagonalen von  $K_2^R$  befinden sich entsprechend die vier Regeln  $\bar{r} \in \bar{M}^2$ , die aus konstitutioneller Sicht jeweils mit dem Minimum des zugehörigen Auszahlungsvektors bewertet werden. Verfehlen die Spieler eine Einigung, so wird eines der beiden Gefangenendilemma-Szenarien gespielt. Jeder Nichteinigungspunkt  $\rho \in M^2 \setminus \bar{M}^2$  kann daher für jeden Spieler  $i$  mit den minimalen Payoffs,  $\min\{g_1[0], g_2[0]\}$ , des Nash-Gleichgewichts dieses GDs bewertet werden. Welche Gleichgewichte hat das konstitutionelle  $n$ -Personen-Spiel  $K_n^R$ ? Die faire Regel  $\bar{r}_f$  ist stets ein *echtes Koordinationsgleichgewicht* im Sinne von Lewis.<sup>35</sup> Ein Koordinationsgleichgewicht ist eine Strategienkombination, bei der kein Spieler wünscht, daß *irgendein* Spieler - weder er selbst noch ein anderer - einseitig von seiner gewählten Strategie abweicht. Denn der Spieler, der allein abzuweichen wünscht, wird nur dann im Vergleich zur Ausgangssituation bessergestellt, wenn beide sich auf ein bestimmtes Gleichgewicht zu einigen vermögen. Ein „echtes“ (striktes) Koordinationsgleichgewicht liegt vor, wenn diese Strategienkombination von jedem Spieler strikt vorgezogen wird, gegeben die Wahl der anderen Spieler.

<sup>35</sup> Vgl. Lewis (1975 [1969]), S. 14 und S. 21.

B:

	$r_{Bf} = (C_1, C_2)$	$r_{Bd1} = (D_1, C_2)$	$r_{Bd2} = (C_1, D_2)$	$r_{Bq} = (D_1, D_2)$
$r_{Af} = (C_1, C_2)$	$W_i^R(\bar{r}_f) = \min\{f_1[2], f_2[2]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$
$r_{Ad1} = (D_1, C_2)$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_{d1}) = \min\{g_1[1], f_2[1]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$
A:				
$r_{Ad2} = (C_1, D_2)$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_{d2}) = \min\{f_1[1], g_2[1]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$
$r_{Aq} = (D_1, D_2)$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$	$W_i^R(\bar{r}_q) = \min\{g_1[0], g_2[0]\}$

mit  $i = A, B$

Abbildung 5.4: Konstitutionelles Spiel  $K_2^R$  unter Rawls' Schleier des Nichtwissens

Diese Bedingung ist im vorliegenden Fall erfüllt, da gemäß der GD-Definition (5.1) stets gilt:  $W_i^R(\bar{r}_f) > W_i^R(\bar{r}_q)$ .<sup>36</sup> Gilt für eine diskriminierende Regel  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  überdies  $W_i^R(\bar{r}_d) \geq W_i^R(\bar{r}_q)$ , so ist auch  $\bar{r}_d$  ein Koordinationsgleichgewicht von  $K_n^R$ .<sup>37</sup>

Die Strategie  $r_{if}$ , mit welcher jeder Spieler  $i$  die faire Regel  $r_f$  wählt, dominiert (schwach) aufgrund der Definition des Gefangenendilemmas die Status-quo-Strategie  $r_{iq}$  für jede mögliche Auszahlungskonstellation, da  $W_i(r_{if}, r_{-i}) = W_i(r_{iq}, r_{-i})$  für alle  $(r_{if}, r_{-i}) \neq \bar{r}_f$  und  $W_i^R(r_{if}, r_{-if}) > W_i^R(r_{iq}, r_{-if})$  mit  $(r_{if}, r_{-if}) = \bar{r}_f$  (mit  $r_{-i}$  als Strategienvektor aller Spieler ohne  $i$ ). Unter bestimmten Payoff-Konstellationen werden auch alle übrigen Strategien  $r_i \in M_i \setminus \{r_{if}, r_{iq}\}$  schwach von  $r_{if}$  dominiert, so daß das konstitutionelle Spiel dominanzlösbar ist.<sup>38</sup>

Da beide Spieler im konstitutionellen Spiel unter dem Schleier des Nichtwissens stets identische konstitutionelle Präferenzordnungen haben (weshalb im folgenden auf eine Indizierung mit  $i$  verzichtet wird), läßt sich  $K_n^R$  immer mit

<sup>36</sup> Nach (5.1) gilt für alle Positionsinhaber  $j$ :  $f_j[n] > g_j[0]$ , so daß auch:

$\min\{f_1[n], \dots, f_n[n]\} > \min\{g_1[0], \dots, g_n[0]\} \Leftrightarrow W^R(\bar{r}_f) > W^R(\bar{r}_q)$ .

<sup>37</sup> Unter bestimmten Payoff-Konstellationen existieren auch (vollständig und unvollständig) gemischte Gleichgewichte des konstitutionellen Spiels, die, da sie in einer konstitutionellen Betrachtung unplausibel erscheinen, im folgenden nicht weiter betrachtet werden.

<sup>38</sup> Vgl. dazu unten § 5 E. I.

*Pareto-Dominanz-* (bzw. *Auszahlungsdominanz-*)Überlegungen lösen, die dem vertragstheoretischen Konsensprinzip genau entsprechen.<sup>39</sup> Weitergehende Rationalitätsanforderungen, wie sie die spieltheoretischen Gleichgewichtsauswahltheorien vorschlagen<sup>40</sup>, werden im folgenden außer Betracht bleiben, da ihre Berücksichtigung über die Rawlsschen (wie Buchananschen) Annahmen hinausginge.

Rawls betrachtet jede unter einem Schleier des Nichtwissens getroffene Vereinbarung als „gerecht“; von der Gerechtigkeit der Situation, in der die Entscheidung fällt, schließt er auf die Fairneß der Entscheidungen:

„[W]ir nehmen an, daß die Fairneß der Bedingungen, unter denen die Übereinkunft getroffen wird, sich auf die Gerechtigkeitsgrundsätze überträgt; da der Urzustand freie und gleiche moralische Personen in eine Situation gegenseitiger Fairneß bringt, ist *jede* von ihnen angenommene Gerechtigkeitskonzeption ebenfalls fair; daher der Name: *Gerechtigkeit als Fairneß*.“<sup>41</sup>

Diese Aussage ist, wenn, wie bei Rawls, Gerechtigkeit im Sinne der Universalisierungsnorm verstanden wird, (logisch) falsch:

*Satz 5.1: Es gibt Payoff-Konstellationen, für die unter Rawls' Schleier des Nichtwissens eine einstimmige Wahl der diskriminierenden Regel  $\bar{r}_d \in \overline{M} \setminus \{\bar{r}_f, \bar{r}_q\}$  das Ergebnis des konstitutionellen Spiels ist.*

Um die von Satz 5.1 behauptete Möglichkeit von Unfairneß unter dem Schleier des Nichtwissens und damit die logische Falschheit der von Rawls behaupteten Allaussage zu beweisen, genügt der Nachweis eines einzigen Beispiels, in dem unter den Rawlsschen Situations- und Rationalitätsannahmen eine diskriminierende, nicht aber eine faire Entscheidung zustande kommt. Im folgenden werden je ein Zahlenbeispiel für den Zwei- und eines für den Multi-Personen-Fall konstruiert, deren einzige Funktion es ist aufzuzeigen, daß die Bedingungen des Rawlsschen Urzustands auch Implikationen haben können, welche den in der „Theorie der Gerechtigkeit“ behaupteten Folgerungen widersprechen.

Betrachten wir zunächst das in Abbildung 5.5a gegebene Zwei-Personen-Gefangenendilemma. Wie ich oben in Abschnitt A. argumentiert habe, spricht nichts gegen die Annahme eines solchen GD-Spiels als im Urzustand zu lösendes Problem. Im resultierenden konstitutionellen Spiel  $K_2^R$  (Abbildung 5.5b)

<sup>39</sup> Vgl. zum Konzept der Pareto- (bzw. Auszahlungs-)Dominanz *Harsanyi/Selten* (1988), S. 80 ff.

<sup>40</sup> Zu denken wäre etwa an das Konzept der Risikodominanz von *Harsanyi/Selten* (1988) oder das Kriterium der Widerstandsdominanz von *Güth/Stephan* (1984).

<sup>41</sup> *Rawls* (1992 [1980]), S. 90 (erste Hervorhebung von mir; *C. M.*); vgl. ähnlich z. B. *Rawls* (1994 [1971]), S. 29; *ders.* (1992 [1985]), S. 274.

wird dennoch die faire Regel  $\bar{r}_f$  nicht gewählt. Nach dem Pareto-Dominanz-Kriterium werden sich die konstitutionellen Entscheidungsträger rationalerweise auf die diskriminierende Regel  $\bar{r}_{d1}$  einigen, obwohl diese im Beispiel - gemessen am zugrunde liegenden Universalisierungsprinzip - zu unfairen Ergebnissen führt. Denn zum einen ist die postkonstitutionelle Verteilung der Payoffs in  $\bar{r}_{d1}$  weit weniger ausgewogen als unter der fairen Regel  $\bar{r}_f$ . Und zum anderen wird - bei Geltung von Regel  $\bar{r}_{d1}$  - der Spieler, der sich in Position 2 des postkonstitutionellen Spiels wiederfindet, sogar gegenüber der Situation der Regellosigkeit schlechter gestellt ( $91 < 92$ ). Während also die Person in Position 2 ohne Regel gegenüber ihrem Gegenspieler leicht im Vorteil gewesen wäre, endet sie mit dem unter dem Schleier des Nichtwissens geschlossenen Verfassungsvertrag als das am schlechtesten gestellte Individuum in einer extrem diskriminierenden Gesellschaft. Universalisierbar, wie behauptet, ist dieses Entscheidungsergebnis nicht.

2:

		$C_2$	$D_2$
1:	$C_1$	$r_f$ 90, 93	$r_{d2}$ 70, 120
	$D_1$	$r_{d1}$ 120, 91	$r_q$ 89, 92

(a)

B:

		$r_{Bf}$	$r_{Bd1}$	$r_{Bd2}$	$r_{Bq}$
	$r_{Af}$	$\bar{r}_f$ 90, 90	89, 89	89, 89	89, 89
A:	$r_{Ad1}$	89, 89	$\bar{r}_{d1}$ 91, 91	89, 89	89, 89
	$r_{Ad2}$	89, 89	89, 89	$\bar{r}_{d2}$ 70, 70	89, 89
	$r_{Aq}$	89, 89	89, 89	89, 89	$\bar{r}_q$ 89, 89

(b)

Abbildung 5.5: Einstimmige Wahl der diskriminierenden Regel  $\bar{r}_{d1}$  im konstitutionellen Spiel  $K_2^R$  unter Rawls' Schleier des Nichtwissens

Als zweites Beispiel zur Demonstration des Ergebnisses von Satz 5.1 sei ein konstitutionelles Spiel auf der Basis eines Elf-Personen-Gefangenendilemmas ( $n = 11$ ) betrachtet, wobei der Einfachheit halber unterstellt wird, daß es nur zwei nicht-leere Teilklassen  $a$  und  $b$  von Positionsinhabern  $j$  mit jeweils identi-

schen Präferenzen gibt:  $j \in I_{11}$  und  $I_{11} = I_a \cup I_b$ . Das postkonstitutionelle Gefangenendilemma sei durch die folgende Payoff-Funktionen präzisiert:

$$(5.4) \quad \pi_j(r) = \begin{cases} g_a[v(r)] = 45 + 30(v(r))^{0,5} & , \text{ falls } j \text{ defektiert} \\ f_a[v(r)] = 10 + (v(r))^{1,5} & , \text{ falls } j \text{ kooperiert} \end{cases} \quad \text{für alle } j \in I_a \text{ und}$$

$$\pi_j(r) = \begin{cases} g_b[v(r)] = 85 + 30(v(r))^{0,5} & , \text{ falls } j \text{ defektiert} \\ f_b[v(r)] = 50 + (v(r))^{1,5} & , \text{ falls } j \text{ kooperiert} \end{cases} \quad \text{für alle } j \in I_b.$$

Wie sich leicht nachprüfen läßt, erfüllen die in (5.4) gegebenen Auszahlungsfunktionen der Gruppen  $a$  und  $b$  von Positionsinhabern die durch (5.1) gegebenen Bedingungen eines Multi-Personen-Gefangenendilemmas: Zum einen gilt das Dominanzprinzip, nach dem für alle  $0 \leq v(r) \leq n-1$  gelten muß:  $g_j[v(r)] > f_j[v(r)+1]$  für  $j = a, b$ . Zum anderen ist für alle Positionsinhaber der Payoff der Situation allgemeiner Kooperation größer als die Auszahlung in der Situation allgemeiner Defektion, d.h.  $f_a[11] = 46 > g_a[0] = 45$  und  $f_b[11] = 86 > g_b[0] = 85$ , so daß die Strategienkombination  $r_f$  die Situation  $r_q$  pareto-dominiert.<sup>42</sup> Schließlich sind die Auszahlungsfunktionen, wie verlangt, streng monoton steigend.

Alle elf Spieler  $A, B, \dots, K$  im Urzustand kennen diese Payoff-Funktionen der postkonstitutionellen Positionsinhaber. Unter dem Schleier der Unwissenheit wissen sie jedoch nicht, in welcher postkonstitutionellen Position und damit in welcher Gruppe von Positionsinhabern ( $a$  oder  $b$ ) sie sich wiederfinden werden. Da die Mitglieder von Gruppe  $b$  nach (5.4) für alle  $v(r)$  eine um 40 Einheiten höhere Auszahlung erzielen als die Mitglieder von Gruppe  $a$ , kalkuliert jeder Spieler nach Rawls als konstitutionellen Payoff für die faire Regel  $\bar{r}_f$  die Auszahlung der Mitglieder von Gruppe  $a$ :

$$W^R(\bar{r}_f) = f_a[11] = 46.$$

Demgegenüber sei  $\bar{r}_a \in \overline{M}^{11} \setminus \{\bar{r}_f, \bar{r}_q\}$  eine diskriminierende Regel, welche die Defektion aller Mitglieder der Gruppe  $a$  und die Kooperation aller Gruppenangehörigen von  $b$  erzwingt. Die Gruppe  $a$  umfasse  $x$  Mitglieder (mit  $1 = x = 10$ ). Die Payoffs der Positionsinhaber, die Gruppe  $a$  angehören, betragen bei Geltung der diskriminierenden Regel  $\bar{r}_a$  folglich jeweils  $g_a[11-x] = 140$ , während die der Gruppe  $b$  angehörenden, zur Kooperation gezwungenen Positionsinhaber unter  $\bar{r}_a$  je eine Auszahlung in Höhe von  $f_b[11-x] = 82$  erzielen. Es läßt sich zeigen, daß für jeden Wert von  $x$   $g_a[11-x] > f_b[11-x]$ ; daher kalkulieren die Spieler im Urzustand unter dem Schleier des Nichtwissens als konstitutionelle

<sup>42</sup> Diese und alle folgenden Funktionswerte sind der Einfachheit halber auf ganze Zahlen gerundet.

Auszahlung von  $\bar{r}_a$  den Payoff der schlechtergestellten Mitglieder von Gruppe  $b$ , die im schlechtesten Fall (für  $x = 10$ )  $f_b[1] = 51$  erhalten:

$$W^R(\bar{r}_a) = f_b[11-x] = 51.$$

Vergleicht man die beiden konstitutionellen Payoffs für die faire und die diskriminierende Regel, so gilt für alle Größen von  $x$ :

$$W^R(\bar{r}_a) = f_b[11-x] = 51 \quad > \quad W_R(\bar{r}_f) = f_a[11] = 46.$$

Die diskriminierende Regel  $\bar{r}_a$  paretodominiert daher die faire Regel  $\bar{r}_f$  unter dem Schleier des Nichtwissens; unter den hier gemachten Monotonieannahmen im zugrunde liegenden Gefangenendilemma paretodominiert sie auch jede andere Regel sowie jeden Nichteinigungspunkt, bei dem es eine konstitutionelle Auszahlung von  $W^R(\bar{r}_a) = \min\{g_a[0], g_b[0]\} = g_a[0] = 45$  gibt. Die diskriminierende Regel  $\bar{r}_a$  ist damit das einstimmige Ergebnis der konstitutionellen Regelsentscheidung.

Zu beachten ist, daß dieses Ergebnis für jeden Wert von  $x$  gilt (mit  $1 = x = 10$ ). Je nach zugrunde gelegten Präferenzverteilungen, ist daher jede denkbare Art der Diskriminierung unter dem Schleier des Nichtwissens möglich. Das Beispiel zeigt insbesondere, daß die Existenz einer Rawlsschen konstitutionellen Unkenntnis auch keine Garantie gegen die in modernen Gesellschaften wohl typischste Form der Diskriminierung, eine „*Tyrannie der Mehrheit*“, darstellt. Für bestimmte  $x$  führen die postkonstitutionelle Kollektiventscheidung nach der Entscheidungsregel der einfachen Mehrheit und die konstitutionelle Entscheidung unter dem Schleier des Nichtwissens im Beispiel zum gleichen diskriminierenden Entscheidungsergebnis.

Um dies zu zeigen, nehme man an, daß im Beispiel die Gruppe  $a$  sechs Mitglieder umfasse ( $x = 6$ ). Auf der *konstitutionellen* Ebene unter dem Schleier des Nichtwissens wählen die Spieler, wie gezeigt, einstimmig die Regel  $\bar{r}_a$ , welche die fünf Positionsinhaber aus Gruppe  $b$  zur Kooperation und die Mitglieder der Gruppe  $a$  zur Defektion zwingt. In der *postkonstitutionellen* Gesellschaft verfügen die sechs Mitglieder von Gruppe  $a$  bei Kollektiventscheidungen über die absolute Mehrheit der Stimmen. Diese Mehrheitskoalition braucht daher die Interessen der Mitglieder aus Gruppe  $b$  nicht zu beachten und kann jene Strategienkombination des Elf-Personen-Gefangenendilemmas herbeiführen, welche die gemeinsamen Interessen der Mitglieder der Mehrheit maximiert. Strategienkombinationen, die aus Sicht der Mehrheitskoalition jedes andere Spielergebnis paretodominieren, werden auch „*mehrheitsdominant*“ genannt.<sup>43</sup>

<sup>43</sup> Zum Begriff der Mehrheitsdominanz vgl. *Buchanan/Congleton* (1998).



Im Beispiel ist die mehrheitsdominante Situation jene, in der alle sechs Mitglieder der Mehrheitsgruppe  $a$  defektieren und alle fünf Mitglieder der Minderheit (Gruppe  $b$ ) zur Kooperation gezwungen werden, d.h. die diskriminierende Regel  $r_a$ . Denn jedes Mitglied der Mehrheitskoalition erhält in dieser Situation einen Payoff von jeweils  $g_a[5] = 112$ . Eine höhere Anzahl von Menschen zur Kooperation zu zwingen ( $v = 6$ ), wäre nach Maßgabe der Entscheidungsregel der einfachen Mehrheit nicht möglich. Eine geringere Anzahl von Individuen auszubeuten ( $v = 4$ ), würde demgegenüber jedem Mitglied von  $a$  einen geringeren Payoff einbringen als im Falle der Ausbeutung von fünf Personen, nämlich  $g_a[v = 4] = 105$ .

Auch ein völliger Verzicht auf die Ausbeutung anderer - d.h. die Wahl der fairen Regel  $r_f$  - wäre nach dem Kriterium der Mehrheitsdominanz nicht rational, weil die Mitglieder der Mehrheitskoalition in der Situation allgemeiner Kooperation einen geringeren Payoff erzielen als  $g_a[5] = 112$ , nämlich nur in Höhe von  $f_a[11] = 46$ .

Die Kosten dieser Wahl der mehrheitsdominanten Situation  $r_a$  tragen die Mitglieder der Minderheit  $b$ , die in der von der Mehrheit herbeigeführten Situation eine Auszahlung in Höhe von  $f_b[5] = 61$  erzielen, die nicht nur erheblich geringer ist als ihr Payoff bei Geltung der Regel allseitiger Kooperation  $r_f$ , der  $f_b[11] = 86$  beträgt; vielmehr werden die Mitglieder von  $b$  sogar noch schlechter gestellt als im Nash-Gleichgewicht des Gefangenendilemmas, in dem sie eine Auszahlung von  $g_b[0] = 85$  erhielten.

Das Beispiel zeigt, daß unter bestimmten Payoff-Konstellationen und Werte von  $x$  ein Schleier des Nichtwissens nicht einmal zu weniger unfairen Entscheidungen führt als eine so diskriminierende (postkonstitutionelle) Entscheidungsregel wie die einfache Mehrheit. Im gewählten Beispiel hat der Schleier des Nichtwissens die Ungleichheit gegenüber der Situation der Regellosigkeit sogar noch verschärft: Würde auf der konstitutionellen Ebene keine Regel für das postkonstitutionelle Elf-Personen-Gefangenendilemma mit den Auszahlungsfunktionen (5.4) gewählt, so stellte sich das Nash-Gleichgewicht dieses Spiels ein, in dem die Gruppe  $a$  einen Payoff von  $g_a[0] = 45$  und Gruppe  $b$  eine Auszahlung in Höhe von  $g_b[0] = 85$  erzielte; beide Gruppen trennte ein Abstand von 40 Grundgütereinheiten. Unter dem Schleier des Nichtwissens - nicht anders als in einer postkonstitutionellen Mehrheitsentscheidung - würde die diskriminierende Regel  $\bar{r}_a$  gewählt, welche die fünf Mitglieder von Gruppe  $b$  zur Ausbeutung durch Gruppe  $a$  freigibt. Jedes Mitglied von Gruppe  $a$  erhielte hier einen Grundgüterindex von  $g_a[5] = 112$ , während jedes Individuum von Gruppe  $b$  eine Auszahlung in Höhe von  $f_b[5] = 61$  erzielte. Die konstitutionelle Entscheidung unter dem Schleier des Nichtwissens hat die Kluft, welche die Begünstigten von den Schlechtergestellten trennt, sogar noch erhöht; statt, wie im

Fall ohne Regel, um 40 Grundgütereinheiten unterscheiden sich die Auszahlungen der Mitglieder der Gruppen  $a$  und  $b$  nun sogar um 51 Grundgütereinheiten.

Wie die voranstehenden Überlegungen zeigen, ist Rawls' Behauptung, daß im Urzustand unter dem Schleier des Nichtwissens stets faire Entscheidungen getroffen würden, (logisch) falsch. Ein Schleier des Nichtwissens kann die Fairneß der Entscheidungsfindung im Sinne der zugrunde gelegten Universalisierungsnorm nicht garantieren; seine Existenz impliziert nicht zwingend, daß für ein Gefangenendilemma die faire Regel  $r_f$  gewählt wird. Nicht einmal so extrem diskriminierende Gesellschaftsverträge, welche die Regelbefolgung allein einer einzigen Person auferlegen, können ausgeschlossen werden. „Gerechtigkeit als Unfairneß“, so läßt sich schließen, ist möglich unter Rawls' Schleier des Nichtwissens.

Damit soll gleichwohl nicht gesagt werden, daß die von Rawls behaupteten Implikationen im vorliegenden Modell niemals eintreffen, denn es gilt auch:

*Satz 5.2: Es gibt Payoff-Konstellationen, für die unter Rawls' Schleier des Nichtwissens eine einstimmige Wahl der fairen Regel  $\bar{r}_f$  das Ergebnis des konstitutionellen Spiels ist.*

2:

		$C_2$		$D_2$	
1:	$C_1$	$r_f$	90, 90	$r_{d2}$	70, 120
	$D_1$	$r_{d1}$	120, 70	$r_q$	80, 80

(a)

B:

		$r_{Bf}$		$r_{Bd1}$		$r_{Bd2}$		$r_{Bq}$	
A:	$r_{Af}$	$\bar{r}_f$	90, 90		80, 80		80, 80		80, 80
	$r_{Ad1}$		80, 80	$\bar{r}_{d1}$	70, 70		80, 80		80, 80
	$r_{Ad2}$		80, 80		80, 80	$\bar{r}_{d2}$	70, 70		80, 80
	$r_{Aq}$		80, 80		80, 80		80, 80	$\bar{r}_q$	80, 80

(b)

Abbildung 5.6: Einstimmige Wahl der fairen Regel  $\bar{r}_f$  im konstitutionellen Spiel  $K_2^R$  unter Rawls' Schleier des Nichtwissens

Interpretiert man die Behauptung, der Schleier des Nichtwissens führe zu fairen Einigungen im Urzustand, nicht als All-, sondern als eine Existenzaussage wie Satz 5.2, so ist diese logisch wahr; nichts spricht dennoch dafür, daß Rawls

eine in dieser Weise eingeschränkte Gültigkeit der Wirkungen des Schleiers des Nichtwissens beabsichtigte.

Um den Existenzsatz 5.2 zu beweisen, genügt wiederum ein einziges Beispiel. Betrachten wir dazu zunächst wieder ein konstitutionelles Zwei-Personen-Spiel,  $K_2^R$ . In Abbildung 5.6a sind die GD-Szenarien aus Abbildung 5.2 zu einem einzigen (entpersonalisierten) Gefangenendilemma-Spiel zusammengefaßt. Im entsprechenden konstitutionellen Spiel  $K_2^R$  (Abbildung 5.6b) pareto dominiert nicht nur die faire Regel  $\bar{r}_f$  jeden anderen Einigungs- und Nichteinigungspunkt; sie ist gleichzeitig das Gleichgewicht in schwach dominanten Strategien des konstitutionellen Spiels.

Als zweites Beispiel zur Demonstration des Ergebnisses von Satz 5.2 sei nochmals ein konstitutionelles Spiel auf der Basis eines Elf-Personen-Gefangenendilemmas ( $n = 11$ ) betrachtet, das sich von dem soeben behandelten Beispiel nur dadurch unterscheidet, daß alle elf postkonstitutionellen Positionsinhaber  $j$  die gleiche Auszahlungsfunktion wie zuvor die Mitglieder von Gruppe  $a$  haben.  $G_{11}$  sei also symmetrisch:

$$(5.5) \quad \pi_j(r) = \begin{cases} g[v(r)] = 45 + 30(v(r))^{0.5}, & \text{falls } j \text{ defektiert} \\ f[v(r)] = 10 + (v(r))^{1.5}, & \text{falls } j \text{ kooperiert} \end{cases}$$

Der konstitutionelle Payoff  $W^R(\bar{r}_f)$  bei Geltung der fairen Regel  $\bar{r}_f$  eines jeden Spielers  $i$  beträgt also:

$$W^R(\bar{r}_f) = 46.$$

Betrachten wir demgegenüber eine beliebige diskriminierende Regel  $\bar{r}_d \in \overline{M}^{11} \setminus \{\bar{r}_f, \bar{r}_q\}$ , unter der eine Anzahl von  $x$  Spielern defektiert (mit  $1 = x = 10$ ). Jeder Positionsinhaber  $j$ , der defektiert, erhält folglich die Auszahlung  $g[11-x]$ , während jeder kooperierende Positionsinhaber  $j$  einen Indexwert von  $f[11-x]$  erzielt. Der konstitutionelle Payoff für die diskriminierende Regel ist daher

$$W^R(\bar{r}_d) = \min \{f[11-x], g[11-x]\} = f[11-x] \leq 42.$$

Auch im für die unter  $\bar{r}_d$  Kooperierenden günstigsten Fall, in dem  $x = 1$  ist, unterschreitet der konstitutionelle Payoff der diskriminierenden Regel,  $f[10] = 42$ , also die Auszahlung für die faire Regel,  $W^R(\bar{r}_f) = 46$ . Damit pareto dominiert die faire Regel  $\bar{r}_f$  in  $K_{11}^R$  jede beliebige diskriminierende Regel  $\bar{r}_d \in \overline{M}^{11} \setminus \{\bar{r}_f, \bar{r}_q\}$ . Da aufgrund der Definition des Gefangenendilemmas

auch stets  $W^R(\bar{r}_f) = 46 > W^R(\bar{r}_q) = 45$  ist, ist das Koordinationsgleichgewicht  $\bar{r}_f$  im Beispiel die Auszahlungsdominanz-Lösung des konstitutionellen Spiels  $K_{11}^R$ .<sup>44</sup>

### D. Das konstitutionelle Spiel unter Buchanans Schleier der Unsicherheit<sup>45</sup>

Rawls' Annahmen über den Urzustand, in dem die Rationalakteure unter einem Schleier des Nichtwissens entscheiden, berechtigen, wie gezeigt, nicht zu der Annahme, daß die unter diesen Bedingungen getroffenen Entscheidungen stets fair im Sinne der Universalisierungsnorm sind; auch grob diskriminierende Regeln können darum von der Rawlsschen Theorie als „gerecht“ ausgewiesen werden. In diesem Abschnitt D. werde ich zeigen, daß die von Buchanan unterstellte Entscheidungssituation unter dem Schleier der Unsicherheit ganz ähnlichen Einwänden unterliegt wie der Urzustand in Rawls' „Theorie der Gerechtigkeit“.

Die konstitutionelle Entscheidungssituation in der Buchananschen Theorie des Gesellschaftsvertrags ist eine Risikosituation. Die Spieler sind in der Lage, den jeweiligen Positionen in der postkonstitutionellen Gesellschaft subjektive Wahrscheinlichkeiten zuzuordnen. Die Existenz eines Buchananschen Schleiers der Unsicherheit kann als ein (nicht zu beobachtender) „Zug der Natur“ vor Spielbeginn interpretiert werden, der die konstitutionellen Spieler den postkonstitutionellen Positionen im Gefangenendilemma „zulost“.<sup>46</sup>

Bezeichne  $\pi_{ij}(r)$  den kardinalen von-Neumann-Morgenstern-Nutzenwert des Spielers  $i$ , bei Geltung von Regel  $r$  in Position  $j$  zu sein, wobei wiederum  $r \in M^n$  der Menge aller möglichen Strategienkombinationen des zugrunde liegenden  $n$ -Personen-Gefangenendilemmas ist. Darüber hinaus stehe  $w_{ij}$  für die Wahrscheinlichkeit, mit der sich Spieler  $i$  in der postkonstitutionellen Position  $j$  wiederfindet. Unter Buchanans Schleier der Unsicherheit maximiert dann jeder Spieler  $i$  seinen Erwartungsnutzen aus verschiedenen Gesellschaftspositionen  $j$ .<sup>47</sup> Es wird im folgenden von einem „dicken“ Schleier der Unsicherheit ausge-

<sup>44</sup> Die faire Regel ist in diesem Fall auch das Gleichgewicht in sukzessiver Dominanz des konstitutionellen Spiels; vgl. dazu unten die Ausführungen zu Satz 5.6.

<sup>45</sup> Vgl. zu diesem Abschnitt auch Müller (1998a).

<sup>46</sup> Vgl. ähnlich Held (1977), S. 733 ff.; Frohlich (1992).

<sup>47</sup> Vgl. ähnlich für den Zwei-Personen-Fall Barbosa (1974), S. 86, ders. (1978), S. 35.

gangen, so daß jedes Individuum  $i$  jede der  $n$  Positionen in dem postkonstitutionellen Gefangenendilemma für gleich wahrscheinlich hält:

$$w_{i1} = w_{i2} = \dots = w_{in} = \frac{1}{n}.$$

Für jeden Spieler  $i$  ergibt sich somit die folgende konstitutionelle Auszahlungsfunktion  $W_i^B(\rho)$ , welche  $i$ 's „konstitutionelle Präferenz“ für jeden Einigungs- und Nichteinigungspunkt  $\rho \in \mathbf{M}^n$  des konstitutionellen Spiels angibt (mit Index „B“ für „Buchanan“):

$$(5.6) \quad W_i^B(\rho) = \begin{cases} \frac{1}{n} \sum_{j=1}^n \pi_{ij}(r), & \text{falls } \rho = \bar{r} \in \bar{M}^n \\ \frac{1}{n} \sum_{j=1}^n g_{ij}[0] & \text{sonst} \end{cases} \quad .^{48}$$

In Buchanans Urzustand versetzt sich also jeder Spieler  $i$  für alle  $\rho \in \mathbf{M}^n$  in eine jede postkonstitutionelle Position  $j$  (mit  $j = 1, \dots, n$ ) und bewertet diese, da

<sup>48</sup> Die (5.6) zugrunde liegende Gleichwahrscheinlichkeitsannahme impliziert *nicht*, daß jedes Individuum  $i$  risikoneutral ist, wie es von Kyriacou (1998) meiner Modellierung des Buchananschen Schleiers der Unsicherheit vorgeworfen wurde. Zur Berücksichtigung einer möglichen Risikoaversion des jeweiligen konstitutionellen Spielers, so diese Kritik, seien dessen mögliche Verluste unter den alternativen Regeln höher zu gewichten als seine Gewinne. Dieser Einwand beruht auf einer - wenn auch recht häufigen - Mißinterpretation der Erwartungsnutzentheorie (wie sie sich etwa auch bei Patainaik, 1968, S. 1158 f., und Sen, 1975 [1973], S. 32-34, findet). Verwechselt werden hier die Risikoaversion des Entscheidungsträgers in bezug auf sein *Einkommen* und seine Risikoaversion in bezug auf seinen *Nutzen* (vgl. Ng, 1984, S. 184). Beide Fragen sind strikt voneinander zu trennen. Der Nutzen eines Entscheiders für irgendein mögliches Geldeinkommen wird von seiner von-Neumann-Morgenstern-Nutzenfunktion gemessen. Ist er beispielsweise risikoavers in bezug auf Einkommen, so wird seine Nutzenfunktion - ebenso wie seine konstitutionelle Auszahlungsfunktion - strikt konkav in Geldeinkommen sein. Diese Nichtlinearität von  $i$ 's Auszahlungsfunktion in Abhängigkeit von Geldeinkommen impliziert indes nicht ihre Nichtlinearität in Nutzen. Die von-Neumann-Morgenstern-Nutzenwerte  $\pi_{ij}(r)$ , von denen die Auszahlungsfunktion (5.6) abhängig ist, berücksichtigen bereits die individuelle Risikoeigung von Spieler  $i$  auf der konstitutionellen Ebene. Eine Berücksichtigung zusätzlicher risikoabhängiger Gewichtungsfaktoren für  $i$ 's Gewinne oder Verluste in den jeweiligen postkonstitutionellen Gefangenendilemma-Positionen bedeutete daher eine logisch inkonsistente und die Regeln der Bayesianischen Entscheidungstheorie verletzende Doppelzählung von Risikoeinstellungen; vgl. ausführlich Harsanyi (1975), S. 318-324; Ng (1984), S. 183 f.; Mueller (1989a), S. 424-432; Jakob (1996), S. 179-191. Für eine ausführliche Replik auf Kyriacou vgl. Müller (1998b).

die Universalisierung, die der Schleier der Unsicherheit induziert, nur partiell ist, mit seinen *eigenen* Nutzenwerten  $\pi_{ij}(r)$  nach Maßgabe von Auszahlungsfunktion (5.1). Anders als bei Rawls sind daher intersubjektive Vergleiche von Wohlstands- (bzw. Grundgüter-)Niveaus nicht erforderlich.<sup>49</sup> Voraussetzung bleibt aber auch im konstitutionellen Spiel unter Buchanans Annahmen, daß die individuellen Nutzenfunktionen für jede Position  $j$  weiter die üblichen Eigenschaften von Gefangenendilemma-Präferenzen gemäß (5.1) aufweisen.

Einigen sich die Spieler auf eine Regel  $\bar{r} \in \overline{M}^n$ , so bewertet  $i$  Position  $j$  mit  $f_{ij}[v(r)]$ , wenn die Regel den Positionsinhaber zur Kooperation zwingt, und mit  $g_{ij}[v(r)]$ , wenn die Regel die Defektion des Positionsinhabers vorsieht. Wählen hingegen nicht alle Spieler die gleiche konstitutionelle Strategie, so wird das postkonstitutionelle  $n$ -Personen-Gefangenendilemma gespielt, und das Nash-Gleichgewicht  $r_q$  stellt sich ein. Jeder Nichteinigungspunkt  $\rho \in M^n \setminus \overline{M}^n$  des konstitutionellen Spiels kann daher mit dem konstitutionellen Nutzenwert  $W_i^B(\bar{r}_q)$  der Status-quo-Regel  $\bar{r}_q$  bewertet werden, der dem erwarteten Nutzen in allen Positionen des Nash-Gleichgewichts  $r_q$  entspricht.

Das konstitutionelle Spiel  $K_n^B$  unter Buchanans Schleier der Unsicherheit sei zunächst wiederum für die Konstellation  $n = 2$  dargestellt. Abbildung 5.7 stellt die beiden GD-Szenarien dar, die sich von jenen im Rawls-Fall nur dadurch unterscheiden, daß die beiden Spieler  $A$  und  $B$  mit ihren eigenen Nutzenfunktionen bewerten, wie sie selbst sich fühlen würden, hätten sie die Position 1 oder die Position 2 inne; die Auszahlungen der GD-Szenarien sind folglich mit den Indizes „A“ und „B“ versehen.

---

<sup>49</sup> Die hier gewählte Modellierung von *Buchanans* Schleier der Unsicherheit enthält als Spezialfall das Gleichwahrscheinlichkeitsmodell von *Harsanyi*. Ließe man nämlich die Möglichkeit intersubjektiver Nutzenvergleiche zu und nähme man, ausgehend von einem „dicken“ Schleier der Unsicherheit, an, daß alle Spieler  $i$  auf der konstitutionellen Ebene jeder Position  $j$  den gleichen intersubjektiv vergleichbaren („wahren“) Nutzenwert des Positionsinhabers zuwiesen, so erhielte man *Harsanys* utilitaristische Wohlfahrtsfunktion. Da folglich alle Individuen  $i$  auch die gleiche konstitutionelle Auszahlungsfunktion annehmen, treffen sie das gleiche „moralische Werturteil“ im Sinne von *Harsanyi*; unter *Harsanys* Variante des Schleiers der Unkenntnis ist es den Individuen also unmöglich, uneinig zu sein. Das Einstimmigkeitserfordernis unter dem Schleier der Unkenntnis ist daher, wie bei Rawls, für *Harsanys* Modell (unter der heroischen Annahme intersubjektiv vergleichbarer und allgemein bekannter Nutzenfunktionen aller postkonstitutionellen Positionsinhaber) stets erfüllt. Die Erfüllung der Bedingungen der vollkommenen Universalisierung ist - vom Fall reiner Koordinationsprobleme im konstitutionellen Spiel abgesehen (siehe dazu näher unten § 5 E. III.) - eine Garantie für das Erreichen von Einstimmigkeit unter dem Schleier der Unkenntnis; vgl. auch *Barbosa* (1978), S. 38; *Eichberger/Pethig* (1994), S. 320 (Proposition 3.1.).

		$B2:$		$A2:$	
		$C_{B2}$	$D_{B2}$	$C_{A2}$	$D_{A2}$
$A1:$	$C_{A1}$	$f_{A1}[2], f_{B2}[2]$	$f_{A1}[1], g_{B2}[1]$	$B1:$	$C_{B1}$
	$D_{A1}$	$g_{A1}[1], f_{B2}[1]$	$g_{A1}[0], g_{B2}[0]$		$D_{B1}$
		mit $g_{ij}[1] > f_{ij}[2] > g_{ij}[0] > f_{ij}[1]$ für alle $i = A, B$ und $j = 1, 2$ (GD-Bedingung)			
		(a)		(b)	

Abbildung 5.7:  $G_2$ -Szenarien 1 und 2

Nach Maßgabe von (5.2) und (5.5) ergibt sich das konstitutionelle Zwei-Personen-Spiel unter Buchanans Schleier der Unsicherheit,  $K_2^B$ , als ein Koordinationsspiel, in dem die vier wählbaren Regeln  $\bar{r} \in \overline{M}^2$ , die mit Hilfe der vier konstitutionellen Strategien  $r_{if}$ ,  $r_{id1}$ ,  $r_{id2}$  und  $r_{iq}$  gewählt werden können, auf der Diagonalen liegen (Abbildung 5.8). Verfehlen beide Spieler die Einstimmigkeit, so wird überhaupt keine Regel gewählt. Dann wird eines der beiden erwarteten Gefangenendilemmata gespielt, und das paretoinferiore Nash-Gleichgewicht dieses GDs stellt sich ein mit dem konstitutionellen Payoff  $W_i^B(\bar{r}_q) = (g_{i1}[0] + g_{i2}[0])/2$ .

Welche Eigenschaften hat das konstitutionelle  $n$ -Personen-Spiel  $K_n^B$  unter Buchanans Schleier der Unsicherheit? Die faire Regel  $\bar{r}_f = (r_{Af}, r_{Bf}, \dots, r_{nf})$  ist, wie im Rawlschen Urzustand, stets ein starkes Koordinationsgleichgewicht des konstitutionellen Spiels, da sich nach (5.1) jeder Spieler schlechter stellt, wenn nur ein einziges Individuum von diesem Einigungspunkt abweicht:

$$f_{ij}[n] > g_{ij}[0] \quad \text{für alle } j = 1, \dots, n$$

$$\Rightarrow W_i^B(\bar{r}_f) = \frac{1}{n} \sum_{j=1}^n f_{ij}[n] > W_i^B(\bar{r}_q) = \frac{1}{n} \sum_{j=1}^n g_{ij}[0] \quad \text{für alle } i = A, B, \dots, n$$

Die Status-quo-Regel  $\bar{r}_q$ , die auszahlungsäquivalent zu allen Nichteinigungspunkten  $\rho \in M^n \setminus \overline{M}^n$  ist, ist entsprechend stets ein Koordinationsgleichgewicht des konstitutionellen Spiels. Darüber hinaus ist jede beliebige diskriminierende Regel  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  genau dann ein Koordinationsgleichgewicht des konstitutionellen Spiels, wenn

$$W_i^B(\bar{r}_d) = W_i^B(\bar{r}_q) \quad \text{für alle } i.^{50}$$

<sup>50</sup> Auch im konstitutionellen Spiel unter Buchanans Schleier der Unsicherheit können (vollständig und unvollständig) gemischte Gleichgewichte des konstitutionellen Spiels existieren, die wiederum - mangels Plausibilität auf der konstitutionellen Ebene - vernachlässigt werden.



	B:			
	$r_{Bf} = (C_1, C_2)$	$r_{Bd1} = (D_1, C_2)$	$r_{Bd2} = (C_1, D_2)$	$r_{Bg} = (D_1, D_2)$
A:	$r_{Af} = (C_1, C_2)$	$W_A^B(\bar{r}_q) = (f_{A1}[2] + f_{A2}[2])/2,$ $W_B^B(\bar{r}_q) = (f_{B1}[2] + f_{B2}[2])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$
	$r_{Ad1} = (D_1, C_2)$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[1] + f_{A2}[1])/2,$ $W_B^B(\bar{r}_q) = (f_{B2}[1] + g_{B1}[1])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$
	$r_{Ad2} = (C_1, D_2)$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$
	$r_{Ag} = (D_1, D_2)$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$	$W_A^B(\bar{r}_q) = (g_{A1}[0] + g_{A2}[0])/2,$ $W_B^B(\bar{r}_q) = (g_{B1}[0] + g_{B2}[0])/2$

mit  $g_{ij}[1] > f_{ij}[2] > g_{ij}[0] > f_{ij}[1]$  für alle  $i = A, B$  und  $j = 1, 2$  (GD-Bedingung)

Abbildung 5.8: Konstitutionelles Spiel  $K^B$  unter Buchanans Schleier der Unsicherheit

Darüber hinaus dominiert auch im konstitutionellen Spiel unter dem Buchananschen Schleier der Unsicherheit die Strategie  $r_{if}$  für jede mögliche Auszahlungskonstellation schwach die Status-quo-Strategie  $r_{iq}$ , da  $W_i^B(r_{if}, r_{-i}) = W_i^B(r_{iq}, r_{-i})$  für alle  $(r_{if}, r_{-i}) \neq \bar{r}_f$  und  $W_i^B(r_{if}, r_{-if}) > W_i^B(r_{iq}, r_{-if})$  gilt. Auch hier kann unter bestimmten Bedingungen  $r_{if}$  schwach dominant über alle übrigen Strategien  $r_i \in M_i \setminus \{r_{if}, r_{iq}\}$  sein, so daß die faire Regel das Gleichgewicht in sukzessiver Dominanz ist.<sup>51</sup> Für den Fall, daß die Spieler identische Payoff-Relationen unter dem Schleier der Unsicherheit haben, läßt sich das konstitutionelle Spiel mit Hilfe des Kriteriums der Pareto-Dominanz lösen.

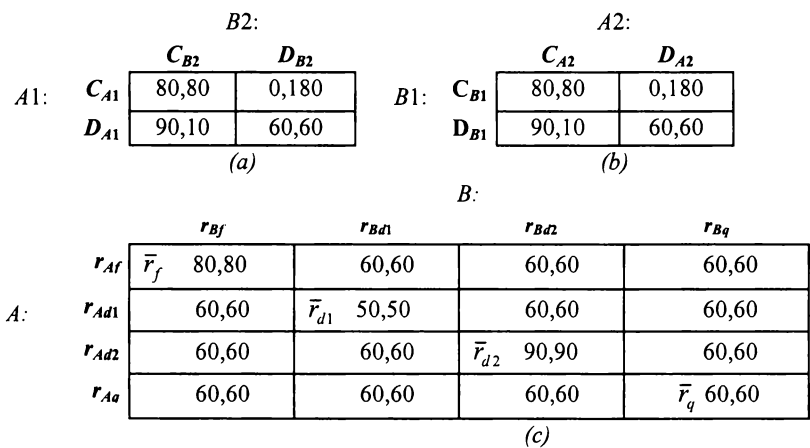


Abbildung 5.9: Einstimmige Wahl der diskriminierenden Regel  $\bar{r}_{d2}$  im konstitutionellen Spiel  $K_2^B$  unter Buchanans Schleier der Unsicherheit

In ihrem neuesten Buch behaupten Buchanan und Congleton:  
„... behind a veil of ignorance/uncertainty, persons could *never* agree to the establishment of political institutions that are predicted to discriminate explicitly in their operation.“<sup>52</sup>  
Diese Behauptung ist (logisch) falsch; im Kontext des vorliegenden Modells läuft sie auf die Allaussage hinaus, daß - bei Rationalität der Entscheidungsträger - alle einstimmigen Entscheidungen unter dem Schleier des Nichtwissens/der Unsicherheit die faire Regel zum Gegenstand haben werden. Statt dessen gilt aber:

<sup>51</sup> Vgl. dazu unten § 5 E. I.  
<sup>52</sup> Buchanan/Congleton (1998), S. 11 (Hervorhebung von mir; C. M.).

*Satz 5.3: Es gibt Payoff-Konstellationen, für die unter einem dicken Buchananschen Schleier der Unsicherheit eine einstimmige Wahl einer diskriminierenden Regel  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  das Ergebnis des konstitutionellen Spiels ist.*

Ein Gegenbeispiel zu Buchanans und Congletons Behauptung (und damit einen Beweis von Satz 5.3) gibt Abbildung 5.9 für ein konstitutionelles Zwei-Personen-Spiel  $K_2^B$ . Offensichtlich pareto dominiert dort - gegeben die Gefangenendilemma-Szenarien (a) und (b) - die diskriminierende Regel  $\bar{r}_{d2}$  jedes andere Spielergebnis, so daß bei Rationalität der Spieler diese Regel (und gerade nicht  $\bar{r}_f$ ) das Ergebnis der Verfassungswahl ist.

Um das Ergebnis von Satz 5.3 für die Regelwahl im Multi-Personen-Fall nachzuweisen, sei noch einmal ein  $n$ -Personen-Gefangenendilemma mit  $n > 2$  betrachtet, wobei aus Gründen der Einfachheit alle Spieler  $i$  im postkonstitutionellen Spiel die gleiche postkonstitutionelle GD-Auszahlungsfunktion (5.1) haben mögen. Alle Spieler  $i$  haben folglich auch die gleiche konstitutionelle Auszahlungsfunktion (5.6).

Zu Zwecken der graphischen Darstellung schreiben wir im folgenden alternativ zu (5.6) die für alle  $i$  identische konstitutionelle Präferenz in unmittelbarer Abhängigkeit von  $v[r]$ , also der Zahl der unter einer Strategienkombination  $r \in M^n$  des Gefangenendilemmas Kooperierenden:

$$(5.7) \quad \bar{W}^B[v(r)] = \begin{cases} g[v(r)] & \text{für } v(r) = 0 \\ \frac{v(r)}{n} f[v(r)] + \frac{n-v(r)}{n} g[v(r)] & \text{für } 0 < v(r) < n \\ f[v(r)] & \text{für } v(r) = n \end{cases}$$

Dabei gibt  $\frac{v(r)}{n}$  die Wahrscheinlichkeit an, mit der ein Spieler - gegeben eine bestimmte Strategienkombination  $r \in M^n$  des postkonstitutionellen Gefangenendilemmas - zu den Kooperierenden gehören wird;  $\frac{n-v(r)}{n}$  bezeichnet die entsprechende Gegenwahrscheinlichkeit, sich in der Gruppe der unter Regel  $r$  Defektierenden wiederzufinden.

Konkret betrachten wir ein symmetrisches Gefangenendilemma  $G_{1000}$  mit tausend Spielern. Jeder Spieler  $i$  habe die folgende Auszahlungsfunktion:

$$(5.8) \quad \pi_j(r) = \begin{cases} g[v(r)] = 5100 + 5v(r) & , \text{ falls } j \text{ defektiert} \\ f[v(r)] = 42 + 160v(r)^{0,5} & , \text{ falls } j \text{ kooperiert} \end{cases}$$

Die konstitutionelle Präferenz unter dem Schleier der Unsicherheit lautet folglich:

$$(5.9) \quad \bar{W}^B[v(r)] = \begin{cases} 5100 + 5v(r) & \text{für } v(r) = 0 \\ \frac{v(r)}{1000}(42 + 160v(r)^{0.5}) + \frac{1000 - v(r)}{1000}(5100 + 5v(r)) & \text{für } 0 < v(r) < n \\ 42 + 160v(r)^{0.5} & \text{für } v(r) = n \end{cases}$$

Die Spieler auf der konstitutionellen Ebene legen im Zuge ihrer Regelwahl die Anzahl  $v$  der Kooperierenden so fest, als ob sie den Erwartungswert (5.9) ihres Nutzens maximierten. Das Maximum von (5.9) liegt bei  $v(r) = 564$  Kooperierenden (siehe Abbildung 5.10); die Spieler erzielen hier einen konstitutionellen Payoff von  $\bar{W}^B[564] = 5620$ . Jede andere Regel des konstitutionellen Spiels ist weniger attraktiv als diese. Die Status-quo-Regel  $\bar{r}_q$  mit dem Payoff  $\bar{W}^B[0] = 5100$  und die faire Regel  $\bar{r}_f$  mit der Auszahlung  $\bar{W}^B[1000] = 5102$  sind Randminima von (5.9).

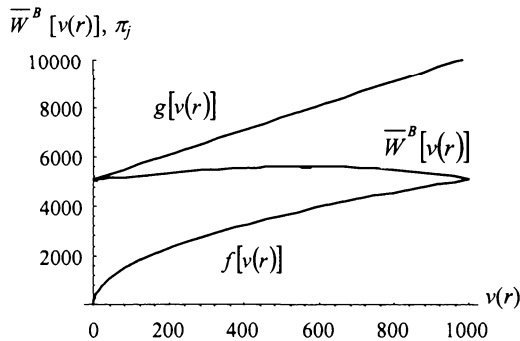


Abbildung 5.10: GD- und konstitutionelle Präferenzen, die die Wahl einer diskriminierenden Regel  $\bar{r}_d$  im konstitutionellen Spiel  $K_{1000}^B$  unter dem Schleier der Unsicherheit implizieren

Stehen nur drei Alternativen zur Abstimmung - die faire Regel  $\bar{r}_f$ , die Status-quo-Regel  $\bar{r}_q$  und eine diskriminierende Regel  $\bar{r}_d[v = 564]$ , die 564 genau bestimmte Positionsinhaber zur Kooperation, die übrigen 436 Positionsinhaber aber zur Defektion zwingt -, so werden sich alle Spieler auf der konstitutionellen Ebene für die diskriminierende Regel entscheiden; von den drei Wahlalternativen ist  $\bar{r}_d[v = 564]$  das einzige paretooptimale Koordinationsgleichgewicht des konstitutionellen Spiels. Offenbar sind im gewählten Beispiel unter dem Schleier der Unsicherheit die Free Rider-Positionen, die von der unfairen Regel

$\bar{r}_d [v = 564]$  einer großen Minderheit der Positionsinhaber eingeräumt werden, so attraktiv, daß alle Individuen für diese Regel stimmen, obwohl die Wahrscheinlichkeit, zu den Zwangskooperierenden zu gehören, deutlich größer ist als jene, sich unter den glücklichen Defektierenden wiederzufinden. Um einer Minderheit Vorteile zu verschaffen, macht Regel  $\bar{r}_d [v = 564]$  die Mehrheit der Gesellschaftsmitglieder zu deren Ausbeutungsobjekt. Trotzdem ist es auf der konstitutionellen Ebene für alle Spieler rational,  $\bar{r}_d [v = 564]$  zu wählen. Auch bei Buchanan ist also eine Entscheidung unter dem Schleier der Unkenntnis keine Garantie für die Fairneß der Entscheidungsergebnisse.

Anders als unter dem Rawlsschen Schleier des Nichtwissens, unter dem anahmegemäß alle ihr Verhalten in vollkommener Weise universalisieren, ist unter Buchanans Schleier des Nichtwissens nicht einmal garantiert, daß sich die Beteiligten überhaupt einigen. Ein solcher Konsens ist aber die notwendige Voraussetzung für die von Buchanan in einigen seiner Arbeiten geübte Verfahrensweise, den Kalkül eines einzigen Individuums unter dem Schleier der Unsicherheit als repräsentativ für alle anderen Beteiligten zu interpretieren und sich damit nur noch auf die Betrachtung eines einzigen, willkürlich ausgewählten Individuums zu konzentrieren.<sup>53</sup> Ein entsprechender Allsatz, daß alle Entscheidungen unter dem Schleier der Unsicherheit einstimmig fallen, ist aber widerlegbar, denn es gilt:

*Satz 5.4: Es gibt Payoff-Konstellationen, für die unter einem dicken Buchananschen Schleier der Unsicherheit eine einstimmige Wahl irgendeiner Regel  $\bar{r} \in \overline{M}$  verfehlt wird.*

Als Beweis für Satz 5.4 wird ein geeignetes Beispiel eines Zehn-Personen-Gefangenendilemmas ( $n=10$ ) gewählt. Der Einfachheit halber wollen wir unterstellen, daß es nur zwei verschiedene, nicht-leere Gruppen von Spielern  $a$  und  $b$  auf der konstitutionellen Ebene gibt. Alle Angehörigen von Gruppe  $a$  erwarten übereinstimmend ein postkonstitutionelles Gefangenendilemma mit der Auszahlungsfunktion (5.10), während die Mitglieder von Gruppe  $b$  ein Gefangenendilemma mit den in (5.11) gegebenen Payoffs erwarten:

$$(5.10) \quad \pi_a(r) = \begin{cases} g[v(r)] = 350 + 200v^{0,5} & , \text{ falls } j \text{ defektiert} \\ f[v(r)] = 260 + 3v^{1,5} & , \text{ falls } j \text{ kooperiert} \end{cases}$$

$$(5.11) \quad \pi_b(r) = \begin{cases} g[v(r)] = 350 + 100v & , \text{ falls } j \text{ defektiert} \\ f[v(r)] = 210 + 15v & , \text{ falls } j \text{ kooperiert} \end{cases}$$

<sup>53</sup> So in *Buchanan/Tullock* (1962) oder in *Brennan/Buchanan* (1988 [1980]).

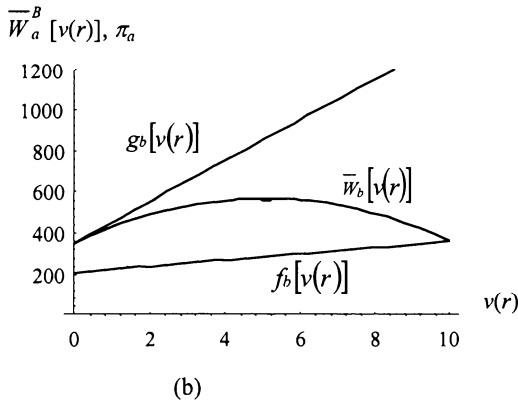
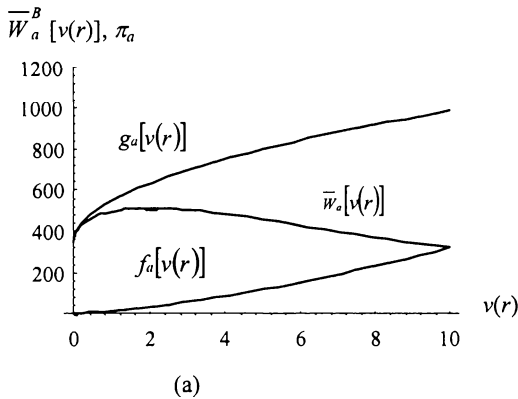


Abbildung 5.11: Gefangenendilemma- und konstitutionelle Präferenzen für Gruppen a und b, die Dissens über die Wahl irgendeiner Regel unter dem Schleier der Unsicherheit implizieren (Zahlenbeispiel für  $n = 10$ )

Die konstitutionellen Präferenzfunktionen für  $a$  und  $b$ , die zusammen mit den Auszahlungsfunktionen in Abbildung 5.11a und b graphisch dargestellt sind, lauten gemäß (5.7):

$$(5.12) \quad \overline{W}_a^B[v(r)] = \begin{cases} 350 + 200v^{0.5} & \text{für } v(r) = 0 \\ \frac{v(r)}{10}(260 + 3v^{1.5}) + \frac{10-v(r)}{10}(350 + 200v^{0.5}) & \text{für } 0 < v(r) < n \\ 260 + 3v^{1.5} & \text{für } v(r) = n \end{cases}$$

$$(5.13) \quad \overline{W}_b^B[v(r)] = \begin{cases} 350 + 100v & \text{für } v(r) = 0 \\ \frac{v(r)}{10}(210 + 15v) + \frac{10-v(r)}{10}(350 + 100v) & \text{für } 0 < v(r) < n \\ 210 + 15v & \text{für } v(r) = n \end{cases}$$

Wie man sieht, haben die Mitglieder von Gruppe  $a$  gemäß (5.12) unter dem Schleier der Unsicherheit eine konstitutionelle Präferenz für eine diskriminierende Regel – nennen wir sie:  $\bar{r}_d[v = 3]$  –, die insgesamt drei Positionsinhaber zur Kooperation und sieben andere Gesellschaftsmitglieder zur Defektion zwingt; jedes Mitglied aus  $a$  erhält dann den insgesamt höchsten Nutzenerwartungswert,  $\bar{W}_a^B[3] = 570$ . Die Individuen in Gruppe  $b$  präferieren demgegenüber gemäß (5.13) auf der konstitutionellen Ebene eine diskriminierende Regel  $\bar{r}_d[v = 5]$ , unter der je fünf Positionsinhaber kooperieren bzw. defektieren; jedes Mitglied von  $b$  erhält hier den höchsten Payoff von  $\bar{W}_b^B[5] = 568$ .

		Mitglied aus Gruppe b:			
		$r_{bf}$	$r_{bd}[v = 3]$	$r_{bd}[v = 5]$	$r_{bq}$
Mitglied aus Gruppe a:	$r_{af}$	355, 360	350, 350	350, 350	350, 350
	$r_{ad}[v = 3]$	350, 350	570, 532	350, 350	350, 350
	$r_{ad}[v = 5]$	350, 350	350, 350	545, 568	350, 350
	$r_{aa}$	350, 350	350, 350	350, 350	350, 350

Abbildung 5.12: Nichteinigung zwischen Mitgliedern von Gruppe a und Gruppe b im konstitutionellen Spiel unter dem Schleier der Unsicherheit (Zahlenbeispiel für  $n = 10$ )

Angenommen, es stehen nur vier Regeln zur Auswahl: die faire Regel  $\bar{r}_f$ , die Status-quo-Regel  $\bar{r}_q$  sowie je eine konkrete diskriminierende Regel  $\bar{r}_d[v = 3]$ , nach der drei bestimmte Positionsinhaber kooperieren müssen, und  $\bar{r}_d[v = 5]$ , nach der fünf Positionsinhaber zur Kooperation gezwungen werden. Dann läßt sich das konstitutionelle Spiel unter dem Schleier der Unsicherheit als Zwei-Personen-Spiel zwischen je einem repräsentativen Individuum aus Gruppe  $a$  und einem aus Gruppe  $b$  darstellen (vgl. Abbildung 5.12). Eliminiert man für beide Gruppen  $a$  und  $b$  die Status-quo-Strategie  $r_{iq}$ , die jeweils von den drei anderen Strategien schwach dominiert wird, so verbleiben drei echte Koordinationsgleichgewichte. Wie in einem „Kampf der Geschlechter“-Spiel präferieren beide Gruppen  $a$  und  $b$  zwar irgendeine Einigung gegenüber jedem Nichteinigungspunkt; auch präferieren beide Gruppen eine Einigung auf eine diskriminierende Regel gegenüber der Wahl der fairen Regel. Jede der beiden Gruppen favorisiert jedoch eine andere diskriminierende Regel, so daß nach dem Pareto-Dominanz-Kriterium nicht entschieden werden kann, auf welche Regel sich die beiden Teilpopulationen einigen werden.

Unter einem Rawlsschen Schleier des Nichtwissens, der zu einer vollkommenen Universalisierung führt, ist eine solche Nichteinigung hingegen ausgeschlossen. Da nämlich bei Rawls beide Spieler die postkonstitutionellen Gesell-



schaftspositionen nicht - wie bei Buchanan - mit ihrer eigenen Nutzenfunktion bewerten, sondern mit den jeweiligen Grundgüterindizes des Positionsinhabers, kann sich hieraus kein Dissens ergeben.<sup>54</sup>

Buchanans Position wurde bislang als Behauptung interpretiert, daß es unter dem Schleier der Unsicherheit *immer* zu einer fairen Einigung kommen werde. Gleichwohl ist auch hier das Werk Buchanans widersprüchlich. Trotz des obigen Zitats Buchanans und Congletons, nach welchem die Existenz eines Schleiers der Unkenntnis *niemals* unfaire Entscheidungen ermögliche, bieten andere Stellen seiner Schriften auch Anhaltspunkte dafür, daß die Behauptung, unsichere Erwartungen über die eigene gesellschaftliche Position implizierten faire und einstimmige Entscheidungen, nicht unter allen Umständen Geltung beansprucht. In bezug auf die unter dem Unsicherheitsschleier induzierte Fairneß (nach Maßgabe ihrer „Allgemeinheitsprinzip“ genannten Universalisierungs-norm) findet sich bei Buchanan und Congleton etwa auch die folgende Passage:

„The veil of ignorance and/or uncertainty is a construction that is helpful in some ultimate evaluation of alternative rules. The construction, as such, however, does not readily allow an objectively identifiable classification of political actions that may or may not meet the abstracted generality norm. Almost any observed political action, no matter how discriminatory in effect, may be rhetorically defended on some veil-of-ignorance argument.“<sup>55</sup>

Passagen in einem Aufsatz von Buchanan und Faith deuten überdies darauf hin, daß auch die Einstimmigkeit keine zwingende Implikation unter einem Schleier der Unsicherheit sein soll.<sup>56</sup>

Es besteht daher die Möglichkeit, daß Buchanans Behauptung, nach welcher Entscheidungen unter dem Schleier der Unsicherheit einstimmig und fair seien, nicht den Status eines Allsatzes, sondern den eines Existenzsatzes haben soll der folgenden Form:

*Satz 5.5: Es gibt Payoff-Konstellationen, für die unter einem dicken Buchananschen Schleier der Unsicherheit die Wahl der fairen Regel  $\bar{r}_f$  das Ergebnis des konstitutionellen Spiels ist.*

<sup>54</sup> Es besteht allerdings auch hier die Möglichkeit eines konstitutionellen Koordinationsproblems; vgl. unten § 5 E. III.

<sup>55</sup> Buchanan/Congleton (1998), S. 7.

<sup>56</sup> „In order to demonstrate that the introduction of subjective elements in the individual calculus [gemeint ist die Annahme einer lediglich partiellen Universalisierung unter dem Schleier der Unsicherheit; C. M.] elevates the contractual process to center stage, it is necessary to show that differences in predictions can lead to differing arrays of institutional alternatives. Without some potential for disagreement on the alternatives from which a collective or group selection is to be made, the contractual process loses its relevance.“ Buchanan/Faith (1980), S. 25.

Um die Wahrheit von Existenzsatz 5.5 zu beweisen, betrachten wir das Beispiel eines konstitutionellen Zwei-Personen-Spiels unter Buchanans Schleier der Unsicherheit,  $K_2^B$ , in Abbildung 5.13. Im konstitutionellen Spiel (c), das aus den beiden Gefangenendilemma-Szenarien (a und b) hergeleitet ist, paretodominiert die faire Regel  $\bar{r}_f$  jeden anderen Einigungs- und Nichteinigungspunkt.

			A2:				A1:
		$C_{B2}$	$D_{B2}$			$C_{A2}$	$D_{A2}$
A1:	$C_{A1}$	60, 50	0, 70	B1:	$C_{A1}$	50, 40	10, 80
	$D_{A1}$	70, 0	10, 10		$D_{B1}$	60, -10	30, 10
(a)				(b)			
B:							
		$r_{Bf}$	$r_{Bd1}$	$r_{Bd2}$	$r_{Bq}$		
A:	$r_{Af}$	$\bar{r}_f$ 50, 50	10, 20	10, 20	10, 20		
	$r_{Ad1}$	10, 20	$\bar{r}_{d1}$ 30, 30	10, 20	10, 20		
	$r_{Ad2}$	10, 20	10, 20	$\bar{r}_{d2}$ 40, 40	10, 20		
	$r_{Aq}$	10, 20	10, 20	10, 20	$\bar{r}_q$ 10, 20		
(c)							

Abbildung 5.13: Einstimmige Wahl der fairen Regel  $\bar{r}_f$  im konstitutionellen Spiel  $K_2^B$  unter Buchanans Schleier der Unsicherheit

Als ein weiteres Zahlenbeispiel für die Pareto-Dominanz der fairen Regel  $\bar{r}_f$  diene ein konstitutionelles Spiel auf der Basis eines Gefangenendilemmas mit 200 Spielern, die alle über die gleiche Auszahlungsfunktion (5.14) verfügen:

$$(5.14) \quad \pi_j = \begin{cases} g[v(r)] = 22500 + 40v & , \text{ falls } j \text{ defektiert} \\ f[v(r)] = 8v^{1,5} & , \text{ falls } j \text{ kooperiert} \end{cases}$$

Unter dem Schleier der Unsicherheit ergibt sich also für jeden Spieler  $i$  gemäß (5.7) folgender Erwartungswert des Nutzens:

$$(5.15) \quad \bar{W}^B[v(r)] = \begin{cases} 22500 + 40v(r) & \text{für } v(r) = 0 \\ \frac{v(r)}{200}(8v^{1,5}) + \frac{200-v(r)}{2000}(22500 + 40v(r)) & \text{für } 0 < v(r) < n \\ 8v^{1,5} & \text{für } v(r) = n \end{cases}$$

Die Funktion (5.15)  $\bar{W}^B[v(r)]$  verläuft über den gesamten Bereich von  $0 \leq v \leq 200$  u-förmig und weist ein Minimum bei  $v = 111$  mit einem konstitutionellen Payoff in Höhe von  $\bar{W}^B[111] = 17181$  auf (vgl. Abbildung 5.14). Es gibt

also keine lokalen Maxima außer am Rand, nämlich bei der Status-quo-Regel  $r_q$  mit  $\bar{W}^B[0] = 22500$ , was zugleich der konstitutionelle Payoff für alle Nichteinigungspunkte ist, und bei der fairen Regel  $r_f$  mit  $\bar{W}^B[200] = 22627$ . Da der konstitutionelle Payoff der fairen Regel für alle Spieler höher ist als bei jedem anderen Einigungs- und Nichteinigungspunkt, ist die faire Regel  $\bar{r}_f$  das einzige pareto dominante Gleichgewicht des konstitutionellen Spiels  $K_{200}^B$ .

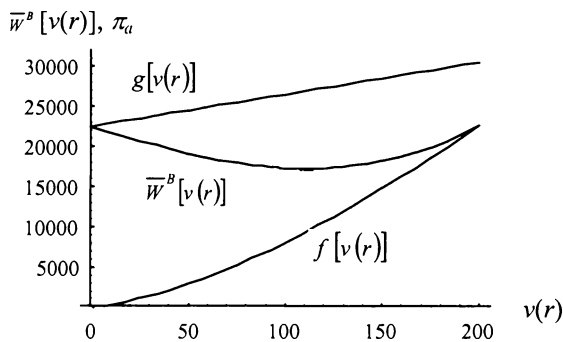


Abbildung 5.14: GD- und konstitutionelle Präferenzen, die die Wahl der fairen Regel  $\bar{r}_f$  im konstitutionellen Spiel  $K_{200}^B$  unter dem Schleier der Unsicherheit implizieren

Die Existenz des Schleiers der Unsicherheit kehrt im Beispiel die Präferenzordnung aller Spieler  $i$  also um. Im Falle *sicherer* Erwartungen schätzt  $i$  nach Maßgabe seiner üblichen Gefangenendilemma-Präferenzen jene Situation, in der er selbst defektiert, die anderen  $v(r) = n-1 = 199$  Individuen aber kooperieren, höher als die Kooperation aller  $n$  Individuen, da gilt:

$$g[199] = 30460 > f[200] = 22627 \quad \Rightarrow \quad \pi(r_d[v = 199]) > \pi(r_f).$$

Unter dem Schleier der Unsicherheit hat jeder Spieler  $i$  demgegenüber im Beispiel eine höhere Präferenz für die faire Regel  $\bar{r}_f$ , unter der alle  $n$  Spieler kooperieren, als für jede andere Strategienkombination des konstitutionellen Spiels. So gilt etwa im Vergleich zu einer Regel  $\bar{r}_d[v = 199]$ , unter der nur  $i$  (oder irgendein anderer Spieler) bei Kooperation aller übrigen  $n-1$  Spieler defektiert:

$$\bar{W}^B[200] = 22627 > \bar{W}^B[199] = 22498 \quad \Rightarrow \quad W^B(\bar{r}_f) > W^B(\bar{r}_d[v = 199])$$

Eine faire Einigung ist also auch unter Buchanans Schleier der Unsicherheit möglich. Die Bedingungen hierfür untersucht der nächste Abschnitt.

### E. Bedingungen einer fairen Einigung

In den vorangegangenen Abschnitten wurde gezeigt, daß der Schleier der Unkenntnis - gleichgültig, ob in der Variante von Rawls oder von Buchanan - eine Einigung der konstitutionellen Entscheidungsträger auf die faire Regel  $\bar{r}_f$  nicht garantieren kann. Gleichwohl, so konnte festgestellt werden, besteht in beiden Fällen die *Möglichkeit* einer fairen Einigung. Im folgenden sollen die Bedingungen eines fairen Vertragsschlusses unter dem Schleier der Unkenntnis genauer analysiert werden. In den Abschnitten I. bis III. werden, in Form von alternativen Fairneßbedingungen, logisch notwendige Erweiterungen der Rawlsschen bzw. Buchananschen Annahmen (und damit eine Einschränkung der Allgemeinheit ihrer Theorien) eingeführt, welche - angesichts eines postkonstitutionellen Gefangenendilemmas - die Behauptung einer fairen Einigung unter dem Schleier der Unkenntnis immer (logisch) wahr machen; die Existenzsätze 5.1, 5.3 und 5.4 sind bei Erfüllung dieser Bedingungen logisch falsch.

#### I. Fairneß und Dominanz

Unter welchen Bedingungen einigen sich die Spieler in einem konstitutionellen Spiel auf die faire Regel  $\bar{r}_f$ ? Satz 5.6 zeigt die Bedingung dafür, daß die faire Regel  $\bar{r}_f$  die Gleichgewichtslösung in sukzessiver Dominanz des konstitutionellen Spiels ist:<sup>57</sup>

*Satz 5.6: Die faire Regel  $\bar{r}_f \in \bar{M}^n$  ist genau dann die Gleichgewichtslösung in sukzessiver (schwacher) Dominanz des konstitutionellen Spiels, wenn gilt:*

$$\forall \bar{r}_d \in \bar{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}: \exists i: [W_i(\bar{r}_q) \geq W_i(\bar{r}_d)].$$

*Beweis:* (a)  $\Leftarrow$ : Es wurde bereits gezeigt, daß Strategie  $r_{if}$  die Status-quo-Strategie  $r_{iq}$  schwach dominiert. Ich zeige: Auch jede beliebige andere Strategie  $r_{id} \in M_i$  (außer  $r_{if}$ ) wird schwach dominiert.

<sup>57</sup> Auf eine Indizierung der konstitutionellen Auszahlung mit „R“ oder „B“ wird verzichtet.

Sei  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  eine beliebige diskriminierende Regel des postkonstitutionellen Gefangenendilemmas.

Nach Voraussetzung:  $\exists i: [W_i(\bar{r}_q) \geq W_i(\bar{r}_d)]$ . Da  $r_{id}$  höchstens gegen  $r_{-id}$  eine andere Auszahlung erzielt als  $r_{iq}$ , gilt:  $W_i(r_{iq}, r_{-i}) \geq W_i(r_{id}, r_{-i})$  für alle  $r_{-i}$ . Da  $r_{iq}$  schwach dominiert wird von  $r_{if}$ , wird folglich auch  $r_{id}$  schwach dominiert. Wenn  $r_{iq}$  gestrichen wird, gilt für das verbleibende Spiel für alle anderen Spieler  $i$ :  $W_i(r_{id}, r_{-i}) = W_i(r_{iq}, r_{-i})$  für alle  $r_{-i}$ . Also wird nun auch für alle anderen Spieler  $i$   $r_{id}$  von  $r_{if}$  schwach dominiert. Weil diese Überlegung für jede beliebige Regel  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  gilt, verbleibt  $r_{if}$  für jeden Spieler  $i$  als einzige nicht (sukzessiv) schwach dominierte Strategie.

(b)  $\Rightarrow$ : Angenommen,  $\bar{r}_f$  ist Lösung in sukzessiver schwacher Dominanz des konstitutionellen Spiels. Zu zeigen ist:  $\exists i: [W_i(\bar{r}_q) \geq W_i(\bar{r}_d)]$ .

Da  $\bar{r}_f$  Lösung in sukzessiver schwacher Dominanz ist, gibt es ein Individuum  $i$ , so daß die Strategie  $r_{id}$  nach Streichung einer endlichen Anzahl anderer Strategien  $r_i \neq r_q$  schwach dominiert wird. Sei  $r_i^*$  also eine Strategie von  $i$ , die  $r_{id}$  schwach dominiert. Dann gilt für alle  $r_{-i}$  auf jeden Fall:  $W_i(r_i^*, r_{-i}) \geq W_i(r_{id}, r_{-i})$  für alle  $r_{-i}$ . Insbesondere für  $r_{-i} = r_{-id}$  gilt daher:  $W_i(r_i^*, r_{-id}) \geq W_i(r_{id}, r_{-id}) \Leftrightarrow W_i(\bar{r}_q) \geq W_i(\bar{r}_d)$ .  $\square$

Nach Satz 5.6 ist die faire Regel  $\bar{r}_f$  genau dann das Gleichgewicht in sukzessiver schwacher Dominanz des konstitutionellen Spiels, wenn  $\bar{r}_f$  von *allen* Spielern als einzige Regel strikt gegenüber  $\bar{r}_q$  präferiert wird. Im Urzustand unter Rawls' Bedingungen ist dies immer erfüllt, wenn - wie in den obigen Beispielen zu Satz 5.2 - alle Spieler im postkonstitutionellen Gefangenendilemma die gleiche Auszahlungs- (Grundgüter-)Funktion haben. Immer dann also, wenn das zugrunde liegende postkonstitutionelle Gefangenendilemma-Spiel *symmetrisch* ist, ist bei Rawls die faire Regel  $\bar{r}_f$  das Gleichgewicht des konstitutionellen Spiels in sukzessiver schwacher Dominanz. Symmetrie schafft Fairneß unter dem Schleier des Nichtwissens.

Unter Buchanans Prämissen kommt es nicht auf die Symmetrie der Auszahlungen im postkonstitutionellen Spiel an. Satz 5.6 impliziert, daß die faire Regel  $\bar{r}_f$  im konstitutionellen Spiel unter dem Schleier der Unsicherheit genau dann das Gleichgewicht in sukzessiver schwacher Dominanz ist, wenn es für alle diskriminierenden Regeln  $r_d \in M^n \setminus \{r_f, r_q\}$  mindestens einen Spieler  $i$  gibt, für den gilt:

$$(5.16) \quad \sum_{j \in I_C(r_d)} (g_{ij}[0] - f_{ij}[v(r_d)]) \geq \sum_{j \in I_D(r_d)} (g_{ij}[v(r_d)] - g_{ij}[0]).$$

Dabei bezeichnet  $I_C(r_d)$  die Menge aller Positionsinhaber  $j$ , die im postkonstitutionellen Spiel in einer Strategienkombination  $r_d$  des Gefangenendilemmas kooperieren, und  $I_D(r_d)$  ist die komplementäre Menge aller unter Regel  $r_d$  defektierenden Positionsinhaber, mit  $I_C(r_d) \cup I_D(r_d) = \{1, 2, \dots, n\}$ .

Immer dann also, wenn es unter jeder diskriminierenden Regel  $r_d$  mindestens einen Spieler gibt, nach dessen GD-Payoffs die Summe seiner Verluste gegenüber dem Nash-Gleichgewicht des Gefangenendilemmas in den Positionen der Kooperierenden (linke Seite der Ungleichung) mindestens so hoch ist wie die Summe seiner Gewinne gegenüber der Situation allgemeiner Defektion in den Positionen der Defektierenden (rechte Seite der Ungleichung), so ist für *alle* Spieler  $r_{ij}$  die einzige nicht-dominierte Strategie des konstitutionellen Spiels.

## II. Fairneß und Pareto-Dominanz

Die von Satz 5.6 formulierte Bedingung dafür, daß die faire Regel  $\bar{r}_f$  die Lösung des konstitutionellen Spiels in sukzessiver schwacher Dominanz ist, ist unter vielen möglichen Payoff-Konstellationen im postkonstitutionellen Gefangenendilemma nicht erfüllt. Gleichwohl kann die faire Regel  $\bar{r}_f$  auch in diesen Fällen das Ergebnis des konstitutionellen Spiels sein, wenn man - wie in den meisten obigen Beispielen - das schwächere, im vertragstheoretischen Kontext jedoch plausible Lösungskonzept der Pareto- (oder Auszahlungs-)Dominanz zugrunde legt:

*Satz 5.7: Die faire Regel  $\bar{r}_f \in \overline{M}^n$  paretodominiert jeden Einigungs- oder Nichteinigungspunkt  $\rho \in M^n$  des konstitutionellen Spiels genau dann, wenn gilt:*

$$\forall \bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}: [\forall i: (W_i(\bar{r}_f) \geq W_i(\bar{r}_d)) \wedge \exists i: (W_i(\bar{r}_f) > W_i(\bar{r}_d))]$$

Ob die faire Regel  $\bar{r}_f$  das einzige paretodominante Gleichgewicht des konstitutionellen Spiels ist, hängt nach Satz 5.7 von den Präferenzrelationen im zugrunde liegenden Gefangenendilemma ab. Im konstitutionellen Spiel unter Rawls' Schleier des Nichtwissens ist die Bedingung von Satz 5.7 dann erfüllt,

wenn für alle diskriminierenden Regeln  $r_d \in M^n \setminus \{r_f, r_q\}$  des postkonstitutionellen Gefangenendilemmas gilt:

$$(5.17) \quad \min \{f_i[n], \dots, f_n[n]\} \geq \min \{\pi_i[r_d], \dots, \pi_n[r_d]\} \quad \text{für alle } i \text{ und} \\ \min \{f_i[n], \dots, f_n[n]\} > \min \{\pi_i[r_d], \dots, \pi_n[r_d]\} \quad \text{für mindestens ein } i.$$

Der minimale Payoff unter der fairen Regel  $r_f$  des Gefangenendilemmas muß also für alle Spieler  $i$  mindestens so hoch sein wie der minimale Grundgüterindex unter allen diskriminierenden Regeln  $r_d$ , wobei jeweils für einen Spieler  $i$  die strikte Ungleichung gelten muß.

Dies läßt sich anschaulich für  $n = 2$  demonstrieren. Nach (5.16) muß für ein Zwei-Personen-Gefangenendilemma im Sinne von Abbildung 5.3 gelten:

$$f_1[2] > f_2[1] \text{ und } f_2[2] > f_1[1].$$

In einem konstitutionellen Zwei-Personen-Spiel unter Rawls' Schleier des Nichtwissens pareto dominiert  $r_f$  also dann jeden anderen Einigungs- und Nicht-einigungspunkt, wenn der Payoff eines einseitig kooperierenden „Suckers“ in der einen Rolle geringer ist als der Payoff des defektierenden Spielers in der anderen Position bei beidseitiger Kooperation. In dem Gefangenendilemma, das dem konstitutionellen Spiel der Abbildung 5.6 zugrunde liegt, ist diese Bedingung erfüllt.

Satz 5.6 impliziert für ein konstitutionelles Spiel unter Buchanans Schleier der Unsicherheit, daß die faire Regel  $\bar{r}_f$  genau dann jede andere Strategienkombination des konstitutionellen Spiels pareto dominiert, wenn für alle Strategienkombinationen  $r_d \in M^n \setminus \{r_f, r_q\}$  des postkonstitutionellen Gefangenendilemmas gilt:

$$(5.18) \quad \sum_{j=1}^n f_{ij}[n] \geq \sum_{j=1}^n \pi_{ij}[r_d] \quad \text{für alle } i \text{ und} \\ \sum_{j=1}^n f_{ij}[n] > \sum_{j=1}^n \pi_{ij}[r_d] \quad \text{für mindestens ein } i.$$

Die Summe der Nutzenwerte eines jeden Spielers  $i$  für alle Positionen  $j$  unter der fairen Regel  $r_f$  muß nach (5.18) also die Summe von  $i$ 's Nutzenwerten unter jeder diskriminierenden Regel  $r_d \in M^n \setminus \{r_f, r_q\}$  übersteigen. Da nach der Definition des Gefangenendilemmas stets  $W_i(\bar{r}_f) > W_i(\bar{r}_q)$ , pareto dominiert  $\bar{r}_f$  auch die Status-quo-Regel  $\bar{r}_q$  und alle Nichteinigungspunkte  $\rho \in M^n \setminus \bar{M}^n$  des konstitutionellen Spiels.



Die in (5.18) formulierte Pareto-Dominanz-Bedingung ist ein Analogon zur Bedingung von Rapoport und Chammah, wonach in einem Zwei-Personen-Gefangenendilemma die Nutzensumme beider Spieler im kooperativen Pareto-Optimum größer sein muß als in den beiden nicht-kooperativen Pareto-Optima des Gefangenendilemmas.<sup>58</sup> Führt nämlich die einstimmige Wahl der kooperativen Strategie im GD nicht zur höchsten Nutzensumme, so bestünde - nach Rapoport und Chammah - die eigentlich kooperative Lösung des Gefangenendilemmas darin, daß sich beide Spieler auf eine der beiden Trittbrettfahrer-Positionen einigen und den dort erzielten Überschuß untereinander aufteilen, etwa indem sie im Falle der Iteration des Spiels sich wechselseitig ausbeuten.

Eine solche konsensuale Lösung der Verteilung der Free-Rider-Gewinne unter den Spielern steht im konstitutionellen Spiel jedoch nicht zur Verfügung, weil die Festlegung einer Verteilungsregel unter den Beteiligten einen Kommunikations- und Verhandlungsprozeß auf der konstitutionellen Ebene voraussetzt. Die konstitutionelle Entscheidungsfindung ist hier jedoch - in Übereinstimmung mit den üblichen Prämissen der Theorien des hypothetischen Gesellschaftsvertrags à la Rawls und Buchanan - als ein Prozeß der *monologischen* Entscheidungsfindung modelliert, da aufgrund eines „Problems der großen Zahl“ in großen, modernen Gesellschaften Verhandlungen von Angesicht zu Angesicht nicht möglich sind. Sollen daher konstitutionelle Modelle Lösungen für heutige Massengesellschaften aufzeigen, so müssen sie darlegen, daß sich die Beteiligten *stillschweigend* - jeder für sich allein - auf die gleiche Lösung „einigen“.<sup>59</sup>

Die Fairneßbedingung gemäß (5.18) bringt die Spieler unter dem dicken Schleier der Unsicherheit dazu, die externen Kosten, welche die Defekteure den kooperierenden Kollektivmitgliedern durch ihr Free Riding im postkonstitutionellen Spiel aufzwingen, in ihren Kalkül mit einzubeziehen. Nach Umstellen von (5.18) erhält man für alle  $r_d \in M^n \setminus \{r_f, r_q\}$ :

$$(5.19) \quad \sum_{j \in I_C(r_d)} (f_{ij}[n] - f_{ij}[v(r_d)]) \geq \sum_{j \in I_D(r_d)} (g_{ij}[v(r_d)] - f_{ij}[n]) \quad \text{für alle } i \text{ und}$$

$$\sum_{j \in I_C(r_d)} (f_{ij}[n] - f_{ij}[v(r_d)]) > \sum_{j \in I_D(r_d)} (g_{ij}[v(r_d)] - f_{ij}[n]) \quad \text{für mindestens ein } i.$$

Nach (5.19) wählt jeder Spieler  $i$  die faire Regel  $\bar{r}_f$ , wenn für alle  $i$  - gemessen in  $i$ 's eigenen Nutzeneinheiten - unter allen diskriminierenden Regeln  $\bar{r}_d \in \bar{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  die Summe der Free-Rider-Gewinne aller Defekteure im

<sup>58</sup> Vgl. Rapoport/Chammah (1965), S. 34 f. Vgl. zu dieser Bedingung auch Kuhn/Moresi (1995); Kuhn (1996); Frohlich/Oppenheimer (1996b).

<sup>59</sup> Vgl. ähnlich Binmore (1994), S. 79.

Vergleich zur fairen Regel (rechte Seite der Ungleichung) nicht höher ist als die externen Kosten in Form von Verlusten, welche die Trittbrettfahrer unter Regel  $\bar{r}_d$  den Kooperierenden im Vergleich zur Situation allgemeiner Kooperation aufbürden (linke Seite der Ungleichung); für mindestens ein  $i$  müssen diese Defektionsgewinne niedriger sein als die Kooperationskosten. Am Fall eines konstitutionellen Spiels auf der Basis eines Gefangenendilemmas mit  $n = 2$  läßt sich dies besonders anschaulich zeigen:

$$(f_{i2}[2] - f_{i2}[1] > g_{i1}[1] - f_{i1}[2]) \text{ und } (f_{i1}[2] - f_{i1}[1] > g_{i2}[1] - f_{i2}[2]).$$

Da Spieler  $i$  als potentielles Ausbeutungsoffer in der einen Position erwartet, mehr zu verlieren (nämlich  $f_{i1}[2] - f_{i1}[1]$  bzw.  $f_{i2}[2] - f_{i2}[0]$ ), als er als potentieller Free Rider in der anderen Position zu gewinnen erwarten kann (nämlich  $g_{i1}[1] - f_{i1}[2]$  bzw.  $g_{i2}[1] - f_{i2}[2]$ ), wird er, um sich selbst zu schützen, für die faire Regel  $\bar{r}_f$  stimmen. Setzt man die Nutzeneinheiten aus Abbildung 5.13 in (5.19) ein, so ergibt sich beispielsweise für Spieler  $A$ :

$$(40 - (-10) > 70 - 60) \quad \text{und} \quad (60 - 0 > 80 - 40).$$

Erfüllen die konstitutionellen Präferenzen also die Fairneßbedingung (5.19), so internalisiert der Schleier der Unsicherheit vollständig die Interessen der „Opfer“ im Kalkül des potentiellen Trittbrettfahrers.

### III. Fairneß und Koordination

Selbst dann, wenn die Fairneßbedingungen aus Satz 5.6 und 5.7 verletzt sind, muß eine Einigung auf die faire Regel  $\bar{r}_f$  nicht völlig ausgeschlossen sein, wenn man die vertragstheoretischen Annahmen um die Möglichkeit von *Fokuspunkt-Lösungen* im Sinne von Schelling erweitert.<sup>60</sup> Ein fokales Gleichgewicht ist ein Koordinationsgleichgewicht, das sich - aufgrund von Traditionen, Kommunikation vor dem Spiel oder anderen Eigenschaften - von jedem anderen Gleichgewicht unterscheidet, ohne daß diese Besonderheiten in die formale Repräsentation des Spiels integriert sind.

*Satz 5.8: Die faire Regel  $\bar{r}_f \in \overline{M}^n$  kann Fokuspunkt-Lösung des konstitutionellen Spiels sein.*

Dieser Fall ist besonders in solchen konstitutionellen Spielen plausibel, die man als „Indifferenzspiele“ bezeichnen könnte. Darunter verstehe ich alle sol-

<sup>60</sup> Vgl. Schelling (1960), Kapitel 3 und 4. Zur spieltheoretischen Modellierung von Fokuspunkt-Gleichgewichten vgl. auch Myerson (1991), S. 108-114.

chen konstitutionellen Spiele, in denen neben der fairen Lösung  $\bar{r}_f$  mindestens ein weiteres Gleichgewicht in reinen Strategien existiert, das alle anderen Gleichgewichte des konstitutionellen Spiels pareto dominiert. Die konstitutionellen Präferenzen können in diesen Fällen - im konstitutionellen Zwei-Personen-Spiel  $K_2^B$  unter Buchanans Schleier der Unsicherheit - für jeden Spieler  $i$  beispielsweise die folgende Form haben:

$$W_i^B(\bar{r}_f) = W_i^B(\bar{r}_{d1}) > W_i^B(\bar{r}_{d2}) \geq W_i^B(\bar{r}_q) \quad (\text{Fall 1}) \text{ oder}$$

$$W_i^B(\bar{r}_f) = W_i^B(\bar{r}_{d2}) > W_i^B(\bar{r}_{d1}) \geq W_i^B(\bar{r}_q) \quad (\text{Fall 2}) \text{ oder}$$

$$W_i^B(\bar{r}_f) = W_i^B(\bar{r}_{d1}) = W_i^B(\bar{r}_{d2}) > W_i^B(\bar{r}_q) \quad (\text{Fall 3})$$

oder irgendeine Kombination der Fälle 1 oder 2 mit Fall 3. Nach Eliminierung der schwach dominierten Strategien  $r_{iq}$  entspricht Fall 3, der auch in Abbildung 5.15 wiedergegeben wird, den wohlbekannten *reinen Koordinationsspielen*.<sup>61</sup> Ähnliche Überlegungen gelten den beiden anderen Fällen von Spielen mit nur zwei auszahlungsdominanten Koordinationsgleichgewichten.

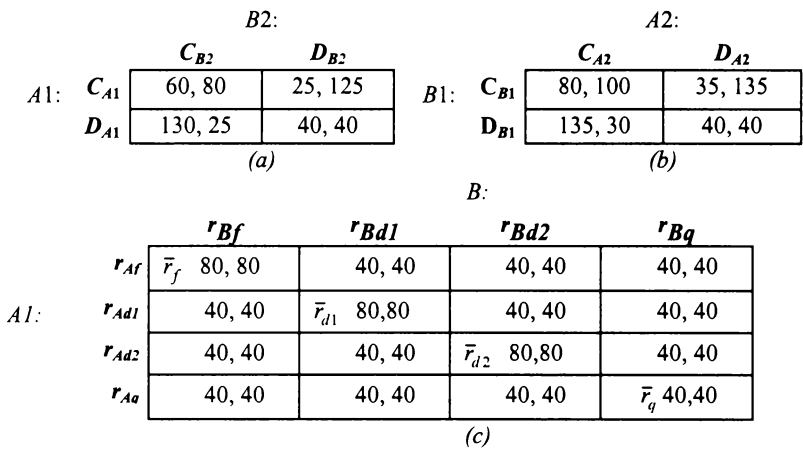


Abbildung 5.15: GD-Szenarien und konstitutionelles Spiel unter Buchanans Schleier der Unsicherheit als reines Koordinationsspiel (Fall 3) für  $n = 2$

In allen diesen Spielen sind die Spieler indifferent zwischen mindestens zwei auszahlungsdominanten Gleichgewichten. Im Fall 3 beispielsweise (Abbildung

<sup>61</sup> Vgl. grundlegend Schelling (1960); Lewis (1975 [1969]); neuere Abhandlungen hierzu finden sich etwa bei Tietzel (1990); Cooper/DeJong/Forsythe/Ross (1990); Metha/Starmier/Sugden (1994a), dies. (1994b); Sugden (1995); Cooper (1999).

5.15) sind beide Spieler indifferent zwischen der fairen Regel  $\bar{r}_f$ , der diskriminierenden Regel  $\bar{r}_{d1}$  und der unfairen Regel  $\bar{r}_{d2}$ . Das konstitutionelle Spiel ist in diesem Fall nicht dominanzlösbar; auch kann nach dem Kriterium der Pareto-Dominanz nicht vorhergesagt werden, ob die beiden Spieler überhaupt eine Einigung treffen können und, wenn ja, welche. Eine monologische Einigung auf die faire Regel  $\bar{r}_f$  braucht aber auch dann nicht ausgeschlossen zu sein.

Während nämlich in Abbildung 5.15 eine diskriminierende Regel wie  $\bar{r}_{d1}$  oder  $\bar{r}_{d2}$  kaum die Eigenschaften eines Fokuspunkt-Gleichgewichts aufweisen dürfte, kann dies für die faire Regel  $\bar{r}_f$  durchaus der Fall sein.

- Auf der einen Seite werden nämlich nur faire Regeln in der Lage sein, die Stabilität eines konstitutionellen Arrangements im Zeitablauf zu gewährleisten. Daß die Entscheidungsträger auf der konstitutionellen Ebene selbst danach streben könnten, die Stabilität ihres Gesellschaftsvertrages zu erhöhen, berücksichtigen auch Vanberg und Buchanan.<sup>62</sup> Unter der Stabilität konstitutioneller Arrangements verstehen sie deren Beständigkeit im Zeitablauf. Nur stabile Verfassungen ermöglichten es den Individuen, die Wohlfahrtssteigerungen, die vom Abschluß eines Gesellschaftsvertrags ausgehen, dauerhaft zu erzielen. Die Entscheidungsträger achteten daher von vornherein darauf, daß die konstitutionellen Verabredungen fair seien. Denn Gesellschaftsverträge, welche die Partikularinteressen bestimmter Individuen fördern, seien zwar grundsätzlich denkbare Ergebnisse des Verhandlungsprozesses; sie würden sich auf Dauer jedoch als „weniger robust“ erweisen als faire Verträge, da von den benachteiligten Kollektivmitgliedern ein dauerhafter Neuverhandlungsdruck zu erwarten sei. Vanberg und Buchanan schließen daraus, daß die konstitutionellen Entscheidungsträger diese Stabilitätsgefahr antizipieren und von vornherein tendenziell faire Verfassungsverträge vereinbaren werden:

„Rational actors can be expected to take considerations of stability into account when engaging in constitutional choice. And to the extent that *fairness* and *stability* of constitutional arrangements are interrelated, the concern for stability will induce a concern for fairness. It will do so ... in addition to and independently of the veil of uncertainty.“<sup>63</sup>

Im vorliegenden Modell läßt sich die Idee, daß neben dem Schleier der Unsicherheit auch eine „Stabilitätsbesorgnis“ der Entscheidungsträger auf der konstitutionellen Ebene konsensfördernd wirken könne, dahingehend interpretieren, daß in den Augen der Beteiligten die faire Regel  $\bar{r}_f$  die Eigenschaften eines Fokuspunkts hat. Denn unter den beiden diskriminierenden

<sup>62</sup> Vgl. Vanberg/Buchanan (1989), S. 54-56.

<sup>63</sup> Vanberg/Buchanan (1989), S. 54 (Hervorhebungen im Original).

Regeln  $\bar{r}_{d1}$  und  $\bar{r}_{d2}$  gibt es stets einen Gewinner und einen Verlierer, der sogar gegenüber dem Status quo schlechter gestellt wird. Von daher ist zu erwarten, daß der Verlierer in der postkonstitutionellen Gesellschaft auf die Neuverhandlung eines Vertrags dringen wird. Er wird in der postkonstitutionellen Gesellschaft das monetäre Äquivalent von bis zu  $(f_i[2] - f_i[1])$ -Payoff-Einheiten für Rent-Seeking-Aktivitäten ausgeben, um von der Geltung einer ihn diskriminierenden Regel zur fairen Regel zu gelangen. Wenn die Entscheidungsträger, wie von Vanberg und Buchanan angenommen, solche Stabilitätsaspekte mit in ihr Kalkül ziehen, werden sie sich im Falle ihrer konstitutionellen Indifferenz zwischen der fairen und mindestens einer diskriminierenden Regel übereinstimmend für das Koordinationsgleichgewicht  $\bar{r}_f$  statt für die diskriminierenden Institutionen  $\bar{r}_{d1}$  und  $\bar{r}_{d2}$  entscheiden.

- Auf der anderen Seite kann die faire Regel  $\bar{r}_f$  - unabhängig von der Stabilitätsbesorgnis der konstitutionellen Entscheidungsträger - eine Fokuspunkt-Eigenschaft auch dadurch haben, daß die Individuen im Rahmen eines vorherrschenden repressiven Meinungsklimas bestimmte Reputationsnutzen dadurch erzielen können, daß sie nach außen hin diejenige Auffassung vertreten, die von ihm „erwartet“ wird, und zwar selbst dann, wenn hierdurch ihre wahren, „privaten“ Präferenzen „falsifiziert“ werden.<sup>64</sup> Hat die öffentliche Meinung eine starke Präferenz für „fares“ Verhalten, so können sich beide Spieler veranlaßt sehen, die faire Regel  $\bar{r}_f$ , nicht aber eine diskriminierende Regel zu wählen. Dies ist um so wahrscheinlicher, als sie für diese ostentative Befolgung der gesellschaftlich akzeptierten Fairneßnorm keine persönlichen Nachteile in Kauf nehmen müssen; denn in einer durch ein Indifferenzspiel beschriebenen „Niedrigkostensituation“ ist „gemeinnütziges“ Verhalten kostenlos.<sup>65</sup>

Im Ergebnis der Überlegungen in § 5 läßt sich festhalten, daß - gegeben ein postkonstitutionelles Gefangenendilemma - auch die Existenz eines Rawlschen oder Buchananschen Schleiers der Unkenntnis keineswegs Anlaß sein kann für den zwangsläufigen Schluß, daß eine Einigung auf die faire Regel  $\bar{r}_f$  zustande kommt. Vielmehr hängt die Richtigkeit dieser Annahme von den *relativen Präferenzintensitäten*, insbesondere von der Erfülltheit der angegebenen Fairneßbedingungen, ab. In Fällen, in denen der Schleier der Unkenntnis ein konstitutionelles Indifferenzspiel induziert, kann es überdies erforderlich sein, daß erst die Existenz (nicht strategisch beeinflussbarer, „gegebener“) zusätzlicher Bedin-

<sup>64</sup> Vgl. Kuran (1995).

<sup>65</sup> Daß es Individuen manchmal wichtiger ist, fair zu erscheinen, als tatsächlich fair zu sein, zeigen auch experimentelle Untersuchungen zum Ultimatum-Spiel; vgl. Kagel/Kim/Moser (1995).

gungen wie mögliche Fokuspunkt-Eigenschaften von Gleichgewicht zu Wahl einer allgemein akzeptablen Institution führen.

Analog zu Rapoport und Chammahs Bedingung könnte man versucht sein, die Fairneß und Einstimmigkeit im konstitutionellen Spiel unter dem Schleier der Unkenntnis dadurch zu retten, Personen mit „unpassenden“ - d.h. die Fairneß-Bedingung verletzenden - Präferenzen einfach nicht zur Entscheidungsfindung zuzulassen. Die Individuen im Urzustand wären dann - unabhängig von den tatsächlich von den Individuen verfolgten Interessen - nur noch im Rahmen vordefinierter, die Fairneßbedingungen erfüllender Präferenzen souverän. Dies wäre jedoch ein eklatanter Verstoß gegen das höchstrangige vertragstheoretische Werturteil des normativen Individualismus, nach dem, in einer Formulierung Brennans und Buchanans, „alle Personen als moralisch gleichwertig angesehen werden, als Subjekte, die alle gleichermaßen in der Lage sind, die jeweiligen Optionen bewerten zu können.“<sup>66</sup>

---

<sup>66</sup> *Brennan/Buchanan* (1993 [1985]), S. 28.

## **§ 6 Fairneß und Einstimmigkeit unter sicheren Erwartungen**

In § 5 wurde gezeigt, daß ein Rawlsscher Schleier des Nichtwissens und ein dicker Buchananischer Schleier der Unsicherheit zwar zu einer einstimmigen Einigung auf die Durchsetzung der kooperativen GD-Lösung führen können, dies aber keineswegs zwangsläufig der Fall sein muß. Denn hierbei sind sogar diskriminierende Regeln, welche die ganze Last der Regelbefolgung nur auf ein einziges Individuum abladen, mögliche Ergebnisse des konstitutionellen Spiels. Unter einem Buchananischen Schleier der Unsicherheit gibt es darüber hinaus nicht einmal eine Garantie, daß überhaupt irgendeine Einigung erzielt wird. Buchanans Methode, das Verhalten eines Entscheidungskollektivs dadurch zu analysieren, daß er - bei Annahme eines Schleiers der Unsicherheit - ein einziges, angeblich repräsentatives Individuum herausgreift, wird angesichts dieses Ergebnisses fragwürdig. Im vorliegenden § 6 wird darüber hinaus gezeigt, daß die Unsicherheit der Entscheidungsträger über ihre eigene postkonstitutionelle Position nicht einmal in allen Fällen faire und einstimmige Entscheidungen auch nur *wahrscheinlicher* macht.

Einer weit verbreiteten Ansicht zufolge, steigt die Wahrscheinlichkeit konsensualer, nicht-diskriminierender Entscheidungen mit der „Dicke“ des Schleiers der Unkenntnis. Für die Abstimmungssituation in Rawls' Urzustand, für die aus normativen Gründen immer die größtmögliche „Dicke“ des Schleiers des Nichtwissens unterstellt wird, ist dies eine Selbstverständlichkeit. In der „Theorie der Gerechtigkeit“ sind stets zur zwei Zustände zu vergleichen: die Situation mit und die ohne konstitutionelle Unkenntnis. Im ersten Fall ist eine faire Einigung mit Sicherheit, im zweiten nur im Ausnahmefall zu erwarten.

Auch bei Buchanan findet sich die Behauptung, faire und einstimmige Kollektiventscheidungsergebnisse im Urzustand seien um so wahrscheinlicher, je „dicker“ der Schleier der Unsicherheit ist. Vor dem Hintergrund von Buchanans normativer Ausgangsposition legen Vanberg und Buchanan den Schwerpunkt vor allem auf den Einigungs- statt auf den Fairneßaspekt:

„The veil of uncertainty can be more or less transparent, or, in other terms, its ‘thickness’ may vary, dependent on certain characteristics of the actual choice situation. As the veil’s ‘thickness’ increases so will the prospect of achieving agreement.“<sup>1</sup>

---

<sup>1</sup> Vanberg/Buchanan, 1989, S. 54 (Hervorhebung von mir; C. M.).



In diesem Sinne behaupten auch Brennan und Buchanan:

„Die Unsicherheit, die mit *jeder* Entscheidung über Regeln und Institutionen einhergeht, läßt die Wahrscheinlichkeit potentieller Übereinkünfte zunehmen. Konfrontiert mit einer echten Unsicherheit darüber, wie eine bestimmte Regel sich auf die eigene Position auswirkt, wird das Individuum durch seine eigeninteressierten Überlegungen dazu veranlaßt, sich auf die Entscheidungsoptionen zu konzentrieren, die die Wahrscheinlichkeit potentiell schädlicher Ergebnisse eliminieren oder minimieren.“<sup>2</sup>

Im folgenden werde ich jedoch darlegen, daß diese Behauptung - zumindest in ihrer Allgemeinheit - falsch ist. Ich werde argumentieren, daß sich die in § 5 aufgezeigten Probleme, unter konstitutioneller Unsicherheit über die eigene gesellschaftliche Position einstimmig eine faire Regel für das Gefangenendilemma zu wählen, unter bestimmten Umständen dadurch lösen lassen, daß man die Prämisse eines Schleiers der Unkenntnis einfach aufgibt.

### A. Das konstitutionelle Zwei-Personen-Spiel unter perfekter Information

Liegt dem konstitutionellen Spiel der Standardfall eines einfachen Zwei-Personen-Gefangenendilemmas zugrunde, so gilt:

*Satz 6.1: Rationale Spieler im Zwei-Personen-Spiel mit GD-Präferenzen einigen sich auf der konstitutionellen Ebene in voller Kenntnis ihrer Partialinteressen immer auf die faire Regel  $\bar{r}_f$ . Das Gleichgewicht  $\bar{r}_f$  ist bei sicheren Erwartungen die sukzessive Dominanzlösung von  $K_2$ .*

Kennt jeder Spieler im Zeitpunkt  $t_c$  mit Sicherheit seine postkonstitutionelle Rolle, dann ergibt sich das in Abbildung 6.1 angegebene konstitutionelle Spiel.

Bei sicheren Erwartungen kalkulieren beide Spieler übereinstimmend nur noch ein einziges Gefangenendilemma-Szenario (Abbildung 6.1a). Offensichtlich dominiert im entsprechenden konstitutionellen Spiel (Abbildung 6.1b) Strategie  $r_{Af}$  von Spieler  $A$  nicht nur dessen Status-quo-Strategie  $r_{Aq}$  schwach, sondern auch seine Strategie  $r_{Ad2}$ ; Analoges gilt für die Strategie  $r_{Bf}$  von Spieler  $B$ , welche die Strategien  $r_{Bq}$  und  $r_{Bd1}$  schwach dominiert. Nach Elimination der inferioren Strategien verbleibt die 2x2-Matrix (Abbildung 6.1c) mit dem starken Gleichgewicht  $\bar{r}_f = (r_{Af}, r_{Bf})$  und dem schwachen Gleichgewicht  $(r_{Ad1}, r_{Bd2})$ . In

<sup>2</sup> Brennan/Buchanan (1993 [1985]), S. 39 (Hervorhebung von mir; C. M.).

diesem Restspiel auszahldominiert die faire Regel  $r_f$  nicht nur jeden anderen Nichteinigungspunkt; die Strategie  $r_{if}$  dominiert überdies schwach die Strategie  $r_{Ad1}$  bzw.  $r_{Bd2}$ .

Spieler B:

	$C_B$	$D_B$
Spieler A:	$C_A$ $r_f$ $f_A[2], f_B[2]$	$r_{d2}$ $f_A[1], g_B[1]$
	$D_A$ $r_{d1}$ $g_A[1], f_B[1]$	$r_q$ $g_A[0], g_B[0]$

mit  $g_i[1] > f_i[2] > g_i[0] > f_i[1]$  und  $i = A, B$  (GD-Bedingung)  
(a)

Spieler B:

	$r_{Bf}$	$r_{Bf1}$	$r_{Bf2}$	$r_q$
Spieler A:	$r_{Af}$ $\bar{r}_f$ $f_A[2], f_B[2]$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$
$r_{Ad1}$	$g_A[0], g_B[0]$	$\bar{r}_{d1}$ $g_A[1], f_B[1]$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$
$r_{Ad2}$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$	$\bar{r}_{d2}$ $f_A[1], g_B[1]$	$g_A[0], g_B[0]$
$r_{Aq}$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$	$\bar{r}_q$ $g_A[0], g_B[0]$

mit  $g_i[1] > f_i[2] > g_i[0] > f_i[1]$  und  $i = A, B$  (GD-Bedingung)  
(b)

Spieler B:

	$r_{Bf}$	$r_{Bd2}$
Spieler A:	$r_{Af}$ $\bar{r}_f$ $f_A[1], f_B[1]$	$g_A[0], g_B[0]$
$r_{Ad1}$	$g_A[0], g_B[0]$	$g_A[0], g_B[0]$

mit  $g_i[1] > f_i[2] > g_i[0] > f_i[1]$  und  $i = A, B$  (GD-Bedingung)  
(c)

Abbildung 6.1: Konstitutionelles Zwei-Personen-Spiel  $K_2$  unter perfekter Information

Das Ergebnis ist also, daß *ohne* einen Schleier der Unkenntnis eine einstimmige und faire Entscheidung immer getroffen wird. Die Wahl der zwangsweisen Durchsetzung der kooperativen GD-Lösung ist das einzige Gleichgewicht in sukzessiver Dominanz des konstitutionellen Spiels. Gegeben die Rationalität aller Spieler, ist die volle Kenntnis der postkonstitutionellen Position eine *Garantie gegen Disharmonie und Diskriminierung*. Im Unterschied dazu ist das konstitutionelle Spiel selbst im Falle vollkommener Unkenntnis der postkonstitutionellen Partialinteressen nicht immer dominanzlösbar. Auch auszahldominiert die faire Lösung  $\bar{r}_f$  nicht in allen Fällen jedes andere Gleichgewicht. Erst der Schleier schafft damit - im Falle eines postkonstitutionellen Zwei-Personen-Gefangenendilemmas - das Problem, das zu lösen er eingesetzt wurde.

Dieses Ergebnis ist nicht wirklich eine Überraschung. Denn die Payoff-Relationen im Zwei-Personen-Gefangenendilemma erfüllen stets die Anforderungen der Dominanzbedingung aus Satz 5.6, nach der für alle diskriminierenden Regeln  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  für jeweils mindestens ein Individuum  $i$  gelten muß:  $w_i(\bar{r}_q) \geq w_i(\bar{r}_d)$ . In  $G_2$  gilt nach der PD-Bedingung sogar für beide Individuen und beide diskriminierenden Regeln  $\bar{r}_{d1}$  und  $\bar{r}_{d2}$  stets die strikte Ungleichung:

$$(6.1) \quad g_i[0] > f_i[1] \text{ für alle } i$$

$$\Rightarrow \quad w_i(\bar{r}_q) > w_i(\bar{r}_{d1}) \quad \text{und} \quad w_i(\bar{r}_q) > w_i(\bar{r}_{d2}) \text{ für alle } i.$$

Bei völlig sicheren Erwartungen wird daher kein Spieler in einen Vertrag einwilligen, der ihn - in der postkonstitutionellen Phase - schlechter stellen wird als ohne diesen Vertrag. „Politik“ bei Einstimmigkeit - also die Annahme einer Regel - ist damit ohne Schleier stets eine Pareto-Verbesserung. Bei Unkenntnis der eigenen postkonstitutionellen Position ist dies hingegen nicht mehr gewährleistet. Vom konstitutionellen Standpunkt können daher auch die diskriminierenden Regeln  $\bar{r}_{d1}$  und  $\bar{r}_{d2}$  als Pareto-Verbesserungen erscheinen, die einen der beiden Spieler in der postkonstitutionellen Situation mit Sicherheit gegenüber dem Status quo schlechter stellen. Die konstitutionelle Unsicherheit *verringert* also die Wahrscheinlichkeit, daß die postkonstitutionelle Situation im Vergleich zum Status quo paretosuperior ist.

Abbildung 6.2 verdeutlicht diesen Zusammenhang graphisch. Dabei wird aus Vereinfachungsgründen davon ausgegangen, daß auch im Rawls-Fall kardinale Nutzenfunktionen statt Grundgüterindizes verwendet werden. Für den Buchananschen Schleier der Unsicherheit wird von einer nicht nur partiellen, sondern vollkommenen Universalisierung ausgegangen.<sup>3</sup> Dazu wird unterstellt, daß beide Spieler  $A$  und  $B$  identische Nutzenfunktionen haben. Somit unterscheiden sich die beiden erwarteten GD-Szenarien (wie bei Rawls) nur durch die Bezeichnung der Positionen.

Auf den Koordinaten sind die Auszahlungen  $\pi_1$  und  $\pi_2$  der Spieler in den postkonstitutionellen Positionen 1 und 2 angegeben. Es wird angenommen, daß aufgrund eines Schleiers des Nichtwissens bzw. der Unsicherheit nicht bekannt ist, in welcher der beiden Positionen 1 und 2 des postkonstitutionellen Spiels sich  $A$  und  $B$  wiederfinden werden. Die gestrichelte Linie beschreibt die konve-

<sup>3</sup> Die *Buchanan*-Indifferenzkurven fallen daher mit jenen *Harsanyis* zusammen.



$r_{d1}$  jedoch auf einer höheren Rawls-Indifferenzkurve ( $R^*R^*$ ) liegt als  $r_f$  ( $R^{**}R^{**}$ ), ist nach dem Maximin-Kriterium die unfaire Regel  $r_{d1}$  vorzuziehen.

Unter dem Buchananschen Schleier der Unsicherheit erscheinen alle Regeln als wohlfahrtsmehrend, die nordöstlich von Indifferenzkurve  $BB$  liegen. Dies gilt nicht nur für die faire Regel  $r_f = (C_1, C_2)$ , sondern auch für die diskriminierenden Regeln  $r_{d1} = (D_1, C_2)$  und  $r_{d2} = (C_1, D_2)$ , obwohl diese außerhalb der Paretogrenzen  $PP$  liegen. Da die Indifferenzkurve  $B^{**}B^{**}$  ein höheres Nutzenniveau repräsentiert als  $B^*B^*$ , wird - wie im Rawls-Fall - statt der fairen die diskriminierende Regel  $r_{d1} = (D_1, C_2)$  gewählt.

## B. Das konstitutionelle $n$ -Personen-Spiel unter perfekter Information

Auch im Falle eines konstitutionellen  $n$ -Personen-Spiels gibt es Konstellationen, in denen ein Schleier der Unkenntnis zu Diskriminierung und Dissens auf der konstitutionellen Ebene führt, während unter sicheren Erwartungen faire und einstimmige Entscheidungsergebnisse garantiert sind.

Unter sicheren Erwartungen hat jeder Spieler  $i$  im konstitutionellen  $n$ -Personen-Spiel (5.2) die folgende Auszahlungsfunktion:

$$(6.2) \quad W_i(\rho) = \begin{cases} \pi_i(r), & \text{falls } \rho = \bar{r} \in \overline{M}^n \\ g_i[0] & \text{sonst} \end{cases}$$

Immer dann, wenn alle Spieler  $i$  die gleiche konstitutionelle Strategie wählen, erhält  $i$  die Auszahlung der jeweiligen Regel; im Falle der Nichteinigung, der das Spielen des postkonstitutionellen  $n$ -Personen-Gefangenendilemmas zur Folge hat, erhält  $i$  hingegen die Auszahlung  $g_i[0]$  des jeweiligen Nash-Gleichgewichts.

Satz 5.6 impliziert, daß die faire Regel  $\bar{r}_f$  genau dann das Gleichgewicht des konstitutionellen Spiels unter sicheren Erwartungen ist, wenn für jede diskriminierende Regel  $\bar{r}_d \in \overline{M}^n \setminus \{\bar{r}_f, \bar{r}_q\}$  gilt:

$$(6.3) \quad g_i[0] \geq f_i[n-1] \quad \text{für mindestens ein } i.$$

Im Gegensatz zum Zwei-Personen-Fall haben nicht alle  $n$ -Personen-Gefangenendilemmata diese Eigenschaft. So ist es durchaus denkbar, daß Individuen, die unter einer diskriminierenden Regel zur Kooperation gezwungen werden, im Vergleich zum Nash-Gleichgewicht von der Existenz dieser Regel

trotzdem profitieren. In solchen Fällen gibt es kritische Werte des Anteils von Defektierenden im Gefangenendilemma, bei deren Überschreiten ein kooperierendes Individuum gegenüber der Situation allseitiger Defektion schlechter gestellt wird.<sup>5</sup> Auf der konstitutionellen Ebene der Regelwahl wären entsprechend auch unter sicheren Erwartungen nicht nur die faire Regel  $\bar{r}_f$ , sondern auch diskriminierende Regeln  $\bar{r}_d$  paretosuperior im Vergleich zur Status-quo-Regel  $\bar{r}_q$ . Unter dieser Bedingung könnte sich dann auch im konstitutionellen Spiel ein Gleichgewichtsauswahlproblem stellen, das mit dem Kriterium der Pareto-Dominanz allein nicht zu lösen ist. Um zu einer Entscheidung zu gelangen, welche Regel aus konstitutioneller Sicht vorzuziehen ist, bedürften die Rationalentscheider im Urzustand daher sehr wohl einer zusätzlichen, die Einigung befördernden Bedingung wie eines Schleiers der Unkenntnis. Auch die unsicheren Erwartungen können aber, wie in § 5 gezeigt wurde, nicht garantieren, daß sich die Individuen auf die faire Regel  $\bar{r}_f$  statt auf eine diskriminierende Regel einigen werden.

Es ist plausibel anzunehmen, daß die Erfülltheit von Bedingung (6.3) wesentlich von der Größe eines Kollektivs abhängig ist. Wird etwa in einer gegebenen Gesellschaft ein Kollektivgut bereitgestellt und mit einer proportionalen Einkommensteuer finanziert, so befinden sich die Individuen insoweit in einem  $n$ -Personen-Gefangenendilemma: Die Steuerpflichtigen haben einen Anreiz, der eigenen Zahlung durch Steuerausweichung zu begegnen. Wird das Kollektivgut vollständig bereitgestellt, so externalisieren sie durch ihre Defektion den auf sie entfallenden Finanzierungsanteil auf ihre kooperierenden Mitbürger; wird das Gut indes nur in einer geringeren Menge bereitgestellt, so tragen die Kooperierenden aufgrund der Steuerhinterziehung der anderen Opportunitätskosten in Form der entgangenen Staatsleistungen.<sup>6</sup>

Entzieht sich in einem kleinen Kollektiv mit nur einem Dutzend Spielern nur ein einziger seiner Beitragspflicht, so kann diese Defektion einen oder sogar alle kooperierenden Mitspieler schlechter stellen als im Nash-Gleichgewicht des Gefangenendilemmas, in dem keiner einen Beitrag leistet; Bedingung (6.3) ist dann erfüllt. Eine diskriminierende Regel  $\bar{r}_d$ , welche die Kooperation nicht aller Individuen erzwingt, kann daher nicht Gleichgewicht in (schwach) undominierten Strategien sein. Jeder ausgebeutete Spieler wird gegen  $\bar{r}_d$  stimmen, da er unter dieser Regel noch schlechter stünde als in der Situation ohne jede Regel;  $\bar{r}_d$  liegt für ihn im konstitutionellen Spiel unter sicheren Erwartungen auf einer dominierten Strategie. Im konstitutionellen Spiel mit kleinem  $n$  ist

<sup>5</sup> Vgl. z. B. Schelling (1973), S. 387 ff.

<sup>6</sup> Vgl. ähnlich Buchanan (1966); Gaertner (1988), S. 119 f.

daher eine Einigung auf die faire Regel  $\bar{r}_f$  unter allen Umständen garantiert. Ein Schleier der Unkenntnis ist hier bestenfalls unschädlich; nützen kann er nicht.

Anders mag dies mit wachsendem  $n$  aussehen. Je mehr Spieler ein Gefangenendilemma spielen, desto weniger dürfte *ceteris paribus* die Gruppe der Kooperierenden dadurch geschädigt werden, daß ein einziger defektiert. Im Extremfall eines sehr großen Kollektivs ergibt sich die völlige Insignifikanz aller Spieler des  $n$ -Personen-Gefangenendilemmas; die Steuerausweichung eines einzigen Kollektivmitglieds hat dann keinen unmittelbaren Einfluß mehr auf die Auszahlungen der kooperierenden Mitbürger. Aus der konstitutionellen Perspektive können darum selbst unter sicheren Erwartungen diskriminierende Regeln Pareto-Verbesserungen gegenüber der Status-quo-Regel  $\bar{r}_d$  darstellen; auch ohne Schleier der Unkenntnis besteht hier die Möglichkeit von Diskriminierung und Dissens im konstitutionellen Spiel.

Das bedeutet jedoch nicht, daß das Vorliegen von Bedingung (6.3) in einem Gefangenendilemma für sehr großes  $n$  völlig unplausibel wäre. Im Naturzustand von Hobbes etwa, der für die heutige Gefangenendilemma-Modellierung der Vertragstheorie Pate gestanden hat, ist diese Bedingung für alle Auszahlungsfunktionen der Spieler erfüllt. So ist im „Leviathan“ stets klar, daß kein Individuum im Naturzustand einen Anreiz hat, *einseitig* zu kooperieren - in Hobbescher Diktion: sein anarchisches „Recht auf alles“ aufzugeben - statt zu defektieren, wenn die anderen Spieler ihrerseits defektieren:

„Verzichten ... andere nicht ebenso wie er auf ihr Recht, so besteht für niemanden Grund, sich seines Rechts zu begeben, denn dies hieße eher, sich selbst als Beute darbieten ...“<sup>7</sup>

Übersetzt in die Sprache des hier entwickelten Modells bedeutet dies: Kooperieren nicht alle Beteiligten unter einer konstitutionell für das Gefangenendilemma zu wählenden Regel, so zahlt es sich für jeden einzelnen aus, gegen diese Regel zu stimmen. Das wird er aber nur dann tun, wenn der Payoff im Nash-Gleichgewicht  $r_q$  des  $n$ -Personen-Gefangenendilemmas höher ist als im Falle jeder einseitigen Kooperation:

$$g_i[0] > f_i[n-1] \quad \text{für alle Spieler } i = A, B, \dots, n.$$

Dieser „Hobbesianischen“ Anforderung an die Payoffs im  $n$ -Personen-Gefangenendilemma, die sogar schärfer ist als Bedingung (6.3), genügen alle in § 5 diskutierten Beispiele für das  $n$ -Personen-Gefangenendilemma. Ein Schleier

---

<sup>7</sup> Hobbes (1976 [1651]), S. 100. Vgl. zur hier vertretenen Hobbes-Interpretation auch Taylor (1987), S. 133.



der Unkenntnis hat hier daher keine positive Funktion: für eine faire Einigung wird er nicht gebraucht; er kann aber, wie wir gesehen haben, das Erreichen von Fairneß und Einstimmigkeit auf der konstitutionellen Ebene verhindern.

Vor diesem Hintergrund verwundert es auch nicht länger, daß sich im „Leviathan“ kein Schleier der Unkenntnis oder irgendeine andere, die Fairneß und Einstimmigkeit der Individuen garantierende Prämisse findet. Die Einigung auf die faire Regel  $\bar{r}_f$  ist bei Hobbes das einzige Gleichgewicht des konstitutionellen  $n$ -Personen-Spiels in sukzessiver Dominanz.<sup>8</sup>

## C. Diskussion

Wann immer Gesellschaftsverträge durch ein Gefangenendilemma bedroht werden, in dem die Spieler in der Lage sind, sich gegenseitig signifikant zu schädigen, ist ein Schleier der Unkenntnis, sei es in der Rawlschen oder der

---

<sup>8</sup> Dieses Ergebnis gilt allerdings nicht, wenn man, wie manche Autoren (vgl. *Taylor* 1976, 1987; *McLean* 1981, 1985), den *Hobbesschen* Naturzustand als ein (unendlich iteriertes) Gefangenendilemma-Superspiel interpretiert. Zum einen ist bei üblichen Rationalitätsannahmen die Gestalt des konstitutionellen Spiels in diesem Fall nicht immer eindeutig, da es nach dem Folk-Theorem im GD-Superspiel nicht, wie im einmal gespielten Gefangenendilemma, nur ein einziges, sondern für genügend hohe Diskontfaktoren unendlich viele teilspielperfekte Nash-Gleichgewichte gibt (vgl. *Fudenberg/Maskin* (1986); *Fudenberg/Tirole* (1991), S. 150-165; *van Damme* (1991), Kap. 8.); es fehlt damit im konstitutionellen Spiel eine eindeutige Status-quo-Regel  $r_q$ . Zwar ist dieses mehrdeutige Ergebnis im Kontext der *Hobbesschen* Theorie unplausibel, derzufolge im Urzustand Defektion unter allen Umständen rational ist; das Konzept der Teilspielskonsistenz (vgl. *Güth/Leininger/Stephan* 1991), das Defektion in allen (isomorphen) Teilspielen des GD-Superspiels verlangt, scheint daher der *Hobbesschen* Rationalitätsdefinition näher zu kommen als das Konzept der Teilspielperfekteit. Selbst dann aber, wenn bei Zugrundelegung der Teilspielskonsistenz-Bedingung das iterierte Gefangenendilemma mit  $(ALL\_D, \dots, ALL\_D)$  ein einziges Gleichgewicht plausibel erscheint und die Status-quo-Regel  $r_q$  mithin eindeutig bestimmt ist, verbleibt unter *sicheren* Erwartungen im konstitutionellen Spiel stets ein Gleichgewichtsauswahlproblem. Ist daher das postkonstitutionelle Spiel ein iteriertes (Zwei- oder  $n$ -Personen-)Gefangenendilemma, so ist die faire Regel  $\bar{r}_f = (r_{Af}, r_{Bf}, \dots, r_{nf})$  mit  $r_{if} = (ALL\_C, \dots, ALL\_C)$  unter keinen Umständen die Gleichgewichtslösung in dominanten Strategien bzw. die Payoff-Dominanz-Lösung des konstitutionellen Spiels. Ein Schleier der Unkenntnis kann in diesem Fall deshalb auch im *Hobbesschen* Naturzustand durchaus eine positive Funktion haben. Gleichwohl kann auch diese Annahme, wie in den in § 5 diskutierten Fällen, Fairneß und (der *Buchanansche* Schleier der Unsicherheit) Einstimmigkeit auf der konstitutionellen Ebene nicht garantieren. Der Fall eines postkonstitutionellen Gefangenendilemma-Superspiels wird im Rahmen dieser Arbeit nicht weiter betrachtet.

Buchananschen Variante, nicht nur überflüssig, sondern die konstitutionelle Unwissenheit der Entscheidungsträger über ihre eigenen postkonstitutionellen Positionen kann sogar einer fairen (und bei Buchanan auch einer einstimmigen) Entscheidungsfindung auf der konstitutionellen Ebene im Wege stehen. Statt dessen, so habe ich in § 6 argumentiert, werden die konstitutionellen Entscheidungsträger in voller Kenntnis ihrer Partialinteressen keine Schwierigkeiten haben, sich fair und einstimmig zu entscheiden.

Die hier erzielten Ergebnisse werfen jedoch zwei Fragen auf: *Erstens* stellt sich die Frage, warum Individuen in der Realität, wo die Entscheidungsfindung über Regeln typischerweise nicht unter einem Schleier der Unkenntnis stattfindet, oft so große Probleme haben, sich in Situationen zu einigen, die üblicherweise als Gefangenendilemma beschrieben werden. Warum etwa fällt es Politikern so schwer, ordnungspolitische Maßnahmen einzuführen, obwohl der ordnungspolitische Regelrahmen, wie häufig angenommen, von einem Gefangenendilemma bedroht wird? Auch wird oft dem Umweltproblem die Anreizstruktur eines Gefangenendilemmas zugeschrieben. Warum aber ist es dann nicht längst gelöst? Zumindest dann, wenn die GD-Auszahlungen Bedingung (6.3) erfüllen, dürften die Individuen bzw. deren politische Sachwalter keine Probleme haben, sich zu einigen, wenn sie nur - wie in Parlamenten - über Bindungsfähigkeit für sich und andere verfügen. *Zweitens* läßt sich fragen, ob nicht auch dann, wenn unter sicheren Erwartungen eine faire und einstimmige Entscheidung garantiert ist, ein Schleier der Unkenntnis irgendeine positive Funktion in bezug auf ein postkonstitutionelles Gefangenendilemma erfüllen kann. Im folgenden will ich einige mögliche Antworten auf diese Fragen skizzieren.

## I. Warum soviel Dissens?<sup>9</sup>

Auf die erste Frage - warum es in der Realität, in der es in der Regel keinen Schleier der Unkenntnis gibt, oft so schwer ist, Einstimmigkeit zu erzielen, obwohl dies nach Maßgabe des Modells so einfach erscheint - gibt es sicherlich eine Vielzahl möglicher Antworten: Eine erste könnte sein, daß Individuen nicht immer rational handeln. Menschen wählen, wie sich in den Forschungen der experimentellen Spieltheorie erwies<sup>10</sup>, häufig auch *dominierte* Strategien (die in der obigen Argumentation einfach als irrelevant ausgestrichen wurden). Handelt auch nur einer der Spieler in dieser Weise irrational, so kann bereits im konstitutionellen Zwei-Personen-Spiel unter perfekter Information das faire Gleichgewicht  $r_f$  verfehlt werden.

<sup>9</sup> Vgl. zu den Überlegungen in diesem Abschnitt auch Müller (1999).

<sup>10</sup> Vgl. Cooper/DeJong/Forsythe/Ross (1990).

Ein zweiter Grund mag sein, daß die Individuen ihrer eigenen Selbstbindungsfähigkeit mißtrauen. So ist natürlich die staatliche Durchsetzung von Regeln niemals vollkommen, sondern ist ihrerseits mit Kosten verbunden. Auch sie unterliegt daher einem ökonomischen Kalkül. Auf diese Weise kann es dazu kommen, daß die Reziprozität der Regelbefolgung nicht mehr gewährleistet ist: Wird aus Kostengründen eine zur Wahl stehende Regel nur gegen *einen* Spieler (oder in großen Gesellschaften: gegen eine Minderheit) durchgesetzt, so kann es - wenn dies vorausgesehen wird - ratsam sein, schon auf der konstitutionellen Ebene gegen diese Regel zu stimmen, um nicht selber in die „Sucker“-Position unilateraler Regelbefolgung gedrängt zu werden, in der man schlechter steht als in der Situation der Regellosigkeit.

Der Schlüssel zur Beantwortung dieser Fragen scheint mir jedoch zu sein, daß das Gefangenendilemma nicht das adäquate Modell ist, um den zu überwindenden Konflikt zu beschreiben. Einerseits ist die Spielstruktur des Gefangenendilemmas in der Realität nicht einfach gegeben, sondern ist selbst Gegenstand von Verhandlungen. Das Gefangenendilemma suggeriert, daß es genau eine paretoverbessernde Regel gibt, um die Situation des sozialen Konflikts zu überwinden. Diese Regel ist einfach da, sie fällt wie „Manna“ vom Himmel, und alle Beteiligten kennen sie. Die Möglichkeit eines *Theoriekonflikts*<sup>11</sup> - also eines Dissenses über das anzuwendende Mittel - wird hierbei nicht berücksichtigt. In der Realität ist ein solcher Theoriekonflikt aber wohl nahezu immer präsent. So läßt sich etwa das Dilemma der Übernutzung von Umwelt nicht nur mit Hilfe der Zertifikatelösung beheben, sondern möglicherweise ebenso gut mit Hilfe von Pigou-Steuern und Umweltauflagen. Auch gibt es - natürlich - nicht *die* Ordnungspolitik. In der Wettbewerbspolitik mögen Entscheidungsträger unterschiedliche Auffassungen haben über die Abgrenzung des relevanten Marktes oder das Verhalten von Oligopolisten. Und natürliche Monopole lassen sich nicht nur durch Preiskontrollen regulieren, sondern etwa auch durch eine Versteigerung des Monopols im Abstand von einigen Jahren.<sup>12</sup> Gibt es aber auf der konstitutionellen Ebene unterschiedliche Ansichten darüber, welche Regel aus einer Menge von paretoverbessernden Handlungsalternativen die adäquate ist, ist das konstitutionelle Spiel - auch unter sicheren Positionserwartungen bei Erfüllung der Payoff-Bedingung (6.3) - nicht mehr dominanzlösbar.

Andererseits scheint ganz allgemein die Einführung von Regeln in der praktischen Politik kaum je paretoverbessernd zu sein. Das Gefangenendilemma verdeutlicht nur die möglichen Wohlfahrtsgewinne einer Ordnung im Vergleich zu einer gedachten Situation, in der sich alle Beteiligten zumindest darin glei-

---

<sup>11</sup> Vgl. zu diesem Begriff *Vanberg/Buchanan* (1989) sowie die Diskussion in *Sutter* (1998).

<sup>12</sup> Vgl. *Demsetz* (1968).

chen, daß sie die Situation ohne Regel für schlechter halten als die Situation mit Regel. Schon Hobbes, der im Grundsatz von völlig verschiedenen Menschen ausging, trug diesem Punkt Rechnung durch seine Annahme der „natürlichen Gleichheit“ aller Menschen im „Krieg aller gegen alle“. „Denn was die Körperstärke betrifft, so ist der Schwächste stark genug, den Stärksten zu töten - entweder durch Hinterlist oder durch ein Bündnis mit anderen, die sich in derselben Gefahr wie er selbst befinden.“<sup>13</sup> In der Hobbesschen Ausgangssituation kann mithin *jeder* seines Mitmenschen „Wolf“ sein. Indem damit alle - auch die Stärksten - schutzbedürftig sind, können alle gewinnen durch die Installation des schutzgebenden „Leviathan“. Eine Kooperationsmöglichkeit - die Möglichkeit zu einer Pareto-Verbesserung - ist somit schon im postkonstitutionellen Spiel gegeben.

Demgegenüber scheint aber die typische politische Maßnahme eben *nicht* paretoverbessernd zu sein. Auch wenn etwa in unserem Umweltbeispiel de jure kein Eigentum an dem Kollektivgut „Luft“ definiert ist, so gibt es doch stets de facto einen Rechtsinhaber. Das Recht auf Nutzung dieses Guts gehört de facto der Fabrik, die bislang ihre Schadstoffe in die Luft entließ. Wenn zum Beispiel die Anwohner politisch durchsetzen, daß die Fabrik Schadstoff-Filter einzubauen hat, so werden zwar die Anwohner besser gestellt, die Fabrikeigner werden demgegenüber schlechter gestellt. Die Einführung des Filters führt daher realiter zu einer *Umverteilung von Rechten*. Beide Gesellschaftszustände - die Situation vor und nach Filtereinbau - sind also nach dem Pareto-Kriterium nicht vergleichbar.

Auch werden ordnungspolitische Maßnahmen nicht im „politikleeren Raum“ getroffen. Soll etwa ein natürliches Monopol - wie die Ordoliberalen fordern - Preiskontrollen und einem Kontrahierungszwang unterworfen werden, so gewinnen zwar die Nachfrager und die Arbeitnehmer dieses Anbieters; zumindest der Monopolist selbst aber verliert seine privilegierte Position. Wieder wird ein De-facto-Recht, das zuvor der Monopolist wahrnahm, an eine andere Personen-Gruppe umverteilt.

Der Grund, warum politische Entscheidungen getroffen werden, dürfte regelmäßig sein, daß die Gruppe der Entscheidungsträger (und die von ihr vertretene Klientel) wünscht, daß eine von ihr präferierte Maßnahme (und insbesondere deren Finanzierung) auch anderen Gruppen aufgezwungen wird, die nicht davon profitieren.<sup>14</sup> Private Bedürfnisse werden über den Staat befriedigt und die hierbei entstehenden Kosten auf andere gesellschaftliche Gruppen externali-

<sup>13</sup> Hobbes (1976 [1651]), S. 94.

<sup>14</sup> Vgl. de Jasay (1995 [1991]), S. 144.

siert.<sup>15</sup> Gäbe es nicht diese Möglichkeit, unbeteiligte Mitbürger über den Steuerzwang zur Finanzierung privater Bedürfnisse heranzuziehen, bestünde wohl vielfach überhaupt kein Anlaß, diese Entscheidung im Bereich der Politik zu tätigen.

Bedeutet aber Politik (fast) immer auch Umverteilung von Rechten und Pflichten durch Zwang, so sind die Zustände, zwischen denen staatliche Entscheidungsträger abzuwägen haben, wohl auch kaum mit dem Pareto-Kriterium vergleichbar, und das heißt zugleich: alle diese Zustände sind selbst paretooptimal, weil den Entscheidern nur die Wahl offensteht zwischen Handlungsalternativen, die allein dann in der Lage sind, einige Personen besser zu stellen, wenn gleichzeitig andere schlechter gestellt werden. Politik bedeutet daher häufig, anders als das Zwei-Personen-Gefangenendilemma suggeriert, zwischen Pareto-Optima zu wählen.

Genau dies ist auch der Grund, warum der Schleier der Unkenntnis angesichts eines Gefangenendilemmas mit Payoffs gemäß (6.3) scheitern muß. Denn nach dem vertragstheoretischen Wohlfahrtskriterium auf der Basis des Schleiers der Unkenntnis sind nur Pareto-Optima miteinander vergleichbar. Koordinationsgleichgewichte im konstitutionellen Spiel unter einem Schleier des Nichtwissens bzw. der Unsicherheit können daher nur solche Regeln sein, die im zugrunde liegenden Gefangenendilemma Pareto-Optima sind, das sind im Zwei-Personen-GD die faire Regel  $r_f$  und die beiden diskriminierenden Regeln  $r_{d1}$  und  $r_{d2}$ . Die Status-quo-Regel  $r_q$  - die dem Nash-Gleichgewicht des Gefangenendilemmas entspricht - kann demgegenüber niemals Koordinationsgleichgewicht des konstitutionellen Spiels unter einem Schleier der Unkenntnis sein, da diese Situation bereits im Gefangenendilemma von der beiderseitigen Kooperation pareto dominiert wird.

## II. Eine partielle Rehabilitierung des Schleiers

Die zweite oben aufgeworfene Frage ist, ob ein Schleier der Unkenntnis angesichts eines postkonstitutionellen Zwei- oder eines  $n$ -Personen-Gefangenendilemmas, das Bedingung (6.3) erfüllt, irgendeine sinnvolle Funktion haben kann. Die Antwort ist eindeutig positiv.

Wenn die Entscheidungsträger, aus welchen Gründen auch immer, es nicht schaffen, sich auf der konstitutionellen Ebene zu einigen, ist der hypothetische Imperativ verletzt, der von jedem GD-Spieler verlangt, dann und nur dann zu

---

<sup>15</sup> Für eine Analyse der Anreizprobleme, die aus einem solchen Verhalten resultieren können, vgl. Müller/Tietzel (1999).

kooperieren, wenn dies in seinem Interesse liegt. Andere Begründungen wären daher erforderlich, um die Spieler zur Kooperation in dem nachfolgenden Gefangenendilemma anzuhalten.

Einer dieser Gründe könnte ein kategorischer Imperativ sein, in sozialen Dilemmasituation immer so zu handeln, *als ob* man unsicher wäre, in welcher gesellschaftlichen Position man sich befindet. In diesem Sinne könnte die Metapher eines Schleiers der Unkenntnis auf der *post*konstitutionellen Ebene nützlich sein. Wenn jeder Spieler nach Maßgabe einer solchen Norm handelte (und alle Präferenzen die jeweiligen Fairneßbedingungen erfüllten), würden die Individuen trotz ihrer konstitutionellen Nichteinigung im Gefangenendilemma kooperieren.

## § 7 Schluß

Normen sind der Zement einer Gesellschaft; ohne sie wäre menschliches Zusammenleben unmöglich. Alle gesellschaftlichen Regeln und Institutionen - die Garantie von Freiheit und Demokratie, das Sozialstaatsgebot oder das Prinzip einer wettbewerblichen Wirtschaftsordnung - basieren auf grundlegenden Werturteilen. Gleichwohl sind Normen nicht wahrheitsfähig; ihr Geltungskriterium ist die subjektive Akzeptanz durch ihren jeweiligen Adressaten. Niemand ist gezwungen, ein Werturteil, das ein anderer vorschlägt, zu teilen oder gar nach diesem zu handeln. „Handle nach Norm  $x$ !“ oder „Präferiere Regel  $y$ !“ sind daher besonders schwache Imperative. Ob man sich nach ihnen richtet oder nicht, obliegt einem persönlichen Willensentschluß.

Unterliegen Normen aber einer allein subjektiven Geltung, so wäre es immerhin ein Erfolg, wenn sich wenigstens gute Gründe finden ließen, welche es allen Beteiligten einfacher machten, sich auf die grundlegenden Normen und Organisationsprinzipien des gesellschaftlichen Miteinanders zu einigen. Als einen solchen guten Grund für die Akzeptanz bestimmter Gerechtigkeitsnormen oder sozialer Institutionen führen die Theorien des Gesellschaftsvertrags traditionell an, daß selbst die krassesten Egoisten sich unter bestimmten Umständen hierauf einigen würden. Wenn sogar ein so buchstäblich „a-soziales Wesen“<sup>1</sup> wie der kühl und leidenschaftslos kalkulierende *Homo oeconomicus* in bestimmten Situationen aus Eigeninteresse wie ein moralisches Individuum handelt, dann liegt - so wollen uns die modernen Varianten dieser Kontrakttheorien nahelegen - ein solches Verhalten erst recht im Interesse so fehlbarer Menschen wie uns selbst. Die vorliegende Arbeit versuchte, die Eignung der in der Ökonomik wohl einflußreichsten Vertragstheorien von John Rawls und James Buchanan im Hinblick auf die Erreichung dieses Begründungsanliegens zu überprüfen.

Zunächst wurde versucht, das vertragstheoretische Wohlfahrtskriterium ganz allgemein als *Collective choice rule* im Sinne der *Social-choice*-Theorie zu formulieren und den in der Literatur weitgehend diffus bleibenden methodologischen Status des vertragstheoretischen Arguments zu klären: Eine beliebige gesellschaftliche Regel (Situation)  $r$  ist nach dem konstitutionellen Wohlfahrtskriterium einer anderen Regel  $q$  dann kollektiv vorzuziehen, wenn  $r$ , nicht aber

---

<sup>1</sup> Tietzel (1985), S. 41.



$q$ , im Vergleich zu einem fiktiven und als (verfahrens-)gerecht angesehenen „Urzustand“  $z$  eine starke Pareto-Verbesserung ist.

In seiner logischen Struktur gleicht dieses kontraktualistische Argument einer üblichen ökonomischen Rationalwählerklärung oder -prognose, in deren Explanans die Situationsannahmen über die Verhältnisse in dem „gerechten“ Urzustand  $z$  sowie die Rationalverhaltenshypothese der neoklassischen Wirtschaftstheorie stehen. Die wesentliche Besonderheit der Vertragstheorie besteht jedoch darin, daß sie die Erfüllung der empirischen Adäquatheitsbedingung, nach der alle Annahmen im Explanans wahr oder zumindest hochgradig bewährt sein sollen, überhaupt nicht anstrebt. Die faktische Falschheit der Annahmen im Explanans dieses Arguments ist vielmehr eine konstitutive Voraussetzung für die Sinnhaftigkeit der Anwendung des hypothetischen Gedankenexperiments.

Wegen dieser Falschheit mindestens einiger Annahmen in seinem Explanans ist das vertragstheoretische Argument selbst im Idealfall äußerst schwach: Selbst wenn sie in sich schlüssig ist, vermag eine hypothetische Vertragstheorie nicht mehr nachzuweisen, als daß Welten möglich sind, in denen selbst der egoistischste *Homo oeconomicus* sich so verhalten würde, *als ob* er eine bestimmte Moralnorm teilte. Zu wissen, daß es eine solche Situation gibt, kann aber allenfalls ein psychologischer Appell sein, die Herbeiführung des fiktiven Vertragsergebnisses auch in realen Entscheidungssituationen für geboten zu halten. In logischer Hinsicht beruht die Ableitung einer realen Verpflichtung aus einem fiktiven Vertrag aber auf einem unzulässigen Induktionsschluß: Aus der Rationalität einer Handlung in einer *bestimmten* Situation folgt nicht die Rationalität dieser Handlung in *allen* Situationen.

Doch das vertragstheoretische Argument, wie es in der ökonomischen Theorie verwendet wird, ist nicht in sich schlüssig. Wie in der vorliegenden Arbeit nachgewiesen werden konnte, verletzen Rawls' „Theorie der Gerechtigkeit“ und die beiden wichtigsten Varianten von Buchanans normativer Verfassungsökonomik auch jene logische Adäquatheitsbedingung deduktiv-nomologischer Erklärungs- und Prognoseargumente, nach der die Explanandum-Aussage eines Arguments logisch aus den Annahmen im Explanans zu folgen hat. Weder die Prämissen der Rawlsschen noch jene der Buchananschen Theorie, so ergab die Analyse, sind hinreichend, um die jeweils behaupteten Explananda - in Form von Aussagen über die Wahl bestimmter sozialer Regeln oder Prinzipien - zu deduzieren. Beide Ansätze sind lediglich partielle (hypothetische) Erklärungs- bzw. Prognoseargumente.

Dieser Befund wurde in § 2 zunächst für die rein paretianische, ohne die Prämisse eines Schleiers der Unkenntnis operierende „starke“ Vertragstheorie belegt, die Buchanan in „Die Grenzen der Freiheit“ formuliert: Der Rechts-

schutzstaat, den Buchanan als wahrscheinliches Vertragsergebnis behauptet, wird nur dann das Ergebnis einer Einigung im Urzustand sein, wenn die Situationsannahmen im Explanans um Annahmen im Sinne solcher externen Werturteile erweitert werden, welche der paretianische Ansatz gerade aus der Vertragstheorie heraushalten soll.

In den §§ 3 und 4 wurde darüber hinaus gezeigt, daß auch die Versuche scheitern müssen, konkrete Gerechtigkeitskonzeptionen aus einer Situation herzuleiten, in der die fiktiven Entscheidungsträger, weil sie unter einen Schleier der Unkenntnis versetzt sind, ihre eigenen Partialinteressen nicht oder nicht vollständig kennen. In der Rawlsschen „Theorie der Gerechtigkeit“ dient diese Annahme dazu, Entscheidungsergebnisse zu generieren, die mit einer ethischen Universalisierungsnorm - dem kantischen kategorischen Imperativ - vereinbar sind. Buchanan indes behauptet die Existenz eines Schleiers der Unsicherheit als ein empirisches Phänomen - eine Behauptung, die sich bei näherer Betrachtung in § 3 nicht nur als hochgradig unplausibel, sondern auch als zirkulär erwies: Die Existenz des Schleiers der Unsicherheit setzt mit der Universalisierbarkeit jene Eigenschaft von Regeln bereits voraus, die diese Annahme erst zur Folge haben soll. Der eigentliche Sinn der Annahme eines Schleiers der Unsicherheit, so habe ich deshalb vermutet und anhand von Textstellen zu belegen versucht, liegt daher auch bei Buchanan vor allem in der Herleitung gewünschter - d.h. universalisierbarer - Theorieergebnisse.

Einerseits habe ich argumentiert, daß keine der beiden Varianten der Urzustandssituation unter dem Schleier der Unkenntnis eine konkrete Gerechtigkeitskonzeption impliziert, sondern - unter entsprechenden Situationsannahmen - mit jedem denkbaren Verlauf der Bergsonschen Wohlfahrtsfunktion vereinbar sind. Ihr insbesondere von Rawls betontes zentrales Anliegen einer „Normbegründung“ - d.h. der Deduktion einer speziellen Gerechtigkeitskonzeption aus den Prämissen des Urzustands - müssen die mit dem Schleier der Unkenntnis argumentierenden Vertragstheorien also verfehlen.

Andererseits ist die stillschweigende Einstimmigkeit der Entscheidungsträger unter dem Schleier der Unkenntnis, die auf den ersten Blick so natürlich zu sein scheint, keine Selbstverständlichkeit: Der Schleier der Unsicherheit kann - da Buchanan eine lediglich partielle Universalisierung unterstellt, um völlig ohne intersubjektive Nutzenvergleiche auszukommen - einen Konsens der Individuen im Urzustand überhaupt nicht garantieren; der Rawlssche Schleier des Nichtwissens hingegen induziert nur deshalb stets einstimmige Entscheidungen, weil die „Theorie der Gerechtigkeit“ mit ihrem untauglichen Grundgüterkonzept auf einer Annahme basiert, die von der Problematik intersubjektiver Nutzenvergleiche lediglich abstrahiert, anstatt sie zu lösen. Die Rawlssche wie die Buchanansche Verfahrensweise, unter dem Schleier der Unkenntnis den Entscheidungskalkül eines beliebigen Individuums als repräsentativ für alle anderen Entschei-

dungsträger anzusehen, wird vor dem Hintergrund dieser Untersuchungsergebnisse äußerst fragwürdig.

In den §§ 5 und 6 konnte überdies gezeigt werden, daß die Rawlssche und die Buchanansche Wohlfahrtsfunktion schon deshalb nicht aus den Situationsannahmen über den Urzustand unter dem Schleier der Unkenntnis folgen können, weil sie Implikationen haben, die mit der Universalisierungsnorm, für die diese Annahmen stehen, nicht vereinbar sind. Wird ein auf der konstitutionellen Ebene geschlossener Gesellschaftsvertrag - wie seit dem Hobbesschen „Leviathan“ traditionell angenommen - von einem einfachen (Zwei- oder Multi-Personen-) Gefangenendilemma bedroht, so kann unter bestimmten Payoff-Konstellationen weder ein Rawlsscher Schleier des Nichtwissens noch ein Buchananscher Schleier der Unsicherheit die einstimmige konstitutionelle Wahl einer diskriminierenden, mit dem Universalisierungsgedanken unvereinbaren Verhaltensregel verhindern. Aus den genannten Gründen kann darüber hinaus Buchanans Schleier der Unsicherheit eine Einigung nicht, Rawls' Schleier des Nichtwissens hingegen nur formal - durch Ausblendung des Problems intersubjektiver Nutzenvergleiche - garantieren.

Statt dessen, so wurde dargelegt, kann es in diesen Situationen gerade die Annahme *sicherer* Erwartungen sein, welche eine faire und einstimmige Einigung ermöglicht. Soll auf der konstitutionellen Ebene eine Regel für das einfache Gefangenendilemma gewählt werden, so ist im konstitutionellen Regelwahlspiel die Einigung auf die faire (universalisierbare) Regel im Zwei-Personen-Fall unter allen Umständen und im  $n$ -Personen-Fall unter einer schon bei Hobbes anzutreffenden Bedingung an die individuellen GD-Payoffs das einzige Nash-Gleichgewicht in sukzessiver Dominanz. Entgegen einer in der Literatur häufig anzutreffenden Auffassung ist der konstitutionell zu lösende Konflikt im Gefangenendilemma allein ein Selbstbindungs- und kein Interessenproblem. Ein Schleier der Unkenntnis schafft hier erst das Problem, das zu lösen er vorgesehen wurde.

Die vorstehend erzielten Ergebnisse reichen über den Gegenstand der beiden untersuchten Theorien hinaus und werfen ein Schlaglicht auf das Verhältnis zwischen dem Einstimmigkeits- und dem konstitutionellen Wohlfahrtskriterium der hypothetischen (schwachen) Vertragstheorien insgesamt. Der hypothetische Kontraktualismus will eine Einigungsmöglichkeit von Individuen aufzeigen, wo diese eigentlich ausgeschlossen ist: in Nichtkooperationssituationen. Indem es postkonstitutionelle Pareto-Optima miteinander vergleichbar macht, erlaubt das vertragstheoretische Argument genau jene Güterabwägung, die das Pareto-Kriterium von vornherein ausschließt. Es soll uns zeigen, unter welchen Umständen staatlicher Zwang und damit die Schlechterstellung einiger legitim sein soll, um andere besser zu stellen. Das Pareto-Kriterium selbst - und damit die moralische Grundposition, daß jeder ein Veto haben soll gegen eine ihm miß-

liebige staatliche Zwangsmaßnahme - scheinen wir dabei nicht einmal aufgeben zu müssen; es wird, so will uns die Vertragstheorie glauben machen, lediglich auf der „konstitutionellen Ebene“ - statt der postkonstitutionellen - angewendet.<sup>2</sup>

Diese Arbeit macht jedoch klar, daß dies eine Illusion ist: Mit der „konstitutionellen Anwendung“ des Pareto-Kriteriums allein ist, wie die Analyse der starken Vertragstheorie in § 2 zeigte, nämlich nicht viel gewonnen. Individuen, die keine gemeinsamen Interessen haben, werden sich auch dann nicht einigen, wenn sie über die vollkommene Fähigkeit zum bindenden Vertragsschluß verfügen. Bestünde das vertragstheoretische Ziel allein in der Lösung von *Commitment*-Problemen, so hätte man den Betroffenen lediglich Selbstbindungsmacht zu geben, und sie würden sich ohne weiteres *tatsächlich* einigen; eine hypothetische Rechtfertigung von Regeln wäre in dieser Situation also überflüssig.

Doch die (schwache) Vertragstheorie will in Wirklichkeit viel mehr als das: Sie will das *Interessenproblem* lösen, das für das „Faktum des Pluralismus“ verantwortlich ist. Die eigentlich zentrale Prämisse der Theorien des hypothetischen Gesellschaftsvertrags ist daher nicht, daß die Individuen in der Lage sind, bindende Verträge zu schließen; entscheidend sind vielmehr die jeweiligen Annahmen über die konstitutionelle Entscheidungssituation. Was so unschuldig als „Tatsache“ im Urzustand behauptet wird, bedeutet in Wahrheit eine fundamentale Verschiebung der ethischen Ausgangsposition, die weit voraussetzungsreicher ist als das vorgeblich zentrale Einstimmigkeitskriterium. Unter der Annahme eines (vollkommenen) Buchananschen Schleiers der Unsicherheit etwa läuft die angebliche „konstitutionelle Anwendung“ des Pareto-Kriteriums letztlich auf nichts anderes hinaus als eine Entscheidung nach Maßgabe des Kompensationskriteriums von Kaldor und Hicks.<sup>3</sup> Denn ebenso wie nach Kaldor und Hicks, gilt nach der Buchananschen Vertragstheorie ein sozialer Zustand *r* dann gegenüber einem anderen Zustand *q* als „sozial präferiert“, wenn erstens einige besser gestellt werden und zweitens der Wohlfahrtszuwachs der Gewinner Nutzeneinbuße der Verlierer übersteigt. Und wie auch bei Kaldor und Hicks bleibt diese Kompensation nach der Theorie des hypothetischen Gesell-

---

<sup>2</sup> Besonders deutlich wird der Anspruch einer lediglich „konstitutionellen Anwendung“ des Pareto-Kriteriums z. B. bei *Buchanan* (1991 [1988]), S. 47: „... we [= *Buchanan/Tullock* (1962); *C. M.*] were simply advancing the Wickseil-Pareto criterion one stage upward in the choice-making hierarchy“. Vgl. ähnlich *Buchanan/Tullock* (1962), S. 94; *Buchanan/Congleton* (1998), S. 17.

<sup>3</sup> Vgl. *Kaldor* (1939); *Hicks* (1939).

schaftsvertrags ein rein fiktives Gedankenexperiment, denn sie muß nur *möglich* sein, nicht aber durchgeführt werden.<sup>4</sup>

Das Pareto-Kriterium aber ist damit aufgegeben. Es ist daher irreführend, wenn Buchanan und Tullock darüber philosophieren, daß man die unter ihrem Schleier der Unsicherheit konstitutionell „gerechtfertigte“ Umverteilung nur deshalb nicht „paretooptimal“ nennen sollte, weil der Terminus einer „pareto-optimalen Umverteilung“ schon anderweitig - auf der *post*konstitutionellen Ebene - belegt sei.<sup>5</sup> In dem Augenblick nämlich, in dem auf der konstitutionellen Ebene unsichere Erwartungen oder irgendwelche anderen „Annahmen“ eingeführt werden, die das Entscheidungsverhalten der Individuen beeinflussen, kommt es auf die Präferenzen der Individuen überhaupt nicht mehr an. Die Situationsannahmen rauben ihnen die Voraussetzungen für das Einlegen von Widerspruch; sie entziehen den Entscheidungsträgern de facto wieder ihr Veto-recht, das die Vertragstheorie ihnen mit dem Einstimmigkeitskriterium zuvor erst so großzügig eingeräumt hatte. Daher täuscht die Harmonie im Urzustand, gleichgültig, ob dieser durch sozio-ökonomische Gleichheit<sup>6</sup>, die fiktive Gleichheit von Handlungs- und Lebenschancen<sup>7</sup>, eine herrschaftsfreie Sprech-situation<sup>8</sup> oder durch einen Schleier der Unkenntnis gekennzeichnet ist: Post-konstitutionell - und damit in der Realität - bedeutet die Umsetzung der vertragstheoretischen Gestaltungsempfehlungen nahezu nie Zustimmung, sondern fast immer Zwang.

Was aber bleibt nach alledem vom vertragstheoretischen Unternehmen? Die Vertragstheorie setzt psychologische Verführung an die Stelle logisch-stringenter Deduktion. Natürlich wird sich mancher - trotz der logisch fehlerhaften Ableitungsbeziehungen - aufgrund der hohen Suggestionskraft des Gedankenexperiments des Urzustands von der Akzeptanz der jeweils favorisierten Gerechtigkeitskonzeption überzeugen lassen; insofern kann das kontraktualistische Begründungsanliegen durchaus erfolgreich sein. Als „wissenschaftlich begründet“ in dem Sinne, daß man bei Akzeptanz der normativen Ausgangsposition nach den Regeln der Logik gezwungen wäre, eine spezielle Wohlfahrts-

---

<sup>4</sup> Einen Unterschied zwischen dem *Buchananschen* Schleier der Unsicherheit und dem Kompensationskriterium nach *Kaldor* und *Hicks* mag man jedoch im Grad der Universalisierung sehen, der bei *Buchanan*, wie ausgeführt, nur partiell ist, so daß nur *eigene* Nutzenwerte des Kalkulierenden hypothetisch miteinander verrechnet werden. Nach *Kaldor* (1939) und *Hicks* (1939) hingegen sind offenbar die objektiven Nutzenwerte unterschiedlicher Individuen hypothetisch zu saldieren, so daß die Universalisierung - wie bei *Rawls* - vollkommen ist.

<sup>5</sup> Vgl. *Buchanan/Tullock* (1962), S. 195.

<sup>6</sup> Vgl. *Rousseau* (1777 [1755]), *ders.* (1777 [1762]).

<sup>7</sup> Vgl. *Wildt* (1996).

<sup>8</sup> Vgl. *Habermas* (1983).

funktion zu teilen, wird man die vertragstheoretischen Konklusionen indes nicht ansehen können.

Einen Platz in der ökonomischen Theorie hat das vertragstheoretische Argument von Rawls oder Buchanan daher höchstens als ein Mittel der *Normillustration*, nicht aber der Normbegründung. Statt etwa zu fragen, welche Regel nach Maßgabe des ethischen Universalisierungsprinzip zu wählen wäre, ließe es sich unter Umständen einfacher feststellen, ob man auch unter einem Schleier der Unkenntnis hierfür stimmen würde. Die Verwendung der ethischen Verallgemeinerungsnorm selbst vermag die Vertragstheorie jedoch nicht zu begründen. Insoweit gilt heute nichts anderes als Jahrzehnte zuvor zu Werner Sombarts Zeiten: „Für Werte“, so Sombart, „*lebt* man, für Werte *stirbt* man, wenn es notwendig ist. Werte aber *beweist* man nicht.“<sup>9</sup>

---

<sup>9</sup> Sombart (1930), S. 83 (Hervorhebungen im Original).

## Literaturverzeichnis \*

- Albert, H.* (1967), Wertfreiheit als methodisches Prinzip. Zur Frage der Notwendigkeit einer normativen Sozialwissenschaft, in: E. Topitsch (Hrsg.), Logik der Sozialwissenschaften, 10. Auflage, Köln/Berlin, S. 181-210.
- (1998), Bemerkungen zur Wertproblematik. Von der Bewertung des Sozialprodukts zur Analyse der sozialen Ordnung. Vortragsreihe Lectiones Jenenses, Schriftenreihe des Max-Planck-Instituts zur Erforschung von Wirtschaftssystemen, Heft 15.
- Alexander, S.* (1974), Social Evaluation through Notional Choice. Quarterly Journal of Economics 88, S. 597-624.
- Arneson, R. J.* (1990), Primary Goods Reconsidered. Nous 24, S. 429-454.
- (1992), Property Rights in Persons, in: E. Frenkel Paul/F. D. Miller/J. Paul (Hrsg.), Economic Rights, Cambridge, S. 201-230.
- Arnold, V.* (1992), Theorie der Kollektivgüter, München.
- Arrow, K. J.* (1963), Social Choice and Individual Values, 2. Auflage, New Haven/London.
- (1973), Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice. Journal of Philosophy, 70, S. 245-263.
- Aumann, R. J.* (1987), Correlated Equilibrium as an Expression of Bayesian Rationality. Econometrica 55, S. 1-18.
- Axelrod, R.* (1984), The Evolution of Cooperation, New York.
- Ballestrem, K.* (1983), Vertragstheoretische Ansätze in der Politischen Philosophie. Zeitschrift für Politik 30 (N.F.), S. 1-17.
- Barbosa, A. S. P.* (1974), Bargaining Costs and Social Choice Under Uncertainty. Public Choice 17, S. 85-91.
- (1978), The Constitutional Approach to the Fiscal Process, Dissertation, Blacksburg, Virginia.

---

\* Die Zusätze in eckigen Klammern verweisen auf eine frühere Erstveröffentlichung (nicht lediglich frühere Auflage) der jeweiligen Quelle.



- Barr, W. F.* (1974), A Pragmatic Analysis of Idealization in Physics. *Philosophy of Science* 41, S. 48-64.
- Barro, R.* (1974), Are Government Bonds Net Wealth? *Journal of Political Economy* 82, S. 1095-1118.
- Barry, B.* (1973), *The Liberal Theory of Justice: A Critical Examination of the Principal Doctrines in 'A Theory of Justice' by John Rawls*, Oxford.
- (1979), Is Democracy Special?, in: P. Laslett, J. Fishkin (Hrsg.), *Philosophy, Politics and Society*, New Haven/Oxford, S. 172-189.
  - (1984), Unanimity, Agreement, and Liberalism. A Critique of James Buchanan's Social Philosophy. *Political Theory* 12, S. 579-596.
  - (1989), *A Treatise on Social Justice. Volume I: Theories of Justice*, Berkeley/Los Angeles.
- Beauchamp, T. L.* (1980), Distributive Justice and the Difference Principle, in: H. G. Blocker/E. H. Smith (Hrsg.): *John Rawls' Theory of Social Justice. An Introduction*, Athens, S. 132-161.
- Beckmann, K.* (1998), *Analytische Grundlagen einer Finanzverfassung. Regelbegründungen und Regelbindungsprobleme aus finanzwissenschaftlicher Sicht*, Frankfurt am Main u.a.
- Benn, S. I./Peters, R. S.* (1959), *The Principles of Political Thought*, New York.
- Bentham, J.* (1970 [1781]), *An Introduction to the Principles of Morals and Legislation*, hrsg. von J. H. Burns und H. L. A. Hart, London.
- Bergson, A.* (1938), A Reformulation of Certain Aspects of Welfare Economics. *Quarterly Journal of Economics* 52, S. 310-334.
- (1948), Socialist Economics, in: H. S. Ellis (Hrsg.), *A Survey of Contemporary Economics*, Band 1, Homewood, S. 412-448.
- Binmore, K.* (1989), Social Contract I: Harsanyi and Rawls. *Economic Journal* 99 (Supplement 1989), S. 84-102.
- (1992), *Fun and Games. A Text on Game Theory*, Lexington/Toronto.
  - (1994), *Game Theory and the Social Contract. Volume I: Playing Fair*, Cambridge/London.
  - (1996), Suppose Everybody Behaved Like That?, in: B. Allen (Hrsg.), *Economics in a Changing World*, Band 2: Microeconomics, Basingstoke, S. 25-62.
- Bohnen, A.* (1964), *Die utilitaristische Ethik als Grundlage der modernen Wohlfahrtsökonomik*, Göttingen.

- Bohnet, I.* (1997), Kooperation und Kommunikation. Ein ökonomische Analyse individueller Entscheidungen, Tübingen.
- Brembs, B.* (1996), Chaos, Cheating and Cooperation: Potential Solutions to the Prisoner's Dilemma. *Oikos* 76, S. 14-24.
- Brennan, G.* (1973), Pareto Desirable Redistribution: The Non-Altruistic Dimension, in: *Public Choice*, 14, S. 43-67.
- Brennan, G./Buchanan, J. M.* (1983), Tax Structure and Social Overhead Capital, in: K. Roskamp (Hrsg.), *Public Finance and Economic Growth*, Detroit, S. 41-54.
- (1988 [1980]), Besteuerung und Staatsgewalt. Analytische Grundlagen einer Finanzverfassung, Hamburg.
  - (1993 [1985]), Die Begründung von Regeln, Tübingen.
- Brennan, G./Lomasky, L.* (1984), Inefficient Unanimity. *Journal of Applied Philosophy* 1, S. 151-163.
- (1993), Democracy and Decision. The Pure Theory of Electoral Preference, Cambridge.
- Brülisauer, B.* (1980), Die Goldene Regel - Aspekte einer dem kategorischen Imperativ verwandten Grundnorm. *Kant-Studien* 71, S. 325-345.
- Buchanan, J. M.* (1960 [1954]), Social Choice, Democracy, and Free Markets, in: J. M. Buchanan, *Fiscal Theory and Political Economy. Selected Essays*, Chapel Hill, S. 75-89.
- (1960 [1959]), Positive Economics, Welfare Economics, and Political Economy, in: J. M. Buchanan, *Fiscal Theory and Political Economy. Selected Essays*, Chapel Hill, S. 105-124.
  - (1966), Externality in Tax Response. *Southern Economic Journal* 23, S. 35-42.
  - (1967), Public Finance in Democratic Process. *Fiscal Institutions and Individual Choice*, Chapel Hill.
  - (1968), What Kind of Redistribution Do We Want? *Economica* 35, S. 185-190.
  - (1972), Rawls on Justice as Fairness. *Public Choice* 13, S. 123-128.
  - (1977 [1962]), The Relevance of Pareto Optimality, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 215-234.
  - (1977 [1972]), Before Public Choice, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 81-93.

- (1977 [1976]), A Hobbesian Interpretation of the Rawlsian Difference Principle, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 194-211.
  - (1977 [1977]), Law and the Invisible Hand, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 25-39.
  - (1977a), Notes on Justice in Contract, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 123-134.
  - (1977b), The Use and Abuse of Contract, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 135-147.
  - (1977c), Democratic Values in Taxation, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 243-253.
  - (1979), *What Should Economists Do?* Indianapolis.
  - (1984 [1975]), *Die Grenzen der Freiheit. Zwischen Anarchie und Leviathan*, Tübingen.
  - (1986), *Liberty, Market and State. Political Economy in the 1980s*, Sussex.
  - (1987), *Constitutional Economics*, in: J. Eatwell u.a. (Hrsg.), *The New Palgrave*, Band 3, London u.a., S. 585-588.
  - (1989), The Relatively Absolute Absolutes, in: J.M. Buchanan, *Essays in the Political Economy*, Honolulu, S. 32-46.
  - (1989 [1986]), Die Verfassung der Wirtschaftspolitik. Nobel-Lesung vom 8. Dezember 1986, in: H. C. Recktenwald (Hrsg.), *Die Nobelpreisträger der ökonomischen Wissenschaft 1969-1988*, Band II, Düsseldorf, S. 932-949.
  - (1990), The Domain of Constitutional Economics. *Constitutional Political Economy* 1, S. 1-18.
  - (1991 [1988]), The Justification of the Compound Republic: A Retrospective Interpretation of The Calculus of Consent, in: J. M. Buchanan, *The Economics and the Ethics of Constitutional Order*, Ann Arbor, S. 43-49 (Kapitel 4).
- Buchanan, J. M./Bush, W. C. (1977 [1974]), Political Constraints on Contractual Redistribution*, in: J. M. Buchanan, *Freedom in Constitutional Contract*, College Station/London, S. 186-193.
- Buchanan, J. M./Congleton, R. D. (1998), Politics by Principle, Not Interest. Toward Nondiscriminatory Democracy*, Cambridge.
- Buchanan, J. M./Faith, R. (1980), Subjective Elements in Rawlsian Agreement on Distributional Rules. Economic Inquiry* 18, s. 23-38.

- Buchanan, J. M./Tullock, G.* (1962), *The Calculus of Consent*, Ann Arbor.
- Bush, W. C./Mayer, L. S.* (1974), Some Implications of Anarchy for the Distribution of Property. *Journal of Economic Theory* 8, S. 401-412.
- Cassel, D./Ramb, B.-T./Thieme, H.-J.* (Hrsg.) (1988), *Ordnungspolitik*, München.
- Chmielewicz, K.* (1979), *Forschungskonzeptionen der Wirtschaftswissenschaft*, 2. Auflage, Stuttgart.
- Cooper, R.* (1999), *Coordination Games. Complementaries and Macroeconomics*, Cambridge.
- Cooper, R./DeJong, D./Forsythe, R./Ross, T.* (1990), Selection Criteria in Coordination Games: Some Experimental Result. *American Economic Review* 80, S. 218-233.
- Corrado, G.* (1980), Rawls, Games and Economic Theory, in: H. G. Blocker/E. H. Smith (Hrsg.): *John Rawls' Theory of Social Justice. An Introduction*, Athens, S. 71-109.
- van Damme, E.* (1991), *Stability and Perfection of Nash Equilibria*, 2. Auflage, Berlin u.a.: Springer-Verlag.
- Darwall, S. L.* (1980), Is There a Kantian Foundation for Rawlsian Justice?, in: H. G. Blocker/E. H. Smith (Hrsg.): *John Rawls' Theory of Social Justice. An Introduction*, Athens, S. 311-345.
- Dasgupta, P./Heal, G. M.* (1979), *Economic Theory and Exhaustible Resources*, Cambridge.
- Demsetz, H.* (1968), Why Regulate Utilities? *Journal of Law and Economics* 11, S. 55-65.
- Dixit, A. K./Nalebuff, B. J.* (1991), *Thinking Strategically. The Competitive Edge in Business, Politics, and Everyday Life*, New York/London.
- Dworkin, R.* (1975 [1973]), The Original Position, in: N. Daniels (Hrsg.), *Reading Rawls. Critical Studies on Rawls' 'A Theory of Justice'*, Oxford, S. 18-53.
- (1981), What is Equality? Part 1: Equality of Welfare, Part 2: Equality of Resources. *Philosophy and Public Affairs*, 10, S. 185-246 und S. 283-345.
- Eichberger, J./Pethig, R.* (1993), Evaluation of Collective Choice Rules. Expected Cost Minimization versus Expected Utility Maximization?, mimeo.
- (1994), Constitutional Choice of Rules. *European Journal of Political Economy* 10, S. 311-337.

- Ellsworth, L.* (1978), Decision-Theoretic Analysis of Rawls Original Position, in: C. A. Hooker/J. J. Leach/E. F. McClennen (Hrsg.), Foundations and Applications of Decision Theory, Band 2 Dordrecht/Boston, S. 29-45.
- Epikur* (1949), Philosophie der Freude. Eine Auswahl aus seinen Schriften. Übersetzt, erläutert und eingeleitet von Johannes Mewaldt, Stuttgart.
- Eschenburg, R.* (1980), Die Legitimation von Ordnungen, in: W. Dettling (Hrsg.), Die Zähmung des Leviathan. Neue Wege der Ordnungspolitik, Baden-Baden, S. 21-38.
- Faber, M.* (1973), Einstimmigkeitsregel und Einkommensumverteilung. *Kyklos* 26, S. 36-57.
- Fetscher, I.* (1968), Rousseaus Politische Philosophie. Zur Geschichte des demokratischen Freiheitsbegriffs, 2. Auflage, Neuwied/Berlin.
- Frey, B. S./Eichenberger, R.* (1989a), Zur Bedeutung entscheidungstheoretischer Anomalien für die Ökonomik. *Jahrbücher für Nationalökonomie und Statistik* 206, S. 81-101.
- (1989b), Anomalies and Institutions. *Journal of Institutional and Theoretical Economics* 145, S. 423-437.
- Frohlich, N.* (1992), An Impartial Reasoning Solution to the Prisoner's Dilemma. *Public Choice* 74, S. 447-460.
- Frohlich, N./Oppenheimer, J. A.* (1992), Choosing Justice. An Experimental Approach to Ethical Theory, Berkeley u.a..
- (1996a), Experiencing Impartiality to Invoke Fairness in the n-PD. Some Experimental Results. *Public Choice* 86, S. 117-135.
  - (1996b), When is Universal Contribution Best for the Group? *Journal of Conflict Resolution* 40, S. 502-516.
- Frohlich, N./Oppenheimer, J. A./Eavey, C. L.* (1987a), Laboratory Results on Rawls's Distributive Justice. *British Journal of Political Science* 17, S. 1-21.
- (1987b), Choices of Principles of Distributive Justice in Experimental Groups. *American Journal of Political Science* 31, S. 606-636.
- Fudenberg, D./Maskin, E.* (1986), The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica* 54, S. 533-554.
- Fudenberg, D./Tirole, J.* (1991), Game Theory, Cambridge/London.
- Gaa, J. C.* (1984), The Stability of Bargains Behind the Veil of Ignorance. *Theory and Decision* 17, S. 119-133.

- Gaertner*, W. (1988), Untergrundwirtschaft, Steuerhinterziehung und Moral, in: H. Hesse (Hrsg.), *Wirtschaftswissenschaft und Ethik*, Berlin 1988, S. 109-130.
- (1994), Pareto-Effizienz und normative Ökonomik, in: K. Homann (Hrsg.): *Wirtschaftsethische Perspektiven I: Theorie, Ordnungsfragen, Internationale Institutionen*, Berlin, S. 31-51.
- Gauthier*, D. P. (1969), *The Logic of Leviathan*, Oxford.
- (1986), *Morals by Agreement*, Oxford.
- (1993), Between Hobbes and Rawls, in: D. Gauthier/R. Sugden (Hrsg.), *Rationality, Justice and the Social Contract. Themes from Morals by Agreement*, New York u.a., S. 24-39.
- Gibbard*, A. (1979), Disparate Goods and Rawls' Difference Principle: A Social Choice Theoretic Treatment. *Theory and Decision* 11, S. 267-288.
- Gibbons*, R. (1992), *Game Theory for Applied Economists*, Princeton.
- Giersch*, T. (1993), *Bergson-Wohlfahrtsfunktion und normative Ökonomie*, Frankfurt am Main u.a..
- Goldberg*, V. P. (1974), Public Choice - Property Rights. *Journal of Economic Issues* 8, S. 555-579.
- Gordon*, S. (1973), John Rawls' Difference Principle, Utilitarianism, and the Optimum Degree of Inequality. *Journal of Philosophy* 70, S. 275-280.
- (1976), The New Contractarians. *Journal of Political Economy* 84, S. 573-590.
- Gough*, J. W. (1957), *The Social Contract*, 2. Auflage, Oxford.
- Graaff*, J. de V. (1957), *Theoretical Welfare Economics*, Cambridge u.a.
- Gray*, J. (1990), Buchanan on Liberty. *Constitutional Political Economy* 1, S. 149-168.
- Grossman*, H. I./*Kim*, M. (1995), Swords or Plowshares? A Theory of the Security of Claims to Property. *Journal of Political Economy* 103, S. 1275-1288.
- Güth*, W./*Leininger*, W./*Stephan*, G. (1991), On Supergames and Folk Theorems - A Conceptual Discussion, in: R. Selten (Hrsg.), *Game Equilibrium Models. Vol. II: Methods, Morals, and Markets*, Berlin u.a., S. 56-70.
- Güth*, W./*Stephan*, J. (1984), Equilibrium Selection in the One Seller and Many Buyers-Game. *Zeitschrift für Nationalökonomie* 44, S. 267-281.
- Habermas*, J. (1983), *Moralbewußtsein und kommunikatives Handeln*, Frankfurt am Main.

- Hamburger, H.* (1973), N-Person Prisoner's Dilemma. *Journal of Mathematical Sociology* 3, S. 27-48.
- (1979), *Games as Models of Social Phenomena*, San Francisco.
- Hampton, J.* (1980), Contracts and Choices: Does Rawls have a Social Contract Theory? *Journal of Philosophy* 77, S. 315-338.
- Hardin, G.* (1968), The Tragedy of the Commons. *Science* 162, S. 1243-1248.
- Hardin, R.* (1971), Collective Action as an Agreeable n-Prisoners' Dilemma. *Behavioral Science* 16, S. 472-481.
- Hare, R. M.* (1981), *Moral Thinking: Its Levels, Method and Point*, Oxford.
- Harsanyi, J. C.* (1953), Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking. *Journal of Political Economy* 61, S. 434-435.
- (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy* 63, S. 309-321.
  - (1958), Ethics in Terms of Hypothetical Imperatives. *Mind* 47, S. 305-316.
  - (1975), Nonlinear Social Welfare Functions. *Theory and Decision* 6, S. 311-332.
  - (1976 [1975]), Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory, in: J. C. Harsanyi, *Essays on Ethics, Social Behavior, And Scientific Explanation*, Dordrecht, S. 37-63.
  - (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge u.a.
  - (1979), Bayesian Decision Theory, Rule Utilitarianism, and Arrow's Impossibility Result. *Theory and Decision* 11, S. 289-317.
  - (1982 [1977]), Morality and the Theory of Rational Behaviour, in: A. Sen/B. Williams (Hrsg.), *Utilitarianism and Beyond*, Cambridge u.a., S. 39-62.
  - (1994), Game and Decision Theoretic Models in Ethics, in: R. J. Aumann/S. Hart (Hrsg.), *Handbook of Game Theory*, Band 1, Amsterdam u.a., S. 669-707 (Kapitel 19).
- Harsanyi, J. C./Selten, R.* (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge.
- von Hayek, F. A.* (1967), Degrees of Explanation, in: F. A. von Hayek, *Studies in Philosophy, Politics, and Economics*, London, S. 3-21.
- Hegselmann, R.* (1988), Wozu könnte Moral gut sein? *Grazer philosophische Studien* 31, S. 1-28.



- (1989), Rationalität und Moral. *Homo Oeconomicus* 4, S. 15-40.
- Held, V.* (1977), Rationality and Reasonable Cooperation. *Social Research* 44, S. 708-744.
- Hempel, C. G.* (1977), Aspekte wissenschaftlicher Erklärung, Berlin/New York.
- Hempel, C. G./Oppenheim, P.* (1948), Studies in the Logic of Explanation. *Philosophy of Science* 15, S. 135-178.
- Hicks, J. R.* (1939), The Foundation of Welfare Economics. *Economic Journal* 49, S. 696-712.
- Hinsch, W.* (1992), Einleitung, in: J. Rawls, Die Idee des politischen Liberalismus. Aufsätze 1978-1989, hrsg. von W. Hinsch, Frankfurt am Main, S. 9-44.
- (1997a), Gerechtfertigte Ungleichheiten. Eine Verteidigung des Differenzprinzips, unveröffentlichte Habilitationsschrift, Philosophisches Seminar der Westfälischen Wilhelms-Universität, Münster.
- (1997b), Die Idee der öffentlichen Rechtfertigung und die Fiktion des Urzustandes, in: Philosophische Gesellschaft Bad Homburg/W. Hinsch (Hrsg.), Zur Idee des politischen Liberalismus. John Rawls in der Diskussion, Frankfurt am Main, S. 67-115.
- Hobbes, T.* (1966 [1647]), Vom Menschen. Vom Bürger, 2. Auflage, Hamburg.
- (1976 [1651]), Leviathan oder Stoff, Form und Gewalt eines bürgerlichen und kirchlichen Staates, Frankfurt am Main.
- Hochman, H. M./Rodgers, J. D.* (1969), Pareto Optimal Redistribution. *American Economic Review* 59, S. 542-557.
- Höffe, O.* (1981), Sittlich-politische Diskurse, Frankfurt am Main.
- (1984), Is Rawls' Theory of Justice really Kantian?, *Ratio* 26, S. 103-124.
- Hoerster, N.* (1977), Utilitaristische Ethik und Verallgemeinerung, 2. Auflage, München.
- (1979), John Rawls' Kohärenztheorie der Normenbegründung, in: O. Höffe (Hrsg.), Über John Rawls' Theorie der Gerechtigkeit, Frankfurt am Main, S. 57-76.
- Homann, K.* (1988), Rationalität und Demokratie, Tübingen.
- Howe, R. E./Roemer, J. E.* (1981), Rawlsian Justice as the Core of a Game. *American Economic Review* 71, S. 880-895.
- Hume, D.* (1993), Die wertlose Fiktion vom Gesellschaftsvertrag, in: N. Hoerster (Hrsg.), Klassische Texte der Staatsphilosophie, 8. Auflage, München, S. 163-176.

- Jakob, T.* (1996), Die Ethik John C. Harsanyis, Dissertation, Fachbereich Philosophie der Gerhard-Mercator-Universität GH, Duisburg 1996.
- de Jasay, A.* (1995 [1991]), *Liberalismus neugefaßt - für eine entpolitisierte Gesellschaft*, Berlin/Frankfurt am Main.
- Johansson, P.-O.* (1991), *An Introduction to Modern Welfare Economics*, Cambridge u.a.
- Kagel, J./Kim, C./Moser, D.* (1995), Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs. *Games and Economic Behavior* 13, S. 100-110.
- Kaldor, N.* (1939), Welfare propositions of economics and interpersonal comparisons of utility, *Economic Journal* 49, S. 549-552.
- Kant, I.* (1983 [1785]), *Grundlegung der Metaphysik der Sitten*, in: I. Kant, Werke in zehn Bänden, Band 6, hrsg. von W. Weischedel, Sonderausgabe 1983 der Ausgabe Darmstadt 1956, Darmstadt, S. 7-102.
- (1983 [1793]), *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis*, in: I. Kant, Werke in zehn Bänden, Band 9, hrsg. von W. Weischedel, Sonderausgabe 1983 der Ausgabe Darmstadt 1956, Darmstadt, S. 125-172.
  - (1983 [1797]), *Die Metaphysik der Sitten*, in: I. Kant, Werke in zehn Bänden, Band 7, hrsg. von W. Weischedel, Sonderausgabe 1983 der Ausgabe Darmstadt 1956, Darmstadt.
- Kavka, G. S.* (1986), *Hobbesian Moral and Political Theory*, Princeton.
- Kaye, D. H.* (1980), Playing Games with Justice: Rawls and the Maximin Rule. *Social Theory and Practice* 6, S. 33-51.
- Kersting, W.* (1993), *John Rawls zur Einführung*, Hamburg.
- (1994), *Die politische Philosophie des Gesellschaftsvertrags*, Darmstadt.
- Kirman, A. P.* (1992), Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives* 6, S. 117-136.
- Kirsch, G.* (1981), Ordnungspolitik als Gegenstand der politischen Auseinandersetzung, in: O. Issing (Hrsg.), *Zukunftsprobleme der sozialen Marktwirtschaft*, Berlin, S. 255-275.
- Kley, R.* (1989), *Vertragstheorien der Gerechtigkeit. Eine philosophische Kritik der Theorien von John Rawls, Robert Nozick und James Buchanan*, Bern/Stuttgart.
- Kliemt, H.* (1977), *Untersuchungen über die Begründbarkeit staatsphilosophischer Legitimitätskriterien*, Dissertation im Fachbereich Philosophie der Johann Wolfgang Goethe-Universität, Frankfurt am Main.

- (1986), Individualism, Libertarianism and Non-Cognitivism. *Analyse & Kritik* 8, S. 211-228.
- (1988), Thomas Hobbes, David Hume und die Bedingungen der Möglichkeit eines Staates, in: O. Weinberger/P. Koller/A. Schramm (Hrsg.), *Philosophie des Rechts, der Politik und der Gesellschaft. Akten des 12. Internationalen Wittgenstein Symposiums 1987*, Wien, S. 152-160.
- (1993), Constitutional Commitment. On the Economic and Legal Philosophy of Rules. *Jahrbuch für Neue Politische Ökonomie* 12, S. 145-173.
- (1998), Distributive Justice, in: P. Newman (Hrsg.), *The New Palgrave Dictionary of Economics and the Law*, Band 1, London/Basingstoke, S. 630-635.
- Kliemt, H./Lahno, B. (1996), Modelle der Moralwissenschaften, Teil 2, unveröffentlichtes Manuskript.*
- Kliemt, H./Schauenberg, B. (1982), Zu M. Taylors Analysen des Gefangenendilemmas. Analyse & Kritik* 4, S. 71-96.
- Koller, P. (1983), Rawls' Differenzprinzip und seine Deutungen. Erkenntnis* 20, S. 1-25.
- (1987), *Neue Theorien des Sozialvertrags*, Berlin.
- Krelle, W. (1968), Präferenz- und Entscheidungstheorie, unter Mitarbeit von D. Coenen, Tübingen.*
- Krueger, A. O. (1974), The Political Economy of the Rent-Seeking Society. American Economic Review* 64, S. 291-303.
- Küttner, M. (1976), Ein verbesserter deduktiv-nomologischer Erklärungs begriff. Zeitschrift für allgemeine Wissenschaftstheorie* 7, S. 274-297.
- Kuhn, S. T. (1996), Agreement Keeping and Indirect Moral Theory. Journal of Philosophy* 93, S. 105-128.
- (1997), Prisoner's Dilemma, in: *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/spr1998/entries/prisoner-dilemma/>.
- Kuhn, S. T./Moresi, S. (1995), Pure and Utilitarian Prisoner's Dilemmas. Economics and Philosophy* 11, S. 333-343.
- Kukathas, C./Pettit, P. (1990), A Theory of Justice and its Critics*, Cambridge.
- Kulenkampff, A. (1979), Methodenfragen der Gerechtigkeitstheorie. Analyse & Kritik* 1, S. 90-104.
- Kuran, T. (1995), Private Truths, Public Lies. The Social Consequences of Preference Falsification*, Cambridge.

- Kyriacou, A.* (1998), A Comment on Müller's 'Unveiling of the Veil of Uncertainty'. *Constitutional Political Economy* 9, S. 335-338.
- Lahno, B.* (1995), *Versprechen. Überlegungen zu einer künstlichen Tugend*, München/Wien.
- Leschke, M.* (1993), *Ökonomische Verfassungstheorie und Demokratie. Das Forschungsprogramm der Constitutional Economics und seine Anwendung auf die Grundordnung der Bundesrepublik Deutschland*, Berlin.
- Lessnoff, M.* (1986), *Social Contract*, Houndsmill u.a.
- Lewis, D.* (1975 [1969]), *Konventionen. Eine sprachphilosophische Abhandlung*, Berlin/New York.
- Lindahl, E.* (1919), *Die Gerechtigkeit der Besteuerung. Eine Analyse der Steuerprinzipien auf Grundlage der Grenznutzentheorie*, Lund.
- Little, I. M. D.* (1952), *Social Choice and Individual Values*. *Journal of Political Economy* 60, S. 422-432.
- Locke, J.* (1992 [1689]), *Zwei Abhandlungen über die Regierung*, herausgegeben und eingeleitet von Walter Euchner, 5. Auflage, Frankfurt am Main.
- Lottenbach, H.* (1988), Eine (partielle) Rehabilitation Hammurabis. *Grazer philosophische Studien* 31, S. 29-44.
- Luce, R. D./Raiffa, H.* (1957), *Games and Decisions. Introduction and Critical Survey*, New York u.a.
- Lutz, D. S.* (1988), *The Origins of American Constitutionalism*, Baton Rouge/London.
- Machina, M. J.* (1987), Choice Under Uncertainty: Problems Solved and Unsolved. *Journal of Economic Perspectives* 1, S. 121-154.
- Mackie, J. L.* (1981 [1977]), *Ethik. Die Erfindung des moralisch Richtigen und Falschen*, Stuttgart.
- McLean, I.* (1981), The Social Contract in Leviathan and the Prisoner's Dilemma Supergame. *Political Studies* 29, S. 339-351.
- (1985), Egoistische Gene und der Stamm von Ik: vom Hobbesschen Naturzustand zum Gefangenendilemmasuperspiel, in: K.-P. Markl (Hrsg.), *Analytische Politikphilosophie und ökonomische Rationalität I*, Opladen, S. 30-53.
- Metha, J./Starmer, C./Sugden, R.* (1994a), Focal Points in Pure Coordination Games: An Experimental Investigation. *Theory and Decision* 36, S. 163-185.
- (1994b), The Nature of Salience: An Experimental Investigation of Pure Coordination Games. *American Economic Review* 84, S. 658-673.

- Milnor, J.* (1964 [1951]), Spiele gegen die Natur, in: M. Shubik (Hrsg.), Welt im Werden. Spieltheorie und Sozialwissenschaften, Hamburg, S. 129-139.
- Müller, C.* (1997), Der 'Schleier des Nichtwissens' und der 'Schleier der Unsicherheit'. Wirtschaftswissenschaftliches Studium (WiSt) 26, S. 245-247.
- (1998a), The Veil of Uncertainty Unveiled. Constitutional Political Economy 9, S. 5-17.
  - (1998b), Risk Aversion in Constitutional Choice. Reply to Kyriacou. Constitutional Political Economy 9, S. 339-343.
  - (1999), Brauchen wir einen Schleier der Unkenntnis? ORDO. Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft 50, S. 207-232.
- Müller, C./Tietzel, M.* (1998), Allmende-Allokationen, in: M. Tietzel (Hrsg.), Ökonomische Theorie der Rationierung, München, S. 163-201.
- (1999), Property Rights and their Partitioning, erscheint in: J. Backhaus (Hrsg.), Elgar Companion to Law and Economics, Band 1, Cheltenham.
- Mueller, D. C.* (1973), Constitutional Democracy and Social Welfare. Quarterly Journal of Economics 87, S. 60-80.
- (1989a), Public Choice II. A Revised Edition of Public Choice. Cambridge.
  - (1989b), Individualism, Contractarianism, and Morality. Social Justice Research 3,, S. 1-19.
  - (1996), Constitutional Democracy, New York/Oxford.
  - (1998a), The Role of Uncertainty in an Economic Theory of Constitutions, mimeo.
  - (1998b), Buchanan, James McGill, in: P. Newman (Hrsg.), The New Palgrave Dictionary of Economics and the Law, Band 1, London/Basingstoke, S. 179-185.
- Mueller, D. C./Tollison, R. D./Willet, T. D.* (1974), The Utilitarian Contract: A Generalization of Rawls' Theory of Justice. Theory and Decision 4, S. 345-367.
- Musgrave, R. A.* (1974), Maximin, Uncertainty, and the Leisure Trade-Off. Quarterly Journal of Economics 88, S. 624-632.
- Myerson, R. B.* (1991), Game Theory. Analysis of Conflict, Cambridge/London.
- Nagel, T.* (1975 [1973]), Rawls on Justice, in: N. Daniels (Hrsg.), Reading Rawls. Critical Studies on Rawls' 'A Theory of Justice', Oxford, S. 1-16.
- Neumärker, K. J. B.* (1995), Finanzverfassung und Staatsgewalt in der Demokratie, Frankfurt am Main u.a.

- Neyman, A.* (1985), Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma. *Economic Letters* 19, S. 227-229.
- Ng, Y.-K.* (1984), Expected Subjective Utility: Is the Neumann-Morgenstern Utility the Same as the Neoclassical's? *Social Choice and Welfare* 1, S. 177-186.
- Nozick, R.* (1976 [1974]), *Anarchie, Staat, Utopia*, München.
- Nutzinger, H.G.* (1975), Vorwort des Herausgebers, in: A. Sen, *Ökonomische Ungleichheit*, Frankfurt am Main/New York, S. 7-9.
- Olson* (1968 [1965]), *Die Logik des kollektiven Handelns. Kollektivgüter und die Theorie der Gruppen*, Tübingen.
- Ophuls, W.* (1973), Leviathan or Oblivion?, in: H. E. Daly (Hrsg.), *Toward a Steady State Economy*, San Francisco, S. 215-230.
- Pattanaik, P. K.* (1968), Risk, Impersonality, and the Social Welfare Function. *Journal of Political Economy* 76, S. 1152-1169.
- Petersen, T.* (1996), *Individuelle Freiheit und allgemeiner Wille. Buchanans politische Ökonomie und die politische Philosophie*, Tübingen.
- Pettit, P.* (1974), A Theory of Justice. *Theory and Decision* 4, S. 311-324.
- Phelps, E.S.* (1976), Social Policy and Uncertain Careers: Beyond Rawls' Paradigm Case, in: R. E. Grieson (Hrsg.), *Public and Urban Economics. Essays in Honor of William S. Vickrey*, Lexington u.a., S. 159-178.
- Pies, I.* (1994), Zum Verhältnis von Spieltheorie und konstitutioneller Ökonomik. *Homo Oeconomicus* 11, S. 47-69.
- Pies, I./Leschke, M.* (Hrsg.; 1996), *James Buchanans konstitutionelle Ökonomik*, Tübingen.
- Pogge, T. W.* (1994), *John Rawls*, München.
- Popper, K.* (1987), *Das Elend des Historizismus*, 6. Auflage, Tübingen.
- (1994), *Logik der Forschung*, 10. Auflage, Tübingen.
- Rae, D.* (1975), Maximin Justice and an Alternative Principle of General Advantage. *American Political Science Review* 69, S. 630-647.
- Rapoport, A.* (1987), Prisoner's Dilemma, in: J. Eatwell u.a. (Hrsg.), *The New Palgrave*, Band 3, London u.a., S. 973-976.
- Rapoport, A./Chammah, A.* (1965), *Prisoner's Dilemma. A Study in Conflict and Cooperation*, Ann Arbor.
- Rawls, J.* (1958), Justice as Fairness. *The Journal of Philosophy* 54, S. 653-662.

- (1967), Distributive Justice, in: P. Laslett, W.G. Runciman (Hrsg.), *Philosophy, Politics, and Society*, Third Series, London, S. 58-82.
  - (1974), Some Reasons for the Maximin Criterion. *American Economic Review* 64 (Papers & Proceedings), S. 141-146.
  - (1982), Social Unity and Primary Goods, in: A. Sen/B. Williams (Hrsg.), *Utilitarianism and beyond*, Cambridge u.a., S. 159-185.
  - (1992 [1977]), Die Grundstruktur als Gegenstand, in: J. Rawls, *Die Idee des politischen Liberalismus. Aufsätze 1978-1989*, hrsg. von W. Hirsch, Frankfurt am Main, S. 45-79.
  - (1992 [1980]), Kantischer Konstruktivismus in der Moraltheorie, in: J. Rawls, *Die Idee des politischen Liberalismus. Aufsätze 1978-1989*, hrsg. von W. Hirsch, Frankfurt am Main, S. 80-158.
  - (1992 [1985]), Gerechtigkeit als Fairneß: politisch und nicht metaphysisch, in: J. Rawls, *Die Idee des politischen Liberalismus. Aufsätze 1978-1989*, hrsg. von W. Hirsch, Frankfurt am Main, S. 255-292.
  - (1992 [1987]), Der Gedanke eines übergreifenden Konsenses, in: J. Rawls, *Die Idee des politischen Liberalismus. Aufsätze 1978-1989*, hrsg. von W. Hirsch, Frankfurt am Main, S. 293-332.
  - (1992 [1989]), Der Bereich des Politischen und der Gedanke eines übergreifenden Konsenses, in: J. Rawls, *Die Idee des politischen Liberalismus. Aufsätze 1978-1989*, hrsg. von W. Hirsch, Frankfurt am Main, S. 333-363.
  - (1993), *Political Liberalism*, New York.
  - (1994 [1971]), *Eine Theorie der Gerechtigkeit*, 8. Auflage, Frankfurt am Main.
- Reiner, H.* (1974), Die 'goldene Regel'. Die Bedeutung einer sittlichen Grundformel der Menschlichkeit, in: H. Reiner, *Die Grundlagen der Sittlichkeit*, 2. Auflage, Meisenheim am Glan, S. 348-379.
- Ricken, F.* (1997), Ist eine moralische Konzeption der politischen Gerechtigkeit ohne umfassende moralische Lehre möglich?, in: *Philosophische Gesellschaft Bad Homburg/W. Hirsch (Hrsg.), Zur Idee des politischen Liberalismus. John Rawls in der Diskussion*, Frankfurt am Main, S. 420-437.
- Riemer, N.* (1951), The Case for Bare Majority Rule. *Ethics* 62, S. 16-32.
- Robbins, L.* (1932), *An Essay on the Nature and Significance of Economic Science*, London.
- Rousseau, J. J.* (1977 [1755]), *Abhandlung über die Politische Ökonomie*, in: J.J. Rousseau, *Politische Schriften*, Band 1, Paderborn, S. 9-57.



- (1977 [1762]), Vom Gesellschaftsvertrag oder Grundsätze des Staatsrechts, Stuttgart.
- Runciman, W. G./Sen, A.* (1965), Games, Justice and the General Will. *Mind* 74, S. 554-562.
- Samuelson, P. A.* (1947), Foundations of Economic Analysis, Cambridge.
- (1964), A.P. Lerner at Sixty. *Review of Economic Studies* 32, S. 169-178.
- (1976), Optimal Compacts for Redistribution, in: R.E. Grieson (Hrsg.), *Public and Urban Economics*, Lexington, S. 179-190.
- (1981), Bergsonian Welfare Economics, in: S. Rosefielde (Hrsg.), *Economic Welfare and the Economics of Soviet Socialism*, Cambridge, S. 223-266.
- Scanlon, T. M.* (1975 [1973]), Rawls' Theory of Justice, in: N. Daniels (Hrsg.), *Reading Rawls. Critical Studies on Rawls' 'A Theory of Justice'*, Oxford, S. 169-205.
- Schelling, T. C.* (1960), *The Strategy of Conflict*. London u.a.
- (1973), Hockey Helmets, Concealed Weapons, and Daylight Saving. A Study of Binary Choices with Externalities. *Journal of Conflict Resolution* 17, S. 381-428.
- Schmidt, J.* (1991), *Gerechtigkeit, Wohlfahrt und Rationalität. Axiomatische und entscheidungstheoretische Fundierungen von Verteilungsproblemen*, München.
- (1995), Rational Choice Fundierungen von Gerechtigkeitsprinzipien. *Analyse & Kritik* 17, S. 167-182.
- Schmidtz, D.* (1991), *The Limits of Government. An Essay on the Public Goods Argument*, Boulder.
- Schoemaker, P. J. H.* (1982), The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature* 20, S. 529-563.
- Schumann, J.* (1994), Wohlfahrtsökonomik, in: O. Issing (Hrsg.), *Geschichte der Nationalökonomie*, 3. Auflage, München, S. 215- 237.
- Schwartz, A.* (1973), Moral Neutrality and Primary Goods. *Ethics* 83, S. 294-307.
- Schweizer, U.* (1988), Unanimous Choice of Rules, Discussion Paper No. A-172, Universität Bonn, Sonderforschungsbereich 303, Bonn.
- (1990), Calculus of Consent: A Game Theoretic Perspective. *Journal of Institutional and Theoretical Economics* 146, S. 28-54.
- Sen, A.* (1970), *Collective Choice and Social Welfare*, San Francisco.
- (1975 [1973]), *Ökonomische Ungleichheit*, Frankfurt am Main/New York.
- (1982 [1972]), *Choice, Orderings and Morality*, A. Sen, *Choice, Welfare and Measurement*, Oxford: Basil Blackwell, S. 74-83.

- (1982 [1977]), On Weights and Measures: Informational Constraints in Social Welfare Analysis, in: A. Sen, Choice, Welfare and Measurement, Oxford, S. 226-263.
- (1982 [1979]), Interpersonal Comparisons of Welfare, in: A. Sen, Choice, Welfare and Measurement, Oxford, S. 264-281.
- Simon, H. A.* (1955), A Behavioral Model of Rational Choice. *Quarterly Journal of Economics* 69, S. 99-118.
- (1993 [1983]), *Homo rationalis. Die Vernunft im menschlichen Leben*, Frankfurt am Main/New York.
- Singer, M. G.* (1963), The Golden Rule. *Philosophy* 38, S. 293-314.
- Smyth, J.* (1972), The Prisoner's Dilemma II. *Mind* 81, S. 427-431.
- Sombart, W.* (1930), *Die drei Nationalökonomien*, München/Leipzig.
- Stegmüller, W.* (1969), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band 1: Wissenschaftliche Erklärung und Begründung*, Berlin u.a.
- Stigler, G. J.* (1961), The Economics of Information. *Journal of Political Economy* 69, S. 213-225.
- Strasnick, S.* (1976), Social Choice and the Derivation of Rawls's Difference Principle. *Journal of Philosophy* 73, S. 85-99.
- Sugden, R.* (1993), Rationality and Impartiality: Is the Contractarian Enterprise Possible?, in: D. Gauthier/R. Sugden (Hrsg.), *Rationality, Justice and the Social Contract. Themes from Morals by Agreement*, New York u.a., S. 157-175.
- (1995), A Theory of Focal Points. *Economic Journal* 105, S. 533-550.
- Sumner, L. W.* (1977), Rawls and the Contract Theory of Civil Disobedience, in: K. Nielsen/R. A. Shiner (Hrsg.), *New Essays on Contract Theory (Canadian Journal of Philosophy, Supplement vol. 3)*, Guelph, S. 1-48.
- Sutter, D.* (1998), Calculation of Self-Interest and Constitutional Consensus: The Role of Ideology. *Constitutional Political Economy* 9, S. 323-333.
- Taylor, M.* (1976), *Anarchy and Cooperation*, London u.a.
- Taylor, M.* (1985), Von Hobbes zu den rationalen Kooperationsbedingungen des Gefangenendilemmasuperspiels, in: K.-P. Markl (Hrsg.), *Analytische Politikphilosophie und ökonomische Rationalität, Band 1: Vom Hobbes'schen Wissenschaftsbegriff zum liberalen Paradox*, Opladen, S. 1-29.
- (1987), *The Possibility of Cooperation*, Cambridge u.a.

- Tietzel, M.* (1981), Die Rationalitätsannahme in den Wirtschaftswissenschaften oder: Der homo oeconomicus und seine Verwandten. *Jahrbuch für Sozialwissenschaft* 32, S. 115-138.
- (1985), Wirtschaftstheorie und Unwissen. Überlegungen zur Wirtschaftstheorie jenseits von Risiko und Unsicherheit, Tübingen.
  - (1986), Idealisierte Erklärungen. *Zeitschrift für allgemeine Wissenschaftstheorie* 17, S. 315-321.
  - (1990), Virtue, Vice, and Dr. Pangloss. On the Economics of Conventions. *ORDO. Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft* 41, S. 251-268.
- Tietzel, M./Müller, C.* (1998), Noch mehr zur Meritorik. *Zeitschrift für Wirtschafts- und Sozialwissenschaften (ZWS)* 118, S. 87-127.
- de Tocqueville, A.* (1976 [1848]), Über die Demokratie in Amerika, München.
- Tollison, R. D.* (1972), Involved Social Analysis, in: J. M. Buchanan/R. D. Tollison (Hrsg.), *Theory of Public Choice*, Ann Arbor, S. 3-10.
- Ullmann-Margalit, E.* (1977), *The Emergence of Norms*, Oxford.
- Vanberg, V.* (1986), Individual Choice and Institutional Constraints: The Normative Element in Classical and Contractarian Liberalism. *Analyse & Kritik* 8, S. 113-149.
- (1990), James M. Buchanan: eine Einführung in Person und Werk, in: J. M. Buchanan, *Politische Theorie als Verfassungstheorie*, Zürich, S. 9-22.
  - (1994), Social Contract Theory, in: P. J. Boettke (Hrsg.), *The Elgar Companion to Austrian Economics*, Cheltenham, S. 337-342.
  - (1997), Protektionismus und Ordnungsevolution. Zum Problem der Anpassung von Wirtschaftssystemen an ihre Umwelt, mimeo.
- Vanberg, V./Buchanan, J. M.* (1988), Rational Choice and Moral Order. *Analyse & Kritik* 10, S. 138-160.
- (1989), Interests and Theories in Constitutional Choice. *Journal of Theoretical Politics* 1, S. 49-62.
- Varian, H.* (1994 [1992]), *Mikroökonomie*, 3. Auflage, München.
- Vickrey, W.* (1945), Measuring Marginal Utility by Reactions to Risk. *Econometrica* 13, S. 319-333.
- (1960), Utility, Strategy, and Social Decision Rules. *Quarterly Journal of Economics* 74, S. 507-535.
- Wagner, G.* (1998), Zentrale Aufgaben beim Um- und Ausbau der Gefahrenvorsorge. Ein Versuch, die Vertragstheorie sowie die Theorie des Staatsversagens für die Sozi-

- alpolitik nutzbar zu machen, in: J. Genesko u.a. (Hrsg.), Reform des Sozialstaats II. Theoretische, institutionelle und empirische Aspekte, Berlin, S. 11-51.
- Wagner*, R. (1988), The Calculus of Consent: A Wicksellian Retrospective. Public Choice 56, S. 153-166.
- Wald*, A. (1945), Statistical Decision Functions which Minimize the Maximum Risk, Annals of Mathematics 46, S. 265-280.
- Weber*, M. (1996), Ökonomie der Arbeitsverfassung. Die Organisation von Arbeitsbeziehungen in Unternehmen und Volkswirtschaften; mit vier Fallstudien, Hamburg.
- Weisser*, G. (1971 [1954]), Grundsätze der Verteilungspolitik, in: B. Külp/W. Schreiber (Hrsg.), Soziale Sicherheit, Köln/Berlin, S. 110-135.
- Wessels*, J. H. (1993), Redistribution from a Constitutional Perspective. Constitutional Political Economy 4, S. 425-448.
- (1994), Verfassungsregeln als glaubwürdige Verpflichtung. Homo Oeconomicus 11, S. 21-45.
- Wicksell*, K. (1969 [1896]), Finanztheoretische Untersuchungen nebst Darstellung und Kritik des Steuerwesens Schwedens, Aalen (Neudruck der Ausgabe Jena 1896).
- Wildt*, A. (1996), Gerechtigkeit, Gleichheit und Optimierung für jeden. Zur Begründung von Rawls' Differenzprinzip, in: K. Bayertz (Hrsg.), Politik und Ethik, Stuttgart, S. 249-276.
- Wimmer*, R. (1980), Universalisierung in der Ethik. Analyse, Kritik und Rekonstruktion ethischer Rationalitätsansprüche, Frankfurt am Main.
- Wolff*, R. P. (1976), On Strasnick's 'Derivation' of Rawls's 'Difference Principle'. Journal of Philosophy 73, S. 849-858.
- (1977), Understanding Rawls. A Reconstruction and Critique of A Theory of Justice, Princeton.
- Zintl*, R. (1983), Individualistische Theorien und die Ordnung der Gesellschaft, Berlin.
- (1986), Vertrag ohne Voraussetzungen: James M. Buchanan, in: L. Kern/H.-P. Müller (Hrsg.), Gerechtigkeit, Diskurs oder Markt? Die neuen Ansätze in der Vertragstheorie, Opladen, S. 111-125.

## Sachwortregister

Adäquatheitsbedingungen s. Erklärung

Allmendeproblem 121, 124

Anarchie, Hobbessche 120

Auswahlregel

- kollektive (soziale) 30
- konstitutionelle soziale 35 f.

Auswahlregel, kollektive 20

Auszahlungsdominanz 140

Autonomie

- Kants Begriff der 65
- rationale 62 f.
- vollständige 63

Chancenprinzip

s. Gerechtigkeitsgrundsätze

Collective choice rule (s. auch

Auswahlregel) 20, 83, 185

Differenzprinzip

- und Weissers Distributionsprinzip 91
- und Utilitarismus 94 f.

Dilemma des „paretianischen“ Ökonomen  
70

- und konstitutioneller Ansatz 70 f.
- und Wohlfahrtsökonomik 70 f.

Dominanzprinzip 129, 131

Einigung

- Bedingungen einer fairen 161 ff.
- stillschweigende 165

Einstimmigkeit 26

- als Entscheidungsregel 17
- hypothetische 32 ff.
- stillschweigende 103
- und inhaltliche Gerechtigkeitsnormen  
69 f.

- und Repräsentationsaspekt 28 ff.

- und Schutzaspekt 29 ff.

Einstimmigkeitsspiel 135

Ergebnisoffenheit 52

Erklärung(en)

- Als-ob- 38, 44
- approximative 46
- deduktiv-nomologische 37
- empirische Adäquatheitsbedingung 20,  
43, 186
- idealisierte 44
- logische Adäquatheitsbedingungen 20,  
42, 46, 186
- logische Struktur der  
vertragstheoretischen Als-ob- 38
- partielle 57, 186
- Rationalwahl- (Rational-choice-) 38 ff.

Fairneß 31

Fokuspunkt 166

Folk-Theorem 179

Gefangenendilemma 23, 120 ff.

- *n*-Personen- 130
- Regeln im 133
- Superspiel 179
- Zwei-Personen- 128
- und Schleier der Unkenntnis 125 ff.
- und Verallgemeinerungsprinzip 134

Gerechtigkeit 15, 31 f.

- konsensuale 15
- Laplace- 96-102
- Maximin- 87-96
- zwischen den Generationen 108 f.

Gerechtigkeitsauffassung (-theorie)

- Buchanans Theorie als inhaltliche 70, 82
  - inhaltliche 70
  - paretianische 49 f.
  - schwach individualistische 50
  - strikt individualistische 49 f.
- Gerechtigkeitsgrundsätze 19, 88-96
- Chancenprinzip 19, 89 f.
  - Differenzprinzip 19, 89 ff.
  - Prinzip gleicher Grundrechte und  
-freiheiten 19, 89 f.

Gerechtigkeitskonzeption 61, 86-102  
 - politische 62, 84  
 - und allgemeine moralische Konzeption 62

Gerechtigkeitsnorm, inhaltliche (externe)  
 66, 75

- und Einstimmigkeit 69 f.

Gesellschaftsvertrag 17, 28, 33

- diskriminierender 108

- hypothetischer 17

- Stabilität des 168

- Theorien des 15

Gleichheit 40

Gleichwahrscheinlichkeitsannahme 148

Gleichwahrscheinlichkeitsmodell

moralischer Werturteile 99

- als Spezialfall von Buchanan's Theorie 149

Goldene Regel 105, 134

- und der Schleier der Unsicherheit 105

- unsinnige Konsequenzen der 110

Grundgüter 61 f.

- Messung und interpersoneller Vergleich von 112-118

- und Weissers Konzept der Lebenslagen 91

Grundsatz gleicher Grundfreiheiten

s. Gerechtigkeitsgrundsätze

Grundstruktur 16, 61

**Harsanyi-Doktrin** 114

Hobbesscher Dschungel 122

Hypothetischer Imperativ 183

**Individualismus, normativer** 25

Induktionsschluß 47, 186

Informationsgehalt 43 ff.

Interdependenzkosten 73

**Jedermann-Regel** 29, 7 f.

**kategorischer Imperativ** 65, 134, 184

Kompensation 48

Kompensationskriterium (Kaldor und Hicks) 189 f.

Konsens, übergreifender 61

Konsensfindungskosten 32, 72 f.

konstitutionelle Sichtweise 16

- und Grundstruktur 16

Konstitutionelles Spiel 127, 131 ff.

- unter Buchanan's Schleier der Unsicherheit 147 ff.

- unter Rawls' Schleier des Nichtwissens 136 ff.

Konsumentensouveränität 26

konvexe Hülle 175

Koordinationsgleichgewicht 138

Kosten, externe 29 f., 72

Kriterien, externe 27

**Laplace-Prinzip** 196 ff.

Legitimationstheorien

- Dilemma individualistischer 69

Legitimität 32

- mutmaßliche 48

**Maximin-Lösung der sozialen**

Gerechtigkeit 60, 84

**Maximin-Prinzip (-Kriterium)**

- der Entscheidungstheorie 88

- und Entscheidung im Urzustand 92-96

- und Utilitarismus 94 f.

**Mehrheit**

- (Abstimmungs-)Regel der einfachen 71-73

- Tyrannei der 31, 143

**Mehrheitsdominanz** 143

**Nash-Gleichgewicht** 129

Norm(en), externe 20, 70

Nutzenvergleich

- intersubjektiver 106

- intrapersoneller 106

**Ordnung, strikte partielle** 54f., 83

**Ordnungspolitik** 123, 181, 182

**Pareto-Dominanz** 140, 163

**Pareto-Effizienz** 30

**Pareto-Kriterium (-Prinzip)** 19, 30

- und Ergebnisoffenheit 58

**Pareto-Optima**

- Wahl zwischen 55, 69, 183

**Pareto-Relation, starke** 30, 54 f.

**Person, moralische** 62, 84

**Pluralismus, Faktum des** 17, 61, 189

**postkonstitutionelle Sichtweise** 15 f.

**Postkonstitutionelles Spiel** 127 ff.

Präferenzfalsifikation 78, 169  
Prognose s. Erklärung

Rationale, das 63  
Rationalitätsbegriff (-hypothese) 41 ff., 62 f.  
Rationalverhaltenstheorie 38  
Realitätsnähe 45  
Regeln 16, 27 f.  
- und Allgemeinheit 76 ff.  
- und Quasi-Permanenz 76 f.  
Relation, starke konstitutionelle 36  
Rent seeking 123, 125  
Risiko 86, 99

Schleier der Unkenntnis 18, 58  
- und Gefangenendilemma 125 ff.  
Schleier der Unsicherheit 18, 59, 71, 74-82  
- als empirisches Phänomen 75-82  
- Diskriminierung unter dem 153 ff.  
- kantische Interpretation des 81 f.  
- Nichteinigung unter dem 155 ff.  
- und Allwissenheitsannahme der Wohlfahrtsökonomik 105  
- und Begründungsthese 76 ff.  
- und Gerechtigkeit zwischen den Generationen 108 f.  
- und Gleichwahrscheinlichkeitsannahme 99  
- und Schleier des Nichtwissens 76 f., 106, 109  
- und Utilitarismus 99 f.  
- und Wirkungsthese 76 f., 81  
- und Zeitablaufvariante 98 f.  
Schleier des Nichtwissens 18, 58, 63-69  
- als Instrument der Normdiskussion 69  
- Diskriminierung unter dem 140 ff.  
- Faineßfunktion 64  
- Konsensfunktion 65  
- und kategorischer Imperativ 65  
- und Schleier der Unsicherheit 76 f., 106, 109  
- und Sklavereivertrag 68  
- und Universalisierung 65-69  
- und Vetorecht 68  
Selbstbindungsfähigkeit (-macht) 16, 27 f., 181, 189  
Sichere Erwartungen 171 ff.  
- und Fairneß 172 ff.

Situationsannahmen 38 ff.  
Sklavereivertrag 52, 68, 79

Tauschgerechtigkeit 32  
Teilspielkonsistenz 179  
Teilspielperfektheit 179  
Theoriekonflikt 181

Überlegungsgleichgewicht 19  
Umverteilung, paretooptimale 18, 190  
Universalisierung  
- partielle 104 ff., 187  
- vollkommene 110 ff.  
Universalisierungsgrad 103 ff.  
Universalisierungsnorm 59, 65 ff.  
- als inhaltliche Gerechtigkeitsnorm 66  
- in Buchanans Theorie 82  
- und Allgemeinheit 80  
- und Fairneß 80  
- und Reziprozität 80  
- und Schleier des Nichtwissens 65-69  
- und Sklavereivertrag 68  
Unsicherheit 86  
- i.e.S. 86  
- und Risiko 86  
Utilitarismus 19  
- Harsanyis reformulierter Nutzensummen- 99 f., 113 f.  
- klassischer 87, 97, 113  
- und Differenz- (Maximin-) Prinzip 94 f.  
- und Wohlfahrtsökonomik 97  
- zentrale Probleme des 111 f.

Verfahrensgerechtigkeit 32  
- Extremposition der 50  
Verfassung, 28  
Vernünftige, das 63  
Verteilungsproblem 124  
Vertrag, fiktiver 47  
Vertragstheorie  
- Bargaining versions der 103  
- Choice versions der 103  
- schwache (unechte) 21, 58 f.  
- starke (echte) 20, 49-57  
Veto(recht)  
- Entzug des 68, 190  
- und Einstimmigkeitsregel 29, 31  
- und Markttausch 69  
- und Schleier des Nichtwissens 68



**Wertfreiheit** 51, 75

**Werturteile**

- ergebnisbezogene 66
- externe 50, 187

**Wirtschaftstheorie, neoklassische** 41, 63

**Wohlfahrtsfunktion, soziale** 22

- Benthamsche, auf der Basis des von-Neumann-Morgenstern-Nutzenkonzepts 99 f.
- Bergsonsche 66 f., 98, 187
- Laplace- 85
- Maximin- 92
- und Kollektivismus 85

**Wohlfahrtsökonomik** 19, 53 ff., 85, 126

- Erstes Theorem der 53
- paretianische 21, 53, 67
- und Allwissenheitsannahme 105
- und klassischer Utilitarismus 97
- Zweites Theorem der 67

**Wohlstandsniveaus**

- interpersoneller Vergleich 113 ff.
- intrapersoneller Vergleich 114 f.
- und Niveauvergleichbarkeit 112

**Zustände (Situationen), gesellschaftliche** 28

**Zustimmungsfähigkeit** 17, 35