

Normative Urteile in der Ökonomie

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2006

vorgelegt von
Andreas Gösele

Referent:	Prof. Dr. Ekkehart Schlicht
Korreferent:	Prof. Ray Rees
Promotionsabschlußberatung:	7. Februar 2007

Inhaltsverzeichnis

1	Einleitung	5
2	Grenzen der Wohlfahrtsökonomie	31
2.1	Einleitung	31
2.2	Das Paretoprinzip	32
2.2.1	Definitionen	32
2.2.2	Die Irrelevanz des Paretoprinzips	37
2.3	Paretianische Wohlfahrtstheorie	39
2.3.1	Die Irrelevanz der Paretoeffizienz	39
2.3.2	Über Pareto hinaus?	43
2.3.3	Samuelson als Retter?	48
2.3.4	Wohlfahrtstheoreme	53
2.4	Kompensationskriterien	61
2.4.1	Das Kaldor-Hicks-Kriterium	61
2.4.2	Das Scitovsky-Kriterium	71
2.4.3	Nochmals Samuelson	73
2.5	Unproblematisch?	80
3	Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?	95
3.1	Paretoeffizienz und Gerechtigkeit: Zur Exposition der Fragestellung.	95
3.2	Was ist ein Trade-off?	97
3.3	Realeinkommen als Maßstab für Effizienz.	98
3.4	Wohlfahrtsverluste: Effizienz und Abstand zur Paretogrenze?	104
3.4.1	Ineffiziente Transfers	104
3.4.2	Effizienz: Abstand formal	106
3.5	Das Sezieren einer Wohlfahrtsfunktion: Eine Alternative?	108
4	Kritische Anmerkungen zum ethischen Nonkognitivismus	115
5	Eine Theorie der Gerechtigkeit	131
5.1	Der Gegenstand der „Theorie der Gerechtigkeit“	131
5.2	Die Methode	132
5.3	Gerechtigkeit als Fairneß	134
5.4	Die zwei Grundsätze der Gerechtigkeit	138
5.5	Von den Grundsätzen zu den Institutionen der Grundstruktur	143

6	Der Urzustand: Ein Rekonstruktion	149
6.1	Das Problem	149
6.2	Entscheiden ohne Reue	154
6.3	Die wissende Partei: eine Vorübung	159
6.4	Die unwissende Partei	162
6.4.1	Das Grundmodell	162
6.4.2	Abweichungen von Rawls	179
6.4.3	Und der Utilitarismus?	187
7	Das RICE-Modell	199
7.1	Einleitung	199
7.2	Effizienz und Utilitarismus	199
7.3	Eine rawlsianische Perspektive	209
7.3.1	Ein Ein-Gut-Modell	209
7.3.2	Die gerechte Sparrate	211
7.3.3	Das Differenzprinzip	213
7.3.4	Ein rawlsianischer Utilitarismus	217
7.3.5	Das Völkerrecht	218
7.3.6	Umverteilung	226
7.3.7	In the long run	230
7.3.8	Unsicherheit, Katastrophen und Tote	232
	Literaturverzeichnis	251

1 Einleitung

Diese Arbeit besteht aus zwei sehr unterschiedlichen Hauptteilen.

Das Ziel des ersten Teils (bestehend aus den Kapiteln 2 und 3) ist im wesentlichen ein *destruktiver*: Es soll gezeigt werden, daß die paretianische Wohlfahrtstheorie für alle praktischen Zwecke so wenig zur Beantwortung wirtschaftspolitischer normativer Fragen beiträgt, daß es, soweit es um substantielle normative Fragen geht, kein Verlust wäre, auf sie vollständig zu verzichten. Substantielle Antworten auf wirtschaftspolitische normative Fragen sind nur möglich, wenn man entweder über die paretianische Wohlfahrtstheorie – sie ergänzend – weit hinausgeht (normativ interessant ist dann die Auseinandersetzung mit dieser weitergehenden normativen Position) oder sie überhaupt durch etwas ganz anderes ersetzt (normativ interessant ist dann die gewählte Alternative).

Unter paretianischer Wohlfahrtstheorie verstehe ich dabei *im Kern* alles das, was sich aus dem ableiten läßt, was ich im folgenden das Paretoprinzip nennen werde: Die normative These, daß gesellschaftliche (oder jedenfalls wirtschaftspolitische) Entscheidungen so zu treffen sind, daß sie einen beliebigen Zustand verwirklichen, der gemäß einer Ordnung oder Teilordnung, die paretoinklusiv ist¹, zu den besten Zuständen gehört.

Unter normativen Fragen verstehe ich Fragen danach, was wir tun sollen. Normative Aussagen sind Antworten auf normative Fragen, und müssen deshalb im hier zugrundegelegten Verständnis immer eine (wenigstens theoretisch) klare Handlungsanweisung enthalten. Die bloße Verwendung einer wertgeladenen Terminologie („besser“, „effizient“, „optimal“ o. ä.) und die bloße Angabe einer (eventuell in wertgeladener Terminologie ausgedrückten) Ordnung konstituieren deshalb noch keine normative Aussagen.

Die destruktive These ergibt sich dann direkt aus einer einfachen Feststellung: Fälle, in denen sich aus dem Paretoprinzip eine Menge der besten Elemente herleiten läßt, kommen im Bereich der Wirtschaftspolitik praktisch nicht vor², klare Handlungsanweisungen sind also aus der paretianischen Wohlfahrtstheorie nicht ableitbar. Mit anderen Worten, wenn wir ohne paretianische Wohlfahrtstheorie nicht wußten, was wir tun oder was wir als Ökonomen empfehlen sollten, wissen wir es für alle praktischen

¹ Die also einen Zustand g , der von allen betroffenen Individuen als zumindest gleich gut wie ein zweiter Zustand g' bewertet wird und von mindestens einem betroffenen Individuum als besser bewertet wird, ebenfalls als besser einordnet. Der Zustand g ist dann gegenüber dem Zustand g' „paretobesser“ oder „paretoüberlegen“.

² Dies ist eine empirische Behauptung, die ich im folgenden durchwegs als Tatsache voraussetze, da ich davon ausgehe, daß sie nicht bestritten werden wird. In jedem Fall gibt es keine *interessanten* wirtschaftspolitischen Fragen, da sich die wirtschaftspolitische Diskussion gerade aus dem Widerstreit der Meinungen *und* Interessen ergibt. Ähnlich Zajac [1995, 77].

1 Einleitung

Fragen der Wirtschaftspolitik mit den der paretianischen Wohlfahrtstheorie genauso wenig.

Diese These ist nicht wirklich neu. Sie entspricht einer Position, die etwa schon Paul Samuelson in der Diskussion um die sogenannte „Neue Wohlfahrtstheorie“³ konsistent vertreten hat: Ohne *ethische* Ergänzung des Paretoprinzips (für Samuelson durch eine Bergson-Samuelson-Wohlfahrtsfunktion⁴) ist auch die „Neue Wohlfahrtsökonomie“ unzureichend, wenn es darum geht, tatsächliche politische Entscheidungen („actual policy decisions“) zu treffen. [Samuelson 1950, 11 f.; vgl. Samuelson 1962 u. Chipman u. Moore 1978, 578-581] Im Zusammenhang eines Hinweises auf die Notwendigkeit einer Bergson-Samuelson-Wohlfahrtsfunktion im Kontext von wirtschaftspolitischen Entscheidungen den Freihandel betreffend kann Samuelson deshalb von sich sagen: „I have always pointed out the illogic of those new welfare economists who used to try to reach normative conclusions on the basis of insufficient norms.“ [Samuelson 1962; nachgedruckt in Bhagwati 1969, 177]

Aus dem bisher Gesagten – wenn es denn richtig ist – ergibt sich eine einfache Konsequenz: Intellektuelle Redlichkeit verlangt von Ökonomen, entweder darauf zu verzichten, wirtschaftspolitische Ratschläge und Empfehlungen zu geben, und sich konsequent auf die „positive“ Ökonomik zu beschränken, oder die über das Paretoprinzip hinausgehenden oder von ihm überhaupt unabhängigen Werturteile anzugeben, auf die sie ihre Empfehlungen stützen. In den Worten von Chipman u. Moore [1978, 381], mit denen sie ihre sorgfältige Untersuchung der Neuen Wohlfahrtstheorie abschließen: „We cannot make policy recommendations except on the basis of value judgments, and these value judgments should be made explicit.“

Leider scheint die Profession dieser Empfehlung nicht allgemein gefolgt zu sein. Jedenfalls in dem (natürlich sehr beschränkten) Teil der ökonomischen Literatur, den ich überblicke, finden sich nur wenige Ökonomen, die eine solche über das Paretoprinzip hinausgehende, konsistente und für ihre Empfehlungen ausreichende normative Basis explizit machen.⁵

Weitverbreitet ist eine pragmatische Position, die sich auf einige wenige vorgegebene intermediäre Ziele (Reduzierung der Arbeitslosigkeit, Wirtschaftswachstum, niedrige Inflation o. ä.) konzentriert, ohne eine sorgfältig begründete und präzise Antwort darauf zu geben, wie diese Ziele untereinander, mit anderen Zielen und mit möglichen letzten Zielen systematisch zusammenpassen bzw. wie die verschiedenen Ziele gegeneinander zu gewichten sind.⁶ Ebenfalls als „Pragmatismus“ wird man es wohl bezeich-

³ Dabei handelt es sich um den Versuch, über sogenannte Kompensationskriterien (s. u. Abschnitt 2.4) normative ökonomische Urteile (relativ) „wertfrei“ (d. h. nur basierend auf dem Paretoprinzip) und ohne Rekurs auf interpersonelle Nutzenvergleiche zu ermöglichen. Vgl. Samuelson [1950], Chipman u. Moore [1978], Chipman [1987, 526-528].

⁴ D. h. durch eine Funktion, die alle relevanten sozialen Zustände ordnet. Vgl. Sen [1970, 33-35] u. Samuelson [1977].

⁵ Der folgende kursorische Überblick ist weit davon entfernt, vollständig zu sein, und beschränkt sich auf die Angabe von einigen wenigen Beispielen oder Verweisen für die erwähnten Möglichkeiten.

⁶ Ein prominentes Beispiel dafür sind die Jahresgutachten des „Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung“, der seine Ziele im wesentlichen aus dem deutschen und europäischen Recht bezieht, relativ frei gewichtet und durch einige weitere Ziele ergänzt (u. a. Effi-

nen müssen, wenn – wie etwa z. T. in der Kosten-Nutzen-Analyse – normative Kriterien weiterverwendet werden, deren Inkonsistenz wohlbekannt ist.⁷

Eine weitere Gruppe von Autoren geht für ihre wohlfahrts-theoretischen Überlegungen von einer zu maximierenden Samuelson-Bergsonschen sozialen Wohlfahrtsfunktion der Form $W = W(u)$ aus (u ist dabei der Vektor der individuellen Nutzen und W monoton wachsend in u). Während das in der Tat die Anforderung, eine hinreichende normative Basis für wirtschaftspolitische Empfehlungen anzugeben, erfüllt, wenn die Wohlfahrtsfunktion (inklusive der Nutzenfunktionen) tatsächlich spezifiziert wird, ist dies *nicht* der Fall, wenn – wie in einem großen Teil dieser Literatur⁸ – nur diese allgemeine Form zugrundegelegt wird. Dann handelt es sich im wesentlichen⁹ um eine Charakterisierung der Nutzenmöglichkeitsgrenze und damit der (unter den gegebenen Beschränkungen!) paretoeffizienten sozialen Zustände, und die angegebene normative Basis (die allgemeine Form der Wohlfahrtsfunktion) ist *nicht* ausreichend, um politische Handlungsempfehlungen herzuleiten: Die normative Basis ist dann nämlich im wesentlichen nichts anderes als das Paretoprinzip. [Vgl. Samuelson 1963, 203-253]

Nicht zuletzt gibt es einen großen Korpus von wirtschaftswissenschaftlicher Literatur, die von Paretosuperiorität und ökonomischer Effizienz – verstanden als Paretoeffizienz¹⁰ – in einer Weise spricht, die – vorsichtig gesagt – vermuten läßt, daß die Autoren annehmen, daß damit ein signifikanter Beitrag zur Beantwortung normativer ökonomischer Fragen geleistet wurde: sei es, daß wirtschaftspolitische Maßnahmen kurzerhand und einfach damit empfohlen werden, daß sie die Effizienz förderten;¹¹ sei es, daß auf Effizienzverbesserungen hingewiesen wird, die aber gegenüber anderen Zielen abzuwägen und deshalb nicht notwendig direkt umzusetzen seien;¹² sei es auch nur, daß die intensive Beschäftigung mit den Bedingungen für Paretoeffizienz in verschiedenen Kontexten und die wertgeladene Sprache zumindest suggerieren, der Autor wolle hier offensichtlich etwas empfehlen, dem man sich, wenn möglich, annähern solle.¹³

In Wirklichkeit ist es aber so, daß der Begriff der Paretoeffizienz – ohne weitere, über

zienz, Standortattraktivität). Vgl. z. B. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung [2003, V f., 324-326, 473-475].

⁷ Zu diesen Inkonsistenzen in der Kosten-Nutzen-Analyse und dem Zusammenhang mit weiteren Kompensationskriterien vgl. z. B. Chipman [1987, 528 f.] und Blackorby u. Donaldson [1985].

⁸ Als prominentes Beispiel aus der Optimalsteuer-Literatur kann etwa Diamond u. Mirrlees [1971] dienen

⁹ Das heißt bei entsprechenden Konvexitätsannahmen.

¹⁰ Ein sozialer Zustand ist paretoeffizient, wenn es keinen verwirklichbaren paretoüberlegenen alternativen sozialen Zustand gibt.

¹¹ Das geschieht etwa, wenn die neoklassische Analyse von Umweltproblemen direkt in politische Handlungsempfehlungen umgesetzt wird, wie z. B. im Bericht der Enquete-Kommission „Schutz der Erdatmosphäre“ des Deutschen Bundestages [1995, 729 f., 1011, 1015 f.]. Kritisch dazu das Minderheitsvotum 1095-1100.

¹² Das gilt etwa für Autoren, die von einem „Trade-off“ zwischen Effizienz und anderen Zielen (vor allem Gerechtigkeit) sprechen. Vgl. dazu Kapitel 3.

¹³ Das gilt m. E. für einen großen Teil der strikt wohlfahrts-theoretischen Literatur. Typischerweise legen sich allerdings viele Autoren, die in der erwähnten Weise von Effizienz, Paretosuperiorität etc. sprechen, nicht darauf fest, welche normativen Konsequenzen nun denn *genau* aus ihren Effizienzbetrachtungen zu ziehen sind, sodaß man bei der Zuordnung zu den genannten Kategorien oft auf Vermutungen angewiesen ist. Zwei Beispiele mögen genügen:

Otani [1972] beginnt seinen Artikel „Gains from Trade Revisited“ folgendermaßen:

1 Einleitung

das Paretoprinzip hinausgehende normative Annahmen – uns nicht einen Schritt weiterbringt in der Beantwortung der normativen Frage. Die bloße Definition eines Begriffs überführt keine einzige Situation, in der das Paretoprinzip uns nicht sagen konnte, was zu tun ist, in eine Situation, in der wir wissen, was zu tun ist. Die bloße Definition eines Begriffs ermöglicht es uns nicht, eine einziges Paar sozialer Zustände, die wir aufgrund des Paretoprinzips nicht ordnen konnten, nun zu ordnen.

In einem ersten Sinn ist Paretoeffizienz also irrelevant, da sich alle rein paretianischen normativen Urteile strikt ohne den Begriff der Paretoeffizienz formulieren lassen.

In der Regel ist aber auch für Anhänger einer bestimmten paretoinklusiven Ordnung sozialer Zustände die Menge paretoeffizienter Zustände irrelevant. Wieder lassen sich die möglichen Antworten auf normative Fragen ganz ohne Rekurs auf „Paretoeffizienz“ formulieren. Unter den paretoeffizienten Zuständen können sich Zustände befinden, die zu den schlechtesten überhaupt möglichen gehören. Maßnahmen, die garantieren, daß sicher ein paretoeffizienter Zustand verwirklicht wird (wie z. B. in der Theorie die Einführung einer der Marktwirtschaft), werden deshalb in der Regel auch Anhänger einer paretoinklusiven Ordnung wenig beeindrucken, solange nicht auch sichergestellt ist, daß durch sie ein Zustand verwirklicht wird, der zu den besten gehört oder wenigstens den besten nahe ist.

Auch die Truismen, daß die Menge der gemäß einer paretoinklusiven Ordnung be-

There have been numerous articles on gains from trade. Two propositions seem to be widely accepted at present. They are, loosely speaking, as follows: (a) Some trade is better than no trade. (b) Some tariff-subsidy ridden trade without retaliation of other countries is better than free trade. [...] In this paper we shall formulate analytical frameworks in a more precise and rigorous fashion than we encounter in the literature on this topic. Then we provide proofs of several versions of (a) and (b) while revealing assumptions we need to establish them.

Ohne weitere Ausführungen dazu, welche Handlungskonsequenzen denn daraus zu ziehen sind, daß eine bestimmte Außenhandelspolitik „besser“ (später u. a. präzisiert durch den Begriff der Paretosuperiorität) ist als eine andere, ist man wohl berechtigt anzunehmen, der Autor empfehle die „bessere“ Politik, wenn die notwendigen Annahmen (für praktische Zwecke hinreichend) erfüllt sind.

Noch klarer wird die Verknüpfung von Paretoeffizienz und wirtschaftspolitischer Handlungsempfehlung von Keen u. Wildasin [2000] schon in der Zusammenfassung ihres Arbeitspapiers formuliert. Aber auch hier bleibt offen (und das gilt für das ganze Papier) welche Empfehlungen die Autoren denn *genau* aus ihren Ergebnissen ableiten:

This paper addresses a key but neglected task in the theory of international taxation, lent increased urgency by growing awareness of the potential gains from tax coordination: the characterization of Pareto-efficient international tax regimes. It shows that the Diamond-Mirrlees theorem on the desirability of production efficiency, which underlies the key tenets of policy advice in international taxation - the desirability of destination basis for commodity taxation, of the residence principle for capital income taxation, and of free trade - is rendered inherently inapplicable to problems of international tax design by the distinctness of national budget constraints that is of the essence in thinking about international taxation. Conditions are established - relating to the availability of explicit or implicit devices for reallocating tax revenues across countries - under which production efficiency is nevertheless desirable, and a general characterization developed of the precise ways in which Pareto-efficient international taxation may require violation of established tenets.

sten sozialen Zustände eine Teilmenge der Menge der paretoeffizienten Zustände ist, oder daß deshalb Paretoeffizienz eine notwendige aber nicht hinreichende Bedingung dafür ist, daß ein Zustand zu den besten sozialen Zustände gehört, ändert an dem allen nichts: Sie machen schlechte paretoeffiziente Zustände keinen Deut besser und Maßnahmen, die Paretoeffizienz aber nicht Nähe zu den besten Zustände garantieren, keinen Deut attraktiver.

Die Karriere des Konzepts der Paretoeffizienz ist vor diesem Hintergrund ein Rätsel. Wenn man nicht annehmen möchte, daß es sich bei der Wohlfahrtstheorie um eine bloße theoretische Spielerei mit wertendem Vokabular handelt, oder daß Ökonomen, die das Konzept der Effizienz hochhalten, einfach verwirrt sind, wenn sie auf normative Fragestellungen stoßen, bleibt eigentlich nur die Vermutung, daß sie in Wirklichkeit – wenn vielleicht auch uneingeständenerweise – über das Paretoprinzip hinausgehende normative Annahmen voraussetzen.¹⁴

Damit ist im wesentlichen die Aufgabe der Kapitel 2 und 3 meiner Arbeit gegeben: Es wird einmal darum gehen, das gerade Ausgeführte noch genauer darzustellen und zu begründen. Zum anderen wird es darum gehen, einige der möglichen Ergänzungen des Paretoprinzips zu diskutieren, die die extensive Beschäftigung mit Effizienz verständlich machen könnten, da sie Paretoeffizienz normativ relevant machen.

Auch in der Diskussion dieser Ergänzungen bleibt das Ziel im wesentlichen ein destruktives. Zwar würden Ökonomen, wenn sie denn diese Ergänzungen explizit als ihre normative Basis angeben würden, nicht mehr dem oben zitierten Verdikt Samuelsons oder der Kritik Chipman und Moores unterliegen. Es wird aber deutlich werden, daß diese Ergänzungen, wenn sie nicht sowieso inkonsistent sind, oder nur unwesentlich über das Paretoprinzip hinausgehen, ethisch unattraktiv sind, ja – offengelegt – geradezu als abstoßend gelten können.

Ausgehend von einer genaueren Definition der Begriffe „Paretorelation“, „Paretokriterium“ und „Paretoprinzip“ (2.2.1) stelle ich zunächst in Abschnitt 2.2.2 noch einmal ausführlicher dar, was ich die „Irrelevanz“ des Paretokriteriums und des Paretoprinzips nenne. Damit meine ich die bekannte Tatsache, daß Paretokriterium und -prinzip in aller Regel nicht ausreichen werden, um wirtschaftspolitische Entscheidungen zu treffen. Es folgen Ausführungen zur „Irrelevanz“ der Paretoeffizienz (2.3.1), die nicht mehr bedeuten soll als zum einen die Selbstverständlichkeit, daß die bloße Addition einer Definition die normative Reichweite von Paretokriterium und -prinzip nicht ausweitet, und zum anderen, daß auch als Hilfskonstruktion auf dem Weg zu normativen Aussagen aufbauend auf Paretokriterium und -prinzip die Menge der paretoeffizienten Zustände wenig bis nichts beiträgt.

In dreierlei Hinsicht werden diese Ausführungen dann noch vertieft: Zuerst in Hinblick auf zwei verschiedene Interpretationen der unter Ökonomen verbreiteten Skepsis gegenüber über das Paretoprinzip hinausgehenden Werturteilen, die entweder so verstanden werden kann, daß solche Werturteile überhaupt zu vermeiden seien, oder so,

¹⁴ Dies gilt unabhängig davon, ob Effizienz als ein ausreichendes Kriterium für die Empfehlung einer wirtschaftspolitischen Maßnahme angesehen wird, oder nur als ein (evtl. *das* ökonomische) Kriterium, das durch andere zu ergänzen und ihnen gegenüber abzuwägen.

1 Einleitung

daß es nur nicht möglich ist, sich auf solche Werturteile zu einigen (2.3.2). In diesem Zusammenhang behandle ich auch – kritisch – das Paretoextensionskriterium, dessen Übernahme durch Ökonomen zwar mit der erwähnten Skepsis unvereinbar wäre, aber doch die extensive Beschäftigung mit der Paretoeffizienz so gut erklären kann, daß man vermuten konnte, es liege – trotz der geäußerten Skepsis, und wenn auch nur implizit – in Wirklichkeit der modernen Wohlfahrtstheorie zugrunde [Sen 1970, 197].

Eine weitere Vertiefung (2.3.3) erfolgt in Hinblick auf die Möglichkeit, die normative Frage nicht in Bezug auf einzelne gesellschaftliche Zustände sondern in Bezug auf Maßnahmen zu verstehen, die sicherstellen, daß Zustände aus einer Menge von Zuständen verwirklicht werden, eine Möglichkeit, die z. B. Samuelson [1950] kurz diskutiert hat. Eine solche Möglichkeit setzt aber wieder eine Ergänzung des Paretoprinzips voraus, die ähnlichen ethischen Bedenken ausgesetzt wäre wie das Paretoextensionskriterium, da sie gegenüber der Verteilungsfrage völlig unsensibel wäre.

Die letzte Vertiefung (2.3.4) erfolgt im Hinblick auf die beiden grundlegenden Wohlfahrtstheoreme, deren normative Irrelevanz eine direkte Folge der normativen Irrelevanz der Menge paretoeffizienter Zustände ist.

Es folgt eine Diskussion der Kompensationskriterien (2.4), die ihren Ausgang nehmen von dem Versuch, „wertfrei“ über die normativen Folgerungen des Paretoprinzips hinausgehen zu können. Ohne zusätzliche normative Annahmen ist das natürlich schon logisch unmöglich. Außerdem hat sich aber im Laufe der Diskussion gezeigt, daß unter den vorgeschlagenen Kriterien nur das Samuelson-Kriterium gewisse Grundanforderungen der Konsistenz (Asymmetrie und Azyklizität) erfüllt. Das Samuelson-Kriterium würde aber nur dann zu über das Paretoprinzip hinausgehenden normativen Folgerungen führen, wenn man die ethische problematische Interpretation der ethischen Frage übernehmen würde, die auf die Verwirklichung von Mengen von Zuständen, statt auf die Verwirklichung von Zuständen abhebt.

Abgeschlossen wird das Kapitel 2 durch eine Diskussion des fragwürdigen Anspruchs, das Paretoprinzip selbst sei unproblematisch oder nicht kontrovers (2.5).

Während in Kapitel 2 die Frage im Vordergrund steht, welche normativen Konsequenzen sich aus der eingestanden Basis der paretianischen Wohlfahrtstheorie, d. h. aus dem Paretoprinzip herleiten lassen, fragt das Kapitel 3 nach dem Verhältnis von auf der Basis des Paretoprinzips definierter Effizienz zur Gerechtigkeit, genauer nach dem behaupteten Trade-off zwischen Effizienz und Gerechtigkeit. Einen solchen Trade-off kann es nicht geben, wenn wir auf der einen Seite die normative Basis der Effizienz strikt auf das Paretoprinzip beschränken, und auf der anderen unsere Gerechtigkeitsvorstellung (wie nur z. B. im Utilitarismus) paretoinklusiv ist. Außerdem ist es wohl nur dann sinnvoll von einem *Trade-off* zu sprechen, wenn man dem im Paretoprinzip ausgedrückten Werturteil zustimmt, und es deshalb in der Tat immer als einen gesellschaftlichen, ethischen Verlust ansieht, wenn eine mögliche Paretoverbesserung nicht verwirklicht wird. Wenn man sich dieses Werturteil aber zu eigen gemacht hat, ist es doch zumindest nicht auszuschließen, daß es auch Teil der eigenen Gerechtigkeitsvorstellung bildet. Die Selbstverständlichkeit, mit der Ökonomen davon ausgehen, daß es einen Trade-off zwischen Effizienz und Gerechtigkeit gibt, ist also unbegründet.

Wenn man denn von einem solchen Trade-off sprechen möchte, müßte man also zu-

mindest genauer und explizit angeben, was man unter Gerechtigkeit versteht¹⁵, und was unter Effizienz, eventuell über der Paretoprinzip hinausgehend oder dieses ersetzend.

Da für eine Vielfalt von Gerechtigkeitsvorstellungen kein Trade-off zwischen Gerechtigkeit und Effizienz vorliegen kann, wenn man die normativen Forderungen der Effizienz allein auf der Basis des Paretoprinzips, ja selbst des Paretoextensionskriteriums formuliert, liegt es nahe zu vermuten, daß diese weitverbreitete Vorstellung eines Trade-offs auf einer weitergehenden normativen Basis beruht. Als Beispiele diskutiere ich die Idee des Realeinkommens¹⁶ (3.3) und die intuitive Idee, daß das Maß der Effizienz vom Abstand zur Paretogrenze abhängt (3.4).

Die meisten Vorschläge für die Messung des Realeinkommens sind nicht einmal paretoinklusiv (3.3). Wer Effizienz auf ihrer Basis bestimmt, sollte also zumindest offenlegen, daß er über etwas ganz anderes als Paretoeffizienz spricht. Eine Ausnahme bildet die Definition des Realeinkommens über die Summe des geldmetrischen Nutzens. Das ist natürlich nichts anderes als eine Version des Utilitarismus, einer substantiellen und umstrittenen ethischen Theorie, und es nicht ganz klar, weshalb Ökonomen gerade Anhänger des Utilitarismus sein sollen. Auch wer seine Effizienzurteile auf dieser Definition des Realeinkommens aufbauen möchte, sollte also zumindest offenlegen, daß seine Idee der Effizienz utilitaristisch und nicht rein paretianisch ist. Die konsistente Bestimmung des Realeinkommens verstanden als Summe geldmetrischer Nutzen verlangt die Angabe eines über verschiedene Entscheidungen festen Basispreisvektors, ohne Angabe eines solchen Vektors ist die Effizienzordnung nicht festgelegt. Die damit verbundenen Schwierigkeiten werden noch dadurch verschärft, daß nur mit sehr einschränkenden Annahmen über die individuellen Präferenzen ausgeschlossen werden kann, daß unausgewogene Verteilungen ausgewogenen vorgezogen werden, und deshalb in bestimmten Bereichen kleine Änderungen des Basispreisvektors große Änderungen in den gemäß der Effizienzordnung optimalen sozialen Zuständen bewirken. [Blackorby u. Donaldson 1988]

Die Idee dagegen, eine Effizienzordnung über den Abstand von der Paretogrenze zu bestimmen (3.4), ist in der Regel (solange diese Ordnung über verschiedene Entscheidungssituationen festgehalten werden soll und nicht ausgeschlossen werden kann, daß sich die Paretogrenzen verschiedener Entscheidungssituationen überkreuzen) nicht konsistent durchführbar. Das gilt auch, wenn man von sozialen Entscheidungen auf der Basis von sozialen Ordnungen allgemeiner auf soziale Entscheidungsfunktionen übergeht, die aber bestimmte Konsistenzbedingungen z. B. über binäre Entscheidungen erfüllen sollen.¹⁷

¹⁵ Wozu man den weitverbreiteten ethischen Skeptizismus zumindest insoweit verlassen müßte, daß man einen genaueren Begriff davon entwickelt, was überhaupt als eine Gerechtigkeitsvorstellung gelten kann, und die Gerechtigkeitsvorstellungen angibt, die einer Auseinandersetzung und Konfrontation mit Effizienzvorstellungen wert sind.

¹⁶ Die im übrigen eng mit den schon früher diskutierten Kompensationskriterien zusammenhängt.

¹⁷ Diese Konsistenzprobleme gelten allgemein für jede Form, soziale Entscheidungen darauf aufzubauen, daß paretooptimale Zustände besser sind als nicht-paretooptimale. Vgl. dazu in Kapitel 3 die Anmerkung 17 und die Minimalbedingungen S. 107 f.

1 Einleitung

Es gäbe deutlich weniger Schwierigkeiten, zu verstehen, was Ökonomen denn nun genau meinen, wenn sie vom Trade-off zwischen Effizienz und Gerechtigkeit sprechen, wenn sie zum einen eine Ordnung über sämtliche möglichen gesellschaftlichen Zustände (d. h. für viele ökonomische Anwendungen über sämtliche Allokationen) angeben würden, die ihre Effizienzurteile beschreibt, und zum anderen eine Ordnung, die Gerechtigkeiturteile beschreibt und zuletzt eine Methode, diese beide Ordnungen für soziale Entscheidungen zu kombinieren. Dies wäre auch ein schöner Ausdruck intellektueller Redlichkeit und Offenheit, der es sofort ermöglichen würde, die Werturteile der Ökonomen einer kritischen Bewertung zu unterziehen: Ist die Effizienzordnung attraktiv, vernünftig, drückt sie auch *unsere* Werturteile aus? Entspricht die Gerechtigkeitsordnung auch unsere Idee von Gerechtigkeit? Sind die Entscheidungen, die sich aus der Kombination der beiden Ordnungen ergeben ethisch akzeptabel?

Leider ist das bis auf Ausnahmen nicht die Vorgehensweise, den Trade-off von Effizienz und Gerechtigkeit zu beschreiben. Es ist es deshalb nicht überraschend, daß schon andere, berufenere Autoren vor mir das Bedürfnis hatten, das Gemeinte zu präzisieren. Dies gilt z. B. für Atkinson u. Stiglitz [1980], mit deren Vorschlag zur Präzisierung ich mich zum Abschluß des Kapitels (3.5) auseinandersetze. Daß die Autoren nicht den Weg einschlagen, zur Präzisierung eine Effizienz- und Gerechtigkeitsordnung anzugeben, macht schon skeptisch. Statt dessen versuchen sie für *eine* Wohlfahrtsfunktion anzugeben, inwieweit sie Effizienz- und inwieweit sie Gerechtigkeitsziele ausdrückt. Das überraschende Ergebnis der Analyse ihres Vorschlags läßt sich dann kurz so zusammenfassen: Gerechtigkeit ist das, was man durch (verlustfreie) Umverteilung erreichen kann. Eine Wohlfahrtsfunktion verrät dann kein Interesse an Gerechtigkeit, wenn die in ihr erfaßten Gerechtigkeitsziele gerade erfüllt sind, weil sich durch Umverteilung nichts mehr verbessern läßt. Diese Interpretation des Verhältnisses von Gerechtigkeitszielen zu Wohlfahrtsfunktionen wird man wohl eigenwillig nennen dürfen. Sie ist in jedem Fall unvereinbar mit der Idee, daß eine Ordnung gesellschaftlicher Zustände unsere Gerechtigkeitsziele *ausdrückt*, und zwar unabhängig davon, welcher Zustand gerade realisiert ist. Daß Atkinson u. Stiglitz [1980] ihren Vorschlag als eine *Präzisierung* des Verständnisses des Trade-offs von Effizienz und Gerechtigkeit verstehen konnten, scheint mir jedenfalls deutliches Anzeichen eines *Präzierungsbedarfs*.

Die Bedeutung, die die paretianische Wohlfahrtstheorie trotz ihrer Probleme in den Wirtschaftswissenschaften gewonnen hat, hängt ohne Zweifel mit der weitverbreiteten Überzeugung zusammen, Werturteile hätten in den Wissenschaften nichts zu suchen, u. a. da sie nicht objektiv begründbar seien. Natürlich ist auch die Wohlfahrtstheorie selbst mit dieser Position nicht vereinbar: Auch die Aussage, eine Situation sei einer zweiten gegenüber besser, da sie paretoüberlegen ist, ist ein Werturteil. Konsequenz wäre nur der vollständige Verzicht auf normative Urteile in den Wirtschaftswissenschaften.

Normative Aussagen in den Wirtschaftswissenschaften sind nur möglich, wenn man den generellen Bann gegen Werturteile in den Wissenschaften aufhebt, was sicher leichter fällt, wenn man der Überzeugung ist, daß objektive Werturteile möglich sind. Das Kapitel 4 meiner Arbeit trägt deshalb einige Argumente dafür zusammen, daß Werturteile nicht weniger objektiv möglich sind als wissenschaftliche Tatsachenurteile, u. a.

da wissenschaftliche Tatsachenurteile von Werturteilen abhängen. Jedenfalls gibt es – wie ich zu zeigen versuche – keine einfache Trennungslinie zwischen objektiv und kognitiv begründbaren Tatsachenurteilen und (angeblich) nur subjektiven Werturteilen. Innerhalb der philosophischen metaethischen Diskussion gibt es gewichtige Stimmen, die die Möglichkeit objektiver, rationaler, kognitiver moralischer Urteile verteidigen.¹⁸ Da das Kapitel als eine Art Scharnier zwischen dem ersten – negativen – Hauptteil und dem zweiten Hauptteil gedacht ist, der die ethische Position Rawls' darstellen, verteidigen und auf seine praktische Tragfähigkeit hin auswerten möchte, werde ich mich bei meiner Verteidigung des ethischen Kognitivismus vor allem auf solche Argumente stützen, die mit dem Rawlsschen *Konstruktivismus* [Rawls 1999c] gut vereinbar sind und sich u.a. bei Putnam [1981], Kutschera [1982] und Rescher [1991] finden.¹⁹

Der zweite Hauptteil meiner Arbeit möchte einen *konstruktiven* Beitrag zur Begründung von Werturteilen in der Ökonomie leisten und schlägt vor, die Rawlssche Theorie der Gerechtigkeit dafür als Basis zu benutzen.²⁰

In Kapitel 5 versuche ich zunächst eine Kurzdarstellung der Rawlsschen Theorie der Gerechtigkeit. Es ist klar, daß eine solche Kurzdarstellung der Komplexität und Vielschichtigkeit einer Theorie, die zunächst in einem Werk von rund 600 Seiten [Rawls 1973] vorgelegt und dann über 25 Jahre in vielen Artikeln und zwei weiteren Büchern ergänzt, präzisiert und weitergeführt wurde, nicht gerecht werden kann. Die Darstellung möchte deshalb vor allem zwei Dinge leisten: Sie soll einen systematischen Überblick über das Gesamte des Rawlsschen Argumentationsgangs geben und dabei deutlich machen, welche Rolle der Rawlssche Urzustand in einer konstruktivistischen Gesamtkonzeption von ethischer Argumentation spielt (insbesondere, wie er mit der Idee des Überlegungsgleichgewichts verknüpft ist), und sie soll die Elemente besonders betonen, die Ansatzpunkte für eine normative Ökonomie bilden können. Diese Konzentration auf die Aspekte, die für eine Skizze des Argumentationsgangs notwendig sind, oder die eine besondere Relevanz für eine normative Ökonomik haben, bringt es mit sich, daß viele andere wichtige Fragestellungen einfach ausgeklammert bleiben werden. Eine weitere wichtige Einschränkung meiner Darstellung ist es, daß sie versucht, die Rawlssche Theorie der Gerechtigkeit als *einen* Theorieentwurf darzustellen, und bis auf einige kurze Hinweise darauf verzichtet, die Entwicklung des Rawlsschen Denkens nachzuvollziehen. Zwei Gründe sind dafür ausschlaggebend: Ziel meiner Ausführungen ist es ja, *einen bestimmten* Ausgangspunkt für normative Urteile in der Ökonomie vorzuschlagen, zu begründen und auf seine Reichweite hin kritisch auszuwerten. Von daher scheint es sinnvoll, die Rawlssche Theorie der Gerechtigkeit als *eine* Theorie vorzustellen. Außerdem scheint mir, daß man die Weiterentwicklung der Rawlsschen Theorie durch Rawls selbst durchaus als Klärung, Präzisierung und Korrektur in Ein-

¹⁸ Neben den in Kapitel 4 zitierten Autoren sind im deutschsprachigen Raum insbesondere auch Apel und Habermas zu nennen.

¹⁹ Womit ich die Autoren selbst nicht als „Konstruktivisten“ bezeichnet haben möchte.

²⁰ Wenn Ökonomen sich für einen expliziten ethischen Entwurf als Basis ihrer normativen Aussagen entscheiden, dürfte (faktisch) vielen eine utilitaristische Position näher liegen als eine rawlsianische. Rawls selbst hat seine Theorie explizit als Gegenentwurf zum Utilitarismus verstanden, seine Kritik am Utilitarismus mache ich mir zueigen. Siehe u. a. Abschnitt 6.4.3 und S. 142 f. und 239.

1 Einleitung

zelpunkten verstehen kann, die das Ganze der Theorie nicht in Frage stellen.²¹ Jedenfalls wird das das hermeneutische Prinzip sein, von dem aus ich die Rawlssche Theorie darstellen werde.

Es ist wohlbekannt, daß Rawls seine Konzeption der Gerechtigkeit, d. h. die zwei Prinzipien der Gerechtigkeit, unter anderem damit begründet, daß sie in einem von ihm beschriebenen Urzustand von den Teilnehmern („Parteien“) dieses Urzustands gewählt würden, wenn diesen bestimmte Informationen, wie vor allem die über ihre Identität, verwehrt sind. Rawls beansprucht, daß er die Interessen der Parteien des Urzustands über die Annahme der schwachen Theorie des Guten hinreichend genau bestimmt habe, um im Urzustand eine Entscheidung im Sinn des „in der sozialwissenschaftlichen Theorie“ üblichen „Standardbegriff“ von Rationalität zu ermöglichen. Mit diesem Standardbegriff von Rationalität meint er, wie die von ihm angeführte Literatur und sein Verweis auf „ein widerspruchsfreies System von Präferenzen bezüglich der ihm offenstehenden Möglichkeiten“ zeigt, den Rationalitätsbegriff der in der Ökonomie verbreiteten Theorie der rationalen Entscheidung.²² [Rawls 1975, 166 f.]

Rawls selbst hat nun allerdings selbst kein formales (mathematisches) Modell vorgelegt, in dem er seinen Anspruch nachweisen könnte. Statt dessen legt er ein ausführliches und komplexes verbales Argument vor, in dem dazu noch die keineswegs zum Standard moderner Theorie rationaler Entscheidung gehörende „Maximinregel“ der Entscheidung unter Unsicherheit eine prominente Rolle spielt [Rawls 1975, 174-185, 201-210, Rawls 2003, 152-187], wie auch ein Appell an eine *abwägende* praktische Vernunft²³, wenn er die Entscheidung für das Differenzprinzip mit einer „Bilanz von Gründen“ rechtfertigt. Dieser Rückgriff auf eine Bilanz von Gründen, und damit auf ein Abwägen, ist für Rawls unvermeidlich, weshalb ein rein deduktives Ableiten der Prinzipien der Gerechtigkeit ein „nicht zur Gänze“ erreichbares Ideal darstellt. [Rawls 2003, 207-210]

Es ist deshalb kein Wunder, daß der Anspruch Rawls', im von ihm beschriebenen Urzustand würden sich die Parteien für seine zwei Prinzipien der Gerechtigkeit entscheiden, bei vielen Ökonomen – vorsichtig formuliert – auf Skepsis stößt.²⁴ Zumal, da vor dem Hintergrund der subjektiven Erwartungsnutzentheorie das Entscheidungsproblem im Urzustand als in der Tat leicht lösbar erscheinen könnte, mit einem allerdings ganz anderen Ergebnis als dem von Rawls behaupteten. Liegt es doch nahe, einfach den Urzustand als eine Situation der Unsicherheit zu modellieren und direkt die subjektive Erwartungsnutzentheorie darauf anzuwenden: Es wird also angenommen, die Parteien im Urzustand hätten eine Nutzenfunktion über die verschiedenen

²¹ Zu einer möglichen Ausnahme vgl. unten S. 131 Anmerkung 2.

²² Seine eigene Darstellung rationaler Entscheidungen [vor allem Rawls 1973, 407-424] ist aber keinesfalls einfach *deckungsgleich* mit diesem ökonomischen Rationalitätsbegriff. Zum m. E. wichtigsten Unterschied siehe gleich im folgenden S. 17.

²³ Hier charakterisiere ich allerdings als „abwägend“ nicht die praktische Vernunft eines Individuums in einer konkreten Entscheidungssituation, sondern die praktische Vernunft der (fiktiven) Parteien im Urzustand und derjenigen, die versuchen, ihre Entscheidungen nachzuvollziehen. Zum Begriff der „abwägenden Vernunft“ siehe ansonsten Ricken [1993b], Ricken [1993a], Bormann u. Schröder [2004].

²⁴ Z.B. Zajac [1995, 88]: “Offering no evidence to support his claims of what individuals in the original position would do, Rawls makes empirical leaps of faith throughout *A Theory of Justice*.”

individuellen Zustände (Individuum und gesellschaftliche Position), in denen sie sich außerhalb des Urzustandes befinden können, und außerdem subjektive Wahrscheinlichkeitseinschätzungen darüber, welches Individuum sie sind.²⁵ Das Ergebnis ist eine Art Utilitarismus, in dem abhängig von den Wahrscheinlichkeitseinschätzungen allerdings die Nutzen der verschiedenen Individuen verschieden gewichtet sein könnten.²⁶

Der Ansatz setzt die Existenz der Nutzenfunktionen der Individuen wie auch ihrer Wahrscheinlichkeitseinschätzungen voraus, aber es ist nicht klar, wie diese Nutzenfunktionen aussehen sollen und warum sie gleich sein sollen. Ähnliches gilt für die subjektiven Wahrscheinlichkeitsverteilungen, wenn auch die Gleichverteilung hier vielleicht als naheliegend angesehen wird.²⁷ Ein weiterer Anspruch Rawls', daß sich nämlich die Überlegungen aller Parteien im Urzustand so gleichen würden, daß alle einstimmig zum gleichen Ergebnis kämen wird damit ebenfalls zweifelhaft. Ein seinen Prinzipien nahes Ergebnis kommt überhaupt nur zustande, wenn die Parteien aufgrund eines schwer zu rechtfertigenden Pessimismus alle annehmen, mit Sicherheit die Position des am schlechtesten Gestellten einzunehmen, und als Nutzenfunktion denselben Index der Grundgüter verwenden.

Angesichts der Bedeutung, die heute in der Ökonomie formalen, mathematischen Modellen zugemessen wird, scheint es mir deshalb sinnvoll, zu versuchen, die Entscheidung im Urzustand formal so zu rekonstruieren, daß sich ein der Rawlsschen Intuition nahes Ergebnis ergibt. Das ist das Ziel des Kapitels 6.

Es ist unvermeidlich, daß eine solche formale *Rekonstruktion* von ihrer Vorlage in vielen Punkten abweichen wird. Die Komplexität der Argumentation Rawls' läßt sich nicht einfach in ein formales Modell zwängen. Der Anspruch, mit der ich die Rekonstruktion vorlege, ist allerdings der, daß sie wesentliche Elemente der Beschreibung des Urzustands durch Rawls aufnimmt, die wichtigsten Voraussetzungen der Rawlsschen Argumentation verdeutlicht und in ihrem Ergebnis den Rawlsschen Prinzipien der Gerechtigkeit hinreichend nahe kommt.

In einem wichtigen Punkt ist die Rekonstruktion dem ursprünglich vorgelegten Argument in Rawls [1973] deutlich näher, als der revidierten Fassung in Rawls [2001]: In der ursprünglichen Fassung hatte Rawls zwischen der allgemeinen und der bestimmten Konzeption der Gerechtigkeit unterschieden: Die erstere verlangt die Gleichverteilung der sozialen Grundgüter, außer wenn eine Ungleichverteilung den am schlechtesten Gestellten zugute kommt. Bei der zweiten handelt es sich um die berühmten zwei Prinzipien der Gerechtigkeit, die dieses Differenzprinzip nur noch auf „soziale und ökonomische Ungleichheiten“ anwenden, und ihm die Grundsätze gleicher Grundfreiheiten und der fairen Chancengleichheit vorordnen. [Rawls 1973, 302 f.] Entscheidend ist, daß Rawls die *besondere* Konzeption als eine Konkretisierung der *allgemeinen* ansah, die sich aus der allgemeinen unter der Voraussetzung weiterer Annahmen über die Situation, für die Grundsätze der Gerechtigkeit aufgestellt werden sollen, ergibt.²⁸ [Rawls

²⁵ Ähnlich Rawls [1973, 154].

²⁶ Als Beispiel vergleiche man die Herleitung des Utilitarismus durch Harsanyi [1977b, 48-83].

²⁷ Eine kurze, kritische Diskussion der Anwendung des Prinzips vom unzureichendem Grund auf den Urzustand findet sich schon in Rawls [1973, 168 f.].

²⁸ Rawls ist allerdings im wesentlichen nur an der besonderen Konzeption interessiert, cf. Rawls [1973,

1 Einleitung

1973, 62 f., 151 f., 302 f., 542-544] Faktisch handelt es sich bei diesen zusätzlichen Annahmen um eine Kombination einer Annahme über den Wohlstand der Gesellschaft, für die entschieden wird, mit einer Annahme über die allgemeinen, menschlichen Präferenzen bezüglich der Austauschbarkeit von Grundfreiheiten und ökonomischer Wohlfahrt:

[...] as the conditions of civilization improve, the marginal significance for our good of further economic and social advantages diminishes relative to the interests of liberty, which become stronger as the conditions for the exercise of the equal freedoms are more fully realized. Beyond some point it becomes and then remains irrational from the standpoint of the original position to acknowledge a lesser liberty for the sake of greater material means and amenities of office.²⁹ [Rawls 1973, 542]

Meine Rekonstruktion ist also zunächst eine Rekonstruktion der Begründung der allgemeinen Gerechtigkeitskonzeption und fragt erst in einem zweiten Schritt, unter welchen Bedingungen sich im gewählten Rahmen ein Vorrang gleicher Grundfreiheiten begründen ließe.

Im ersten Unterkapitel von Kapitel 6 zähle ich [S. 152 ff.] eine ganze Reihe weiterer Unterschiede meiner Rekonstruktion gegenüber dem Rawlsschen Original auf und versuche, sie kurz zu rechtfertigen.³⁰ Der positive Anspruch, trotz der Unterschiede eine *Rekonstruktion* der Argumentation im Urzustand geleistet zu haben, stützt sich u. a. folgende Gesichtspunkte:

1. Die von mir hergeleitete „beschränkte Leximin-Ordnung“³¹ besitzt eine unverkennbare „Familienähnlichkeit“ mit der Maximinordnung und ist eine Anpas-

63].

²⁹ Schon in der deutschen Ausgabe der Theorie der Gerechtigkeit hat Rawls diese ursprüngliche Fassung des Arguments aufgegeben. Statt dessen argumentiert er dort ausgehend von dem Selbstverständnis freier und gleicher Bürger einer wohlgeordneten Gesellschaft (d. h. „einer Gesellschaft, die durch eine öffentliche Gerechtigkeitskonzeption wirksam geregelt wird“ [Rawls 2003, 25]) und ihren mit diesem Selbstverständnis gegebenen fundamentalen und höherrangigen Interessen, wie dem nach Religionsfreiheit und dem Interesse höchster Stufe, die Freiheit zu erhalten, die letzten Ziele abändern zu können. Während in der ursprünglichen Fassung die Parteien eine Hierarchie ihrer Interessen im Urzustand *ausarbeiten konnten*, sind sie jetzt von vornherein von ihr *geleitet*. Vgl. Rawls [1973, 151 f., 542 f.] und Rawls [1975, 176, 588 f.]. Auch in seinen späteren Arbeiten folgte Rawls dann dieser Argumentationslinie und hat sie weiter systematisiert. Vgl. u. a. Rawls [1992b] und Rawls [2003, 81 f., 178]. Auch später geht Rawls zwar weiter davon aus, daß die zwei Prinzipien der Gerechtigkeit ein gewisses Maß an materiellen Wohlstand voraussetzen, er begründet das jetzt aber nur noch über (u. a. ökonomischen) Voraussetzungen leistungsfähiger und demokratischer Institutionen.

³⁰ Meine Rekonstruktion des Differenzprinzips ist etwas komplizierter als die Maximin-Verteilungsregel und wird nicht auf Repräsentanten von Gruppen, sondern auf Individuen angewendet. Ich verzichte vollständig auf eine Diskussion der Chancengleichheit. Ich setzte eine gegebene Gesellschaftsgröße voraus, die bei Rawls unbekannt ist. Meine Rekonstruktion ist im engen Sinn axiomatisch und verzichtet auf das für Rawls entscheidende Abwägen verschiedener Argumente.

³¹ Wie die Maximinordnung baut die beschränkte Leximinordnung auf einem Index der sozialen Grundgüter auf. Die beschränkte Leximinordnung ordnet die Grundstrukturen von Gesellschaften (formal verstanden als Vektoren bzw. Mengen von gesellschaftlichen Positionen, ihrerseits verstanden als Vek-

sung der letzteren an die Situation verschiedener Präferenzen bei mehreren Grundgütern.

2. Im Unterschied zu anderen Rekonstruktionsversuchen, die zu einer Art von Nutzenrawlsianismus gelangen und damit die Lösung des Problems, interpersonell vergleichbare Nutzenfunktionen für die Mitglieder einer Gesellschaft zu finden, voraussetzen – eine Möglichkeit, die Rawls mit größter Skepsis betrachtet [Rawls 1973, 320-325] – benutzt meine Rekonstruktion, wie Rawls, einen Index von Grundgütern.
3. Die Rekonstruktion versucht nahe an der Rawlsschen Beschreibung des Urzustands zu sein, die wichtigsten Elemente und Axiome der Rekonstruktion haben – wie ich jeweils kurz zu belegen versuche – einen Rückhalt in Voraussetzungen, die auch Rawls explizit oder implizit macht.
4. Die Rekonstruktion ist *unabhängig* davon, ob man annimmt, daß die Bedingungen für subjektive oder objektive Wahrscheinlichkeitseinschätzungen vorliegen oder nicht.³²

Im wesentlichen stützt sich die Rekonstruktion auf drei Elemente, deren Verknüpfung zu einem dem Differenzprinzip nahen Prinzip führt:

1. Es gibt allgemein Situationen und bezogen auf die Entscheidung im Urzustand „gesellschaftliche Positionen“, in denen Menschen sich befinden können, die – ex post – „unerträglich“, „inakzeptabel“ sind, und die deshalb *rationalerweise* – ex ante – mit Sicherheit ausgeschlossen werden, wenn dies möglich ist.

Dies ist eine Überzeugung Rawls' und eine Voraussetzung seiner Darstellung der Argumentation im Urzustand, die das ganze Werk Rawls' zumindest seit der „Theorie der Gerechtigkeit“ durchzieht: Ohne sie sind Rawls' Argumente für die zwei Prinzipien der Gerechtigkeit, für die Anwendbarkeit der Maximin-Entscheidungsregel und für den Vorrang der Freiheit nicht nachvollziehbar. [6.2] Auszuschließen, je – ex post – eine Entscheidung zu bereuen, ist für Rawls Teil seines Verständnisses von Rationalität. [Rawls 1973, 422 f.] Eine Rekonstruktion „seiner“ Darstellung der Argumentation im Urzustand ist deshalb *zumindest* berechtigt, ebenfalls von dieser Voraussetzung auszugehen. Ja, von einer der Rawlsschen Intention treuen Rekonstruktion, wird man sogar verlangen können, daß sie diese Voraussetzung macht.

toren von Mengen von Grundgütern) gemäß der dem Index entsprechenden Leximinordnung der (im Sinne der auf Vektoren bezogenen Kleinerrelation) minimalen Elemente der gesellschaftlichen Positionen. Zur genauen Definition s. S. 172.

³² In Rawls [2001, 98, 101 f.] zeigt Rawls ein deutliches Bemühen, sein Argument für die zwei Prinzipien von der Voraussetzung, daß Wahrscheinlichkeitseinschätzungen nicht möglich sind, möglichst unabhängig zu machen, auch wenn ihm dies – in seiner eigenen Auffassung – nicht vollständig gelingt. Vgl. dagegen noch Rawls [1973, 154-156, 172 f.].

1 Einleitung

Natürlich begibt sich Rawls mit diesem Verständnis von Rationalität in direkten Gegensatz zur modernen Theorie der Entscheidung unter Risiko oder Unsicherheit, die davon ausgeht, daß, wenn die „objektive“ oder subjektive Wahrscheinlichkeitseinschätzung für ein beliebig schlechtes „Ergebnis“ nur klein genug ist, die Möglichkeit dieses Ergebnisses immer ex ante akzeptabel sein wird.

Rawls verwendet also eine von der Erwartungsnutzentheorie ganz verschiedene Theorie der Rationalität, in der es *irrational* wäre, eine Entscheidung zu treffen, die man, falls man Pech hat, bereuen würde, und setzt diese Theorie für die Parteien im Urzustand voraus. Die Auseinandersetzung über diese Theorie der Rationalität findet also nicht im Urzustand statt (dort wird sie vorausgesetzt), sondern in der Diskussion darüber, ob die Rawlssche Gestaltung des Urzustands und seine Ergebnisse mit unseren wohlwollenden Überzeugungen übereinstimmen.

Eine Rekonstruktion der Rawlsschen Argumentation im Urzustand muß sich deshalb nicht schämen, eine von der Erwartungsnutzentheorie abweichende Theorie der Rationalität zu benutzen. Im Gegenteil: Nur so kann die Rawlssche Intuition bewahrt bleiben, und nur so kann der Urzustand ein Ergebnis liefern, das dem nahe ist, das Rawls aus ihm ableitet: Eine Anwendung der modernen Entscheidungstheorie im Urzustand wird ja nur unter extremen Annahmen über den Pessimismus der Parteien, die Rawls zurückweist, zu einem Rawls-nahen Ergebnis führen.

In der Art, in der Rawls die rationale Entscheidung im Urzustand mit den sozial-ethischen Prinzipien einer Theorie der Gerechtigkeit verknüpft, hat die von ihm vorausgesetzte Theorie rationaler Entscheidung ein unmittelbares sozialetische Korrelat: Es gibt für Menschen unerträgliche Situationen oder soziale Positionen. Wenn es möglich ist, zu vermeiden, daß irgendein Mensch eine solche Position einnehmen muß, dann ist das auch ein sozialetische Gebot der Gerechtigkeit.

2. Im Urzustand wissen die Parteien, daß in ihren Gesellschaften „moderate Knappheit“ herrscht, d. h. daß (a) Zusammenarbeit *notwendig* ist, (b) Zusammenarbeit zu „allseitig vorteilhaften Regelungen“ *möglich* ist und (c) das Ergebnis der Zusammenarbeit nicht allen konkurrierenden Ansprüchen genügt. [Rawls 1975, 149 f.] Daß allseitig vorteilhafte Regelungen möglich sind, heißt insbesondere, daß Regelungen möglich sind, die allen ein „zufriedenstellendes Minimum“ bieten, und keinem eine Situation zumuten, die „unerträglich“ wäre.³³ [Rawls 1973, 156, 175]

Rawls versteht das Problem der Gerechtigkeit als die Frage nach der Grundstruktur einer Gesellschaft, die man sich so vorstellt, „daß sie verschiedene soziale Po-

³³ In meiner Rekonstruktion werde ich das einfach voraussetzen. Auch Rawls setzt das für die Entscheidung im Urzustand voraus, es gehört zur Beschreibung der Entscheidungssituation. Ein großer Teil seines Werkes ist aber dem Versuch gewidmet, eine den Gerechtigkeitsprinzipien entsprechende Gesellschaft genauer zu beschreiben, und zu *zeigen*, daß eine solche Gesellschaft, in der Tat eine „überaus befriedigende soziale Welt“ [Rawls 2003, 188] darstellt. Vgl. u. a. den Verweis, Rawls [1973, 180], auf den zweiten Teil der Theorie der Gerechtigkeit, den entsprechenden Verweis, Rawls [2003, 159 f.], auf den Teil IV des „Neuentwurfs“ und außerdem Rawls [2003, 181-188].

sitionen enthält, und daß die Menschen, die in sie hineingeboren werden, verschiedene Lebenschancen haben, die teilweise vom politischen System und von den wirtschaftlichen Verhältnissen abhängen.“ [Rawls 1973, 23] Er antwortet mit einem System der reinen Verfahrensgerechtigkeit, die es unnötig macht „den unendlich verschiedenen Einzelumständen und den sich verändernden Stellungen einzelner Menschen“ nachzugehen und explizit nicht mit einem System der allokativen Gerechtigkeit, die danach fragt wie „eine gegebene Gütermenge auf bestimmte Menschen mit bekannten Wünschen und Bedürfnissen zu verteilen ist“. [Rawls 1975, 108 f.] Anders als etwa im Utilitarismus findet eine Zuordnung von gesellschaftlichen Positionen zu Menschen mit als bekannt vorausgesetzten Wünschen und Bedürfnissen (modern: Präferenzen und kardinalen Nutzenfunktionen) also nicht statt, eine Gesellschaftsordnung kann also nur dann für alle zufriedenstellend sein, wenn jede Position für jedes Individuum zufriedenstellend ist.

3. Zwar wissen die Parteien im Urzustand, daß ihre Gesellschaft sich einem Zustand der „moderaten Knappheit“ im gerade beschriebenen Sinn befindet, aber ihnen ist kein *gesellschaftsunabhängiges* „natürliches“ Niveau der Grundgüter bekannt, von dem sie sicher sein könnten, daß es, wenn es allen zur Verfügung gestellt würde, auch für alle akzeptabel wäre. [Vgl. Rawls 2003, 160] Wäre ein solches Niveau bekannt, könnte man annehmen, daß in einer Gesellschaft, in der „moderate Knappheit“ herrscht, es möglich sein sollte, allen Mitgliedern mindestens dieses Niveau anzubieten. Eine einfache „rawlsrationale“ Entscheidungsregel wäre dann einfach die, allen Mitglieder eben mindestens dieses Niveau anzubieten. Während Rawls' Explikation der Grundfreiheiten durchaus den Eindrücken erwecken können, er ginge davon aus, wir wüßten, wie ein ausreichendes und akzeptables Maß an Grundfreiheiten aussieht [vgl. etwa Rawls 2003, 175-181]³⁴, ist in Bezug auf die sozialen und ökonomischen Grundgüter die Vagheit und Unbestimmtheit des Begriffs eines ausreichenden sozialen Minimums, „das ohnehin zum Teil vom Niveau des gesellschaftlichen Wohlstands abhängen wird“, ein wichtiges Argument Rawls' gegen Konzeptionen der Gerechtigkeit, in denen ein solches Minimum das Differenzprinzip ersetzt.³⁵ [Rawls 1973, 316 f., Rawls 2003, 201 f.]

In Kapitel 6 geht es mir im wesentlichen darum, für diese beiden Grundvoraussetzungen eine axiomatische Formulierung zu finden, und aus ihnen – zusammen mit ei-

³⁴ Dabei handelt es sich allerdings nicht um ein aus empirischen Präferenzen abgeleitetes „natürliches“ Niveau oder System von Grundfreiheiten, sondern vielmehr ein aus der *politischen* Konzeption freier und gleicher Bürger abgeleitetes. [Rawls 2003, 44-52, 81 f.]

³⁵ Das Argument baut im wesentlichen darauf auf, daß es *für uns*, außerhalb des Urzustands kaum möglich ist, ein solches Niveau in befriedigender Weise zu bestimmen. Wenn man der Überzeugung wäre, ein solches Niveau wäre in Wirklichkeit hinreichend leicht zu bestimmen, bliebe im Rahmen des Gesamtarguments Rawls' noch immer die Möglichkeit, dieses Wissen, das uns bekannt wäre, den Parteien im Urzustand zu verweigern. [Vgl. allgemein zu dieser Möglichkeit Rawls 2003, 81] Jedenfalls wird von Rawls, wie angesichts seiner Skepsis gegenüber der Bestimmbarkeit eines solchen Minimums zu erwarten, nirgends ein solches Wissen den Parteien unterstellt.

1 Einleitung

nigen weiteren wohl weniger kontroversen und z. T. auch eher technischen Voraussetzungen – die „beschränkte Leximin-Ordnung“ herzuleiten, genauer, formal zu zeigen, daß die Anwendung dieser Ordnung als gesellschaftliches Entscheidungskriterium sicherstellt, daß unter den gegebenen Voraussetzungen in der Tat ein Zustand verwirklicht wird, der für jedes Individuum zufriedenstellend ist, unabhängig davon, welche gesellschaftliche Position es einnimmt („Rawlsrationalität“).³⁶ [6.4.1]

Negativ zeige ich dann, daß die einfache Rawlssche Maximin-Entscheidungsregel die Bedingung der „Rawlsrationalität“ nicht erfüllt. [6.4.2.1] Wie zu erwarten, ist der Vorrang der gleichen Freiheit nur zu begründen, wenn wir über die ursprünglichen, eine Fassung der allgemeinen Gerechtigkeitskonzeption begründenden Axiome hinausgehend weitere Voraussetzungen machen: Anders, als die Ausführungen oben vermuten lassen, müssen wir zusätzlich annehmen, daß es nicht „notwendig“ ist, eine zu geringe Ausstattung mit anderen Grundgütern durch ein Mehr an Grundfreiheiten auszugleichen, oder durch die Gewährung eines Mehr an Grundfreiheiten, die gesellschaftliche „Produktion“ der übrigen sozialen Grundgüter zu fördern. [6.4.2.2]

Rawls selbst verstand seine Theorie ausdrücklich als eine Alternative zum Utilitarismus, weshalb es ihm wichtig war, zu zeigen, daß im Urzustand der Utilitarismus nicht gewählt würde. Sein entscheidendes Argument ist, daß der Utilitarismus die Parteien nicht gegenüber der Möglichkeit, sich in eine inakzeptablen Position zu finden, absichert (also die Bedingung der „Rawlsrationalität“ nicht erfüllt), während seine zwei Prinzipien eine solche Absicherung gewähren (also die Bedingung der „Rawlsrationalität“ erfüllen). Über Rawls hinausgehend kann zunächst gezeigt werden, daß unter bestimmten schwachen Annahmen über die möglichen gesellschaftlichen Zustände Utilitarismus überhaupt mit Rawlsrationalität unvereinbar ist, solange wir den axiomatischen Rahmen nur insoweit verändern, daß wir für die Lösungen des Entscheidungsproblems – wie im Utilitarismus – feste Zuordnungen zwischen gesellschaftlichen Positionen und Nutzenfunktionen annehmen, aber ansonsten von den gleichen Voraussetzungen ausgehen, insbesondere also annehmen, daß es gesellschaftliche Zustände gibt, so daß alle soziale Positionen für alle Individuen akzeptabel sind. Aber auch wenn wir annehmen, daß sich die Parteien in einer Entscheidungssituation befinden, in der sie davon ausgehen können, daß es für alle möglichen Präferenzen und für alle Möglichkeiten dessen, was sie – mit den übrigen Axiomen vereinbar – als inakzeptabel ansehen werden, Nutzenfunktionen gibt, sodaß der Utilitarismus der Bedingung der Rawlsrationalität erfüllt, folgt nicht, daß die Parteien den Utilitarismus wählen würden: Ohne eine Spezifikation des Verfahrens, mit dem die Nutzenfunktionen bestimmt werden, ist der Utilitarismus unterbestimmt, und die Nutzenfunktionen müssen sehr bestimmte und keineswegs plausible Eigenschaften erfüllen, wenn die Rawlsrationalität gewährleistet sein soll. Ein entsprechend spezifizierter Utilitarismus wurde aber noch nicht vorgelegt, er steht deshalb auch nicht zur Wahl, während mit dem Differenzprinzip formalisiert als „beschränkte Leximin-Ordnung“ ein Entscheidungsprinzip vorliegt, das die Bedingung der Rawlsrationalität erfüllt. [6.4.3]

Die Entscheidung, die Rawlssche Theorie der Gerechtigkeit zur Grundlage von nor-

³⁶ Das schließt nicht aus, daß es andere Ordnungen gibt, die diese Bedingung ebenfalls erfüllen!

mativen Aussagen in der Ökonomie zu machen, hat unmittelbar weitreichende Konsequenzen, solange man davon ausgeht, daß die Fragestellung gut durch ein statisches ökonomisches Modell erfaßt werden kann und man sich auf die im engeren Sinn ökonomischen Konsequenzen – d. h. die Versorgung von Menschen mit wirtschaftlichen Gütern – wirtschaftspolitischer Entscheidungen beschränken kann, wie es in einem Großteil der ökonomischen Literatur ja geschieht: Sowohl von den Auswirkungen auf die gewährten Grundfreiheiten wie von den Auswirkungen auf die Chancengleichheit kann man dann absehen, und man hat es nur noch mit einem *Grundgut* (Einkommen) zu tun.³⁷ Die Rawlssche Vereinfachung, die soziale Position allein anhand von Einkommen und Vermögen zu bestimmen [Rawls 1975, 116-121], erscheint für solche Fragestellungen als gerechtfertigt, und es stellt sich nur noch die Frage, wie in Mehrgütermodellen sinnvoll das Realeinkommen bestimmt wird.³⁸

Die Konsequenzen, die sich aus der Übernahme der Rawlsschen Theorie der Gerechtigkeit für die angesprochenen Fragestellungen ergeben, sind zunächst allerdings zu einem großen Teil negativ oder Hinweis auf ein Desiderat ökonomischer Forschung: Wenn man davon ausgeht, daß eine wirtschaftspolitische Maßnahme Verteilungswirkungen hat, dann ist die relevante soziale Position, von der aus die Maßnahme zu beurteilen ist, diejenige der bezogen auf Einkommen und Vermögen am schlechtesten gestellte Gruppe von Bürgern: Auch wenn man die praktischen Probleme bei der Bestimmung dieser Gruppe berücksichtigt [Rawls 1975, 118 f.], sind Modelle, die auf eine Modellierung von Verteilungswirkungen überhaupt verzichten, oder ihre normativen Konsequenzen aus einer Gesamttaggregation der Wirkungen einer Maßnahme ableiten, zur Beurteilung aufgrund des Differenzprinzip völlig ungeeignet. Das Desiderat sind also theoretische und empirische Modelle, die es bei einem (praktisch) ausreichenden Grad der Disaggregation erlauben, die Verteilungswirkungen auf die am schlechtesten Gestellten nachzuvollziehen.

Ein markantes Beispiel für die „negativen“ Konsequenzen der Übernahme der Rawlsschen Theorie der Gerechtigkeit auf normative Überzeugungen von Ökonomen ist die Frage des Freihandels: Es gibt wenige Fragen, in denen Ökonomen so übereinstimmen, wie in der Überzeugung, daß Freihandel etwas Gutes ist: 92,6 % eines repräsentativen Samples der Mitglieder der American Economic Association stimmten der Aussage zu: „Tariffs and import quotas usually reduce the general welfare of society.“ [Fuller u. Geide-Stevenson 2003, 372] Dies ist zunächst einmal bemerkenswert, weil eine Zustimmung zu dieser Aussage eine Vorstellung davon voraussetzt, was „die allgemeine Wohlfahrt einer Gesellschaft“ denn ist, und das Sprechen von einer „Abnahme“ auch die Vorstellung einer Ordnung oder zumindest einer für die Feststellung der Abnahme hinreichenden Teilordnung. Rein auf der Basis des paretianischen Werturteils der

³⁷ Aus der Perspektive der Rawlsschen zwei Prinzipien der Gerechtigkeit ist es dafür allerdings entscheidend, daß eine wirtschaftspolitische Entscheidung *tatsächlich* die Gewährung gleicher Grundfreiheiten und der fairen Chancengleichheit unberührt läßt. Die Frage, wie Grundfreiheiten und Chancengleichheit von wirtschaftspolitischen Entscheidungen beeinflußt werden, wären also aus einer Rawlsschen Perspektive vorrangige Fragestellungen für die Wirtschaftswissenschaften.

³⁸ Eine Frage, die allerdings nicht einfach als gelöst angenommen werden kann. Vgl. dazu das Unterkapitel 3.3.

1 Einleitung

Pareto-Relation läßt sich aber eine Überlegenheit des Freihandels auch in statischen Modellen, die Verteilungswirkungen berücksichtigen, im allgemeinen nicht begründen. Im Gegenteil: Schon allein das Stolper-Samuelson-Theorem zeigt, daß unter in der Außenhandelstheorie gängigen Idealannahmen der Übergang von prohibitiven Handelsbeschränkungen zu Freihandel keine Paretoverbesserung darstellen kann. Wenn Ökonomen davon ausgehen, daß Zölle und Importquoten „normalerweise“ die Gesamtwohlfahrt einer Gesellschaft reduzieren, müssen sie entweder (was unwahrscheinlich ist) davon ausgehen, daß diese „normalerweise“ zu Paretoverschlechterungen führen, oder sie müssen (was wahrscheinlicher ist) eine über das Paretoprinzip hinausgehende Wertbasis zugrundelegen.³⁹

Das Gut einer wohlgeordneten Gesellschaft, genauer einer Gesellschaft, die entsprechend Rawlsscher Theorie der Gerechtigkeit gestaltet ist, ist eben die wohlgeordnete Gesellschaft, oder mit anderen Worten die Verwirklichung der in den zwei Prinzipien formulierten Gerechtigkeitskonzeption. [Rawls 2003, 302-308] In einer Rawlsschen Perspektive wäre also dieses Gut das *Gemeinwohl* und unter der Voraussetzung, daß die gleichen Grundfreiheiten und die Chancengleichheit gesichert bleiben, wären Gemeinwohl, allgemeine Wohlfahrt und Wohlfahrt der am schlechtesten gestellten Gruppe identisch. [Vgl. Rawls 1975, 120]

Vor diesem Hintergrund kann gesagt werden, daß im Sinne der Rawlsschen Theorie der Gerechtigkeit die These, Freihandel sei gegenüber Beschränkungen des Freihandels durch Zölle oder Importquoten „normalerweise“ mit einer höheren allgemeinen Wohlfahrt verbunden, weder theoretisch noch empirisch hinreichend begründet ist: Was die Theorie angeht, ist es folgend dem Stolper-Samuel-Theorem ein Leichtes, einfache Modelle zu konstruieren und so zu spezifizieren, daß Handelsliberalisierungen zu einem Realeinkommensverlust für Bezieher von Arbeitseinkommen führen. Was neuere empirische oder empirisch angereicherte Studien zu den Auswirkungen von Handelsliberalisierungen auf die Armen angeht, so ist das Ergebnis gemischt und negative Auswirkungen auf das Realeinkommen armer Haushalte können in keiner Weise ausgeschlossen werden.⁴⁰

Zwar sind sich Ökonomen wohl bewußt, daß Handelsliberalisierung in der Regel Gewinner und Verlierer hat⁴¹, aber das Rawlssche Kriterium, d. h. die Frage, wie die Ärmsten von ihr betroffen werden, spielt bei der positiven Gesamtbewertung keine Rolle. So stellt Coughlin [2002, 2-10] die wichtigsten Gründe zusammen, derentwegen Ökonomen Freihandel unterstützen: Zwar geht er dabei auch auf das Stolper-Samuel-Theorem ein, und präsentiert empirische Schätzungen für Konsumentengewinne und Produzentenverluste unterschiedlicher Wirtschaftssparten, aber er geht – solange er die

³⁹ Folgt man Coughlin [2002] handelt es sich dabei um Varianten des Realeinkommens.

⁴⁰ Für Entwicklungsländer siehe z. B. die Übersichtsartikel Winters u. a. [2004] und Hertel u. Reimer [2004]. Für Industrieländer gibt es Hinweise auf (meist als unbedeutend eingeschätzte) negative Auswirkungen auf die Einkommensungleichheit (Lohndifferentiale), die bei schwach steigenden oder gar sinkenden durchschnittlichen Reallöhnen eine größere Armut für schlecht qualifizierte Arbeiter bedeuten könnten. Vgl. z. B. Wood [1995] und Slaughter [1999].

⁴¹ Trotzdem hält es Roberts [2000] für notwendig, davor zu warnen, für Freihandel mit der Behauptung zu werben, er sei gut für *alle*, was seiner Beobachtung nach Verteidiger des Freihandels oft tun.

Gründe der Ökonomen darstellt – selbstverständlich davon aus, daß die Gewinne und Verluste gegeneinander aufgerechnet werden müssen (um eine Gesamtgewinn zu ergeben), und er geht nicht darauf ein, daß die Verluste besonders arme, verletzlichere Teile der Bevölkerung treffen könnten. Diese Idee wird von ihm erst eingeführt, als es darum geht, zu verstehen, weshalb Nicht-Ökonomen Handelsliberalisierungen ablehnen [Coughlin 2002, 11 f.]. Nur in diesem Kontext erwähnt er auch die *Möglichkeit*, daß „eine dritte Partei“ eine Änderung der Handelspolitik ablehnen könnte, wenn der Verlust Individuen mit niedrigem Einkommen betrifft.

Aus einer Rawlsschen Perspektive ist also jedenfalls beim heutigen Forschungsstand eine allgemeine und bedingungslose Befürwortung von Handelsliberalisierungen nicht gerechtfertigt. Gleichzeitig ist aber positiv anzumerken, daß wenigstens in Bezug auf Entwicklungsländer die Frage der Auswirkungen von Handelsliberalisierungen auf die Armen auch auf einem angemessen disaggregierten Niveau zu einem wichtigen Forschungsgebiet geworden ist, das sich gut an die normative Perspektive der Theorie der Gerechtigkeit anschließt, und helfen kann, zu entscheiden, in welchen Situationen und unter welchen Bedingungen Handelsliberalisierungen zum Gemeinwohl einer den zwei Prinzipien der Gerechtigkeit folgenden wohlgeordneten Gesellschaft beitragen.⁴²

Soweit sich die Diskussion im Rahmen (komparativ-)statischer Modelle bewegt (wie nach wie vor ein großer Teil der Freihandelsdebatte), die wirtschaftspolitische Maßnahmen (wie etwa eine vorgeschlagene Handelsliberalisierung) aus der Perspektive *eines* betroffenen Landes betrachtet werden, und ihre Auswirkungen auf Einkommen und Konsum im Vordergrund stehen, kann die normative Beurteilung einer solchen Maßnahme durch die Rawlssche Theorie der Gerechtigkeit in einer (theoretisch) unkomplizierten und direkten Anwendung der zwei Prinzipien der Gerechtigkeit bzw. des Differenzprinzips bestehen.

In Kapitel 7 versuche ich, die Rawlssche Theorie der Gerechtigkeit mit einer Fragestellung zu konfrontieren, die sich für eine einfache „Anwendung“ der zwei Prinzipien der Gerechtigkeit *nicht* eignet, die Frage der internationalen Klimapolitik.

Um die Überlegungen möglichst konkret zu halten, beziehe ich mich dabei durchgängig auf das RICE-Modell, das William Nordhaus (zusammen mit seinen Mitarbeitern) beginnend mit Nordhaus u. Yang [1996] in verschiedenen Versionen vorgelegt hat. Es handelt sich dabei um ein „integriertes Klimamodell“ (ein Modell also, das ein Klimamodell mit einem ökonomischen Modell unmittelbar verbindet), das für einen Zeitraum von 350 Jahren sowohl die globale Klimaentwicklung aufgrund des Treibhauseffekts als auch die wirtschaftliche Entwicklung in acht Weltregionen simuliert.⁴³

Nordhaus versucht, seine normativen Voraussetzungen explizit offenzulegen und bekennt sich dabei zu einer Form von Utilitarismus. [Nordhaus u. Boyer 2000c, 14-16] Konsequenter konstruiert er zunächst in utilitaristischer Manier die Zielfunktionen der Weltregionen direkt als intertemporale, abdiskontierte Summen der von ihm angenommenen Nutzenfunktionen – mit den von ihm geforderten prinzipiellen Eigenschaf-

⁴² Siehe dazu z. B. die oben in Anmerkung 40 genannten Übersichtsartikel.

⁴³ Eine weitere Disaggregation findet nicht statt. Schon deshalb ist eine einfache Anwendung der zwei Prinzipien der Gerechtigkeit bzw. des Differenzprinzips als Gerechtigkeitsprinzipien *einer* Gesellschaft ausgeschlossen.

1 Einleitung

ten, darunter der eines abnehmenden Grenznutzens von Einkommens. In der weiteren Durchführung gibt es dann allerdings – wie ich in Unterkapitel 7.2 zeige – einen Bruch, da die weltweite Zielfunktion dann in der Summe der Gegenwartswerte des Konsums besteht, mit regional unterschiedlichen Diskontierungssätzen, so daß *zwischen* den Regionen höherer Konsum faktisch höher bewertet wird. Eine konsequent utilitaristische Fortsetzung des regionalen Ansatzes wäre die einfache, *gleichgewichtete* Summierung der von Nordhaus in der formalen Darstellung angegebenen regionalen Wohlfahrtsfunktionen gewesen.⁴⁴ Ein entsprechendes Vorgehen stellt die normativen Ergebnisse der Studie Nordhaus' auf den Kopf [S. 204 ff.]: Unter anderem hat seine kritische Evaluation des Kyotoprotokolls keine Basis mehr. Alle von ihm betrachteten Varianten des Kyotoprotokolls sind nach diesem Kriterium *besser* als die von ihm so genannte „optimale“ Klimastrategie, und die dem tatsächlichen Protokoll entsprechende ist nach dem einfachen Verschwindenlassen der Klimaschäden („Geoengineering“) die beste Strategie. Dieser pointierte Unterschied in den normativen Konsequenzen unterstreicht noch einmal die Bedeutung, die die klare Explizierung der normativen Grundlagen besitzt.

Eine erste und wichtige Schwierigkeit für den Versuch, die Konsequenzen der Rawlschen Theorie der Gerechtigkeit für die internationale Klimapolitik anhand des Nordhausschen Modells herzuleiten, ergibt sich aus dem intertemporalen Charakter des Modells (und der Fragestellung). [7.3.2]

Rawls selbst ist zwar im Zusammenhang mit der Frage der gerechten Sparrate auf die Fragen intertemporaler Gerechtigkeit eingegangen, aber aus mehreren Gründen helfen seine Überlegungen wenig dazu, die verschiedenen von Nordhaus betrachteten Umweltstrategien normativ zu bewerten: Für Gesellschaften, die ökonomischen Voraussetzungen zur Errichtung einer wohlgeordneten Gesellschaft besitzen, legen seine Überlegungen eine aufgrund der Anforderungen der Gerechtigkeit verlangte Sparrate von 0 nahe. Für Gesellschaften, die diese Voraussetzungen noch nicht besitzen, sind seine Ausführungen sehr vage; über die Vermutung, daß in frühen Perioden, wenn die Gesellschaften noch sehr arm sind, eine geringere Sparrate gefordert ist als in späteren, wenn die Gesellschaften schon reicher sind, gehen sie nicht hinaus.⁴⁵ Beide Aussagen lassen sich mit dem Nordhausschen Modell und seinen endogen bestimmten Sparraten nicht vereinbaren: Zum einen sparen darin auch reiche Gesellschaften noch in erheblichem Umfang, zum andern beginnen alle Gesellschaften mit hohen Sparraten, die mit der Zeit und damit bei wachsendem Wohlstand fallen.

Eine Möglichkeit, die Rawlsschen Überlegungen und das Modell zu versöhnen, ist es, die Sparraten des Modells konsequent als das Ergebnis von Verhaltensannahmen zu interpretieren und nicht normativ: Die Sparraten sind nicht die von der Gerechtigkeit geforderten, sondern liegen über diesen. Solche höheren Sparraten sind in der Rawlschen Theorie erlaubt, solange sie nicht mit anderen Anforderungen der Gerechtigkeit

⁴⁴ In den Exceltabellen, die das Modell begleiten, berechnet Nordhaus auch den Gegenwartsnutzen für die einzelnen Regionen, den er allerdings jeweils mit dem Pro-Kopf-Konsum in der ersten Periode gewichtet. Eine so gewichtete Summe der regionalen Wohlfahrtsfunktionen führt qualitativ zu denselben Ergebnissen wie Nordhaus' faktische Vorgehensweise. [S. 203]

⁴⁵ Eine exaktere Bestimmung des aufgrund des Rawlsschen Kriteriums für gerechte Sparraten geforderten gerechten intertemporalen Schemas von Sparraten bleibt ein Desiderat.

(insbesondere den zwei Prinzipien) in Konflikt geraten. Auch so tragen allerdings die Rawlsschen Überlegungen zur gerechten Sparrate wenig für den Zweck einer Bewertung der verschiedenen Umweltstrategien im RICE-Modell bei.

Es könnte naheliegend erscheinen, die unterschiedlichen Umweltstrategien auf der Basis eines internationalen Differenzprinzips zu beurteilen. [7.3.3] Rawls selbst lehnt eine solche Anwendung des Differenzprinzips auf internationale – wie auch auf intergenerationale – Fragestellungen ab.⁴⁶

Allerdings kann man fragen, ob nicht wichtige Teile der Rawlsschen Beschreibung von Gesellschaften heute nicht besser auf die Welt als Ganzes passen, insbesondere ob nicht die Welt als Ganzes als ein System sozialer Kooperation zu verstehen ist. Konsequenz wäre es dann, im Urzustand die Bewohner der ganzen Welt vertreten zu sehen, und aufgrund einer dem Rawlsschen Urzustand parallelen Logik würde daraus ein internationales Differenzprinzip folgen.⁴⁷

Aufgrund des intertemporalen Charakters der Fragestellung hier kann aber ein internationales Differenzprinzip – nur ergänzt durch ein einfaches Dominanzprinzip – keine beste Strategie bestimmen. Ohne die Strategie des Geoengineerings gibt es sieben Klimastrategien die zumindest einmal der ärmsten Region den höchsten Pro-Kopf-Konsum gewähren. Auch wenn man eine Strategie als Modellartefakt ausschließen kann und andere vielleicht mit der *Intuition*, daß sehr kurze und relativ kleine Vorteile wenig wiegen gegen längerfristige und größere, bleibt doch die Entscheidung offen zwischen der sogenannten Optimalstrategie und einer Variante des Kyotoprotokolls, die beide über eine Reihe von Perioden im Sinne des Differenzprinzips die besten sind. Rawls geht zwar davon aus, daß in vielen praktischen ethischen Entscheidungen ein Punkt erreicht wird, wo die Berufung auf die Intuition unausweichlich wird [Rawls 1975, 59, 354 f.], aber eine intuitive Begründung, weshalb eine der letzteren beiden Strategien die beste sein soll, die nicht sehr nahe an die von Rawls abgelehnte intertemporale Summierung von Periodenkonsum oder -nutzen herankommt, ist nicht offensichtlich.

Eine Möglichkeit in noch weitergehender Abweichung von Rawls zu einem definiten Ergebnis im gegebenen Fall zu kommen, wäre es, das Differenzprinzip auf die regionalen Wohlfahrtsfunktion der am schlechtesten gestellten Region anzuwenden. [7.3.4] Da es sich im gegebenen Fall dabei immer um die Region der ärmsten Länder handelt, ist nach der Bestimmung der am schlechtesten gestellten Region in jeder Periode aufgrund eines Rawlsschen (Einkommens-)Kriteriums kein weiterer interregionaler Nutzenvergleich notwendig. Das gilt aber nicht allgemein, und vor allem ist dieses Verfahren allen Einwänden ausgesetzt, die Rawls gegen die intertemporale Summierung eines abdiskontierten Nutzens vorbringt. Die Anwendung dieses Mischkriteriums würde die Variante des Kyotoprotokolls ohne Zertifikathandel als die beste erscheinen lassen. Das tatsächliche Protokoll würde immer noch als leicht besser als die „Optimalstrategie“ abschneiden.

Fragen der Gerechtigkeit zwischen den Völkern behandelt Rawls selbst in seiner

⁴⁶ Vgl. S. 211 und 214

⁴⁷ Ähnlich argumentieren auch Beitz [2002] und Pogge [1989, 246-259].

1 Einleitung

Theorie des Völkerrechts, vor allem in Rawls [1999d]. [7.3.5] In unserem Zusammenhang kommt es dabei zunächst auf die von Rawls formulierte Beistandspflicht an, in deren Rahmen „wohlgeordnete Gesellschaften“, d. h. (im Sinne Rawls') gerechte liberale Gesellschaften oder Gesellschaften, die zumindest die Menschenrechte, das Völkerrecht und eine Gemeinwohlvorstellung, die die grundlegenden Interessen aller berücksichtigt, respektieren („achtbare Gesellschaften“), die Pflicht haben, Gesellschaften, denen die Voraussetzungen zur Verwirklichung einer wohlgeordneten Gesellschaft fehlen („belastete Gesellschaften“) auch materiell solange zu unterstützen, bis diese Voraussetzungen verwirklicht sind. [S. 218 ff.]

Für Rawls ist es entscheidend, daß es einen *Endpunkt* dieser Beistandspflicht gibt. Wenn man die Rawlssche Theorie des Völkerrechts praktisch anwenden möchte, stellt sich dann natürlich die Frage, wann genau die Beistandspflicht erlischt.

Um mich einer Beantwortung dieser Frage anzunähern, schlage ich vor, den Endpunkt der *materiellen* Beistandspflicht als das Niveau des Bruttosozialprodukts pro Kopf des Jahres 1995 (auf das hin das RICE-Modell kalibriert ist) zu bestimmen, bei dem mindestens ein Land das Ideal einer gerechten liberalen Gesellschaft, wie von Rawls [1999d, 14, 50] beschrieben, wenigstens annähernd erfüllt. Aufgrund der von mir vorgeschlagenen Operationalisierung dieser Kriterien (die auch die Zugänglichkeit entsprechender Daten berücksichtigen muß) ergibt sich dafür – abhängig von Varianten der Operationalisierung – ein Korridor von 1900,- zu 3600,- Dollar.

Rawls [1999d, 105 f.] hat – allerdings ohne auf die Begründung näher einzugehen – historische Beispiele von Ländern angeführt, die wohl seines Erachtens die materiellen Bedingungen wohlgeordneter Gesellschaften erfüllt haben. Während das Beispiel Deutschlands der dreißiger Jahre des letzten Jahrhunderts gut zu dem angegebenen Rahmen paßt, ist es mehr als zweifelhaft, daß je ein Land mit dem Pro-Kopf-Einkommen Spaniens um 1500 oder selbst Frankreichs um 1820 die Kriterien einer gerechten liberalen Gesellschaft auch nur annähernd erfüllt hat. Jedenfalls sind Gesellschaften, die 1995 ein entsprechendes Bruttosozialprodukt hatten, weit davon entfernt.

Die Beistandspflicht ernstzunehmen, bedeutet wohl, das Ziel der Schaffung der Bedingungen einer wohlgeordneten Gesellschaft – abgewogen allerdings gegen die Interessen der früheren Generationen der belasteten Gesellschaften – möglichst früh zu erreichen. Wenn man, um einen konkreten Eindruck zu gewinnen, aus dem angegebenen Korridor den Wert von 2750 Dollar herausgreift, so ergibt sich, daß abgesehen von der Strategie des Geoengineering (die nichts anderes ist als das kostenlose Verschwindenlassen der Folgen der Emission von Treibhausgasen) dieses Ziel als erstes von einer Variante des Kyotoprotokolls erreicht wird, bei der keinerlei Zertifikatshandel stattfindet, dicht gefolgt vom tatsächlichen Protokoll.⁴⁸ Die sogenannte Optimalstrategie erreicht das Ziel fünf Monate später. Drei Tage früher erreicht die um zehn Jahre verschobene Optimalstrategie das Ziel. Diese Strategie ist die einzige, die einer früheren Generation der am schlechtesten gestellten Region ein höhere Bruttosozialprodukt pro Kopf gewährt als die des Kyotoprotokolls ohne Zertifikatshandel. Da es sich dabei aber

⁴⁸ Die genauen Daten der Zielerreichung sind unter der Voraussetzung eines exponentiellen Wachstums interpoliert. Größenordnungsmäßig ergibt eine lineare Interpolation keine wesentlichen Unterschiede.

nur um die erste Periode handelt und der Unterschied viel weniger als ein Cent beträgt, wird das die frühere Zielerreichung kaum aufwiegen.

Die Unterschiede im Zeitpunkt der Zielerreichung sind z. T. sehr gering. Dies ist zum einen ein Korrelat der Kombination der als relativ niedrig modellierten ökonomischen Folgen des Treibhauseffekts und der beachtlichen Wachstumsraten, und zum anderen der Tatsache, daß die von Nordhaus betrachteten Strategien nur ein geringes Maß von Umverteilung enthalten. [7.3.6] Bei nur relativ geringen Belastungen durch den Treibhauseffekt ist aber zu erwarten, daß die Zielerreichung der Beistandspflicht (wenigstens im Modell) im wesentlichen über Umverteilung zu beschleunigen ist. Es ist deshalb keine Überraschung, daß die Modellierung einer Strategie, die erhebliche finanzielle Transfers mit „Effizienz“ (in Sinne der Autoren) verbindet, in der Tat zu einem wesentlich früheren Termin der Zielerreichung (70 Jahre vor dem der Kyotostrategie) führt. Auch nach dem Kriterium eines egalitären Utilitarismus wäre diese Strategie bei weitem allen anderen betrachteten gegenüber überlegen.

Eine Illustration der Harmlosigkeit des Treibhauseffekts im RICE-Modell liefert der Blick auf die Entwicklung des Modells in der „mittleren langen Frist“, womit ich die Entwicklung in ein paar Hunderttausend von Jahren meine, also einen Zeitraum, über den wir heute wohl nicht mehr sinnvollerweise planen können. [7.3.7] Als Beispiel betrachte ich einen Zeitpfad, der sich ergibt, wenn bis in das Jahr 2285 keine Klimapolitik erfolgt, und von da an in jeder Region der Pro-Kopf-Konsum konstant gehalten wird: Im Jahr 1001985 läge die globale Durchschnittstemperatur 1,628 °C über der vorindustriellen Temperatur und in dieser Periode würde die Region der ärmsten Länder weniger als 2,29 % der Bruttoproduktion aufgrund des Treibhauseffekts verlieren.

Die Harmlosigkeit des Treibhauseffekts im RICE-Modell hängt zum einen ab von der Weise, in der die Autoren den Klimaprozeß modellieren. Da es uns hier um die Untersuchung der Ergebnisse unterschiedlicher normativer Annahmen geht, gehe ich auf die Modellierung des Klimaprozesses nicht weiter ein. Zum anderen hängt die Harmlosigkeit aber auch von der Weise ab, in der die Autoren den durch den Treibhauseffekt bewirkten Schaden modellieren, insbesondere auch davon, wie sie die verschiedenen Schadenskategorien *bewerten*.

Zwei dieser Kategorien greife ich besonders heraus, um sie genauer zu untersuchen: die Auswirkung des Treibhauseffekts auf die menschliche Gesundheit [7.3.8.1] und die Möglichkeit katastrophischer Entwicklungen [7.3.8.2].

Nordhaus u. Boyer [2000c] berücksichtigen die gesundheitlichen Folgen des Treibhauseffekts, indem sie auf der Basis von Murray u. Lopez [1996] schätzen, wieviele Lebensjahre aufgrund des Treibhauseffektes verloren gehen, und diese verlorenen Lebensjahre (YLLs) dann jeweils mit dem zweifachen Wert des jährlichen Pro-Kopf-Einkommens der einzelnen Regionen bewerten. [7.3.8.1] Mit einem egalitären ethischen Ansatz wie der Rawlsschen Theorie der Gerechtigkeit ist eine solche Vorgehensweise nicht vereinbar.

Rawls' Überlegungen zur medizinischen Versorgung im Rahmen seiner Theorie der Gerechtigkeit läßt sich nicht direkt auf die Frage der Bewertung des Verlustes von Lebensjahren übertragen: Er beschränkt sich auf die Möglichkeit *zeitlich begrenzter Phasen* von Krankheit im Rahmen *einer* Gesellschaft und *einer* Generation, und gibt der Wie-

1 Einleitung

derherstellung der Gesundheit (d. h. der Fähigkeit, als normal kooperierendes Mitglied der Gesellschaft ein ganzes Leben zu leben) einem der fairen Chancengleichheit parallelen Grad des Vorrangs. [Rawls 2003, 265-268]

Hier aber handelt es sich um eine Situation, in der der (theoretisch vermeidbare) vorzeitige Tod einer großen Zahl von Menschen praktisch nicht vermeidbar ist, und Ansprüche auf ein „volles Leben“ zwischen Menschen verschiedener Generationen und Gesellschaften abzuwägen sind.

Die Rawlssche Theorie der Gerechtigkeit wäre also um Kriterien zu ergänzen, die es ermöglichen, diese verschiedenen, jeweils elementaren Ansprüche gegeneinander abzuwägen: ein schwierige und ethisch herausfordernde Aufgabe, da es sich hierbei um eine Art weltweites und intergeneracionales Triageproblem handelt, bei dem sozusagen per definitionem berechnigte Ansprüche unerfüllt bleiben.

Auch wenn ich einen (unvollständigen) Vorschlag mache, wie eine Rawlssche Theorie der Gerechtigkeit ergänzbar wäre, so daß sie Antworten gibt auch auf Probleme, die mit dem Verlust von Lebensjahren in einem internationalen und intergenerationalen Rahmen zusammenhängen, geht es mir dabei doch mehr um eine Klärung der substantiellen ethischen Fragen, die zu beantworten sind, als um den Vorschlag selbst, den ich als eher vorläufig verstehe.

Die Grundidee ist, aufgrund der globalen Bedrohung für die menschliche Gesundheit neben der Beistandspflicht mit dem Ziel der Erreichung der Bedingungen einer wohlgeordneten Gesellschaft und dieser strikt nachgeordnet eine weitere Kooperationspflicht einzuführen, die auf die Verhinderung des frühzeitigen Todes aufgrund globaler Bedrohungen zielt. Die Nachordnung entspricht der Nachordnung der Chancengleichheit und Gesundheitsfürsorge im Fall einzelner Gesellschaften und setzt voraus, daß die Errichtung einer wohlgeordneten Gesellschaft als solches schon die Gesundheit der Mitglieder fördert. [Daniels u. a. 1999; Daniels 2001] Genauer schlage ich vor, dazu den Indikator der (nicht diskontierten, nicht altersgewichteten) YLLs zu verwenden, und jedenfalls dann von einer eindeutigen Verbesserung zu sprechen, wenn in allen Perioden sowohl die weltweite Summe der durch globale Bedrohungen verursachten YLLs als auch die YLLs pro Einwohner in der am schlechtesten gestellten Region abnimmt.⁴⁹

Die anspruchsvolleren Umweltstrategien können (wie zu erwarten) in erheblichen Umfang Leben retten. Dies gilt insbesondere für die anspruchsvollste Strategie, die der Beschränkung des globalen Temperaturanstiegs auf 1,5 °C: Gegenüber der sogenannten Optimalstrategie würden z.B. schon im Jahr 2105 6 963,3 Tausend Lebensjahre (pro Jahr) weniger verloren gehen. Im Jahr 2305 sogar schon 31 481 Tausend.

Meinem Vorschlag entsprechend würde allerdings dies nichts daran ändern, daß von den Strategien, die ich bis dahin betrachtet habe, die radikale Umverteilungsstrategie gewählt würde, die vergleichsweise wenig Lebensjahre retten würde (ab der sechsten Periode durchgängig weniger als die „Optimalstrategien“, aber in der Regel⁵⁰ mehr als

⁴⁹ Diese Dominanzregel dürfte kaum umstritten sein, solange man überhaupt die YLLs als sinnvolles Kriterium betrachtet. Sie gibt allerdings keine vollständige Ordnung an, insbesondere umgehe ich das schwierige Problem der intertemporalen Gewichtung von frühzeitigem Verlust des Lebens.

⁵⁰Für das tatsächliche Kyotoprotokoll in jeder Periode, für die übrigen Varianten in den meisten.

die verschiedenen Varianten des Kyotoprotokolls).

Eine der Rechtfertigungen für die Vorrangung der Erreichung der Bedingungen einer wohlgeordneten Gesellschaft war, daß zu erwarten ist, daß zum einen wohlgeordnete Gesellschaften ein Interesse an der Gesundheit ihrer Bürger haben, und zum anderen, daß schon einfach höherer Reichtum einer Gesellschaft tendenziell mit einer höheren Lebenserwartung verbunden ist. Der letzte Punkt läßt sich illustrieren, wenn man in einem ähnlichen Verfahren, wie Nordhaus u. Boyer [2000c] es verwenden, die YLLs schätzt, die als armutsbedingt gelten können: In der Tat werden bei Wahl der radikalen Umverteilungsstrategie in den frühen Perioden (bis etwa 2105) durch diese Strategie in hohem Maß armutsbedingte YLLs reduziert (sowohl global, wie im Blick auf die am schlechtesten gestellte Region). In späteren Perioden dominiert aber, auch was den gesellschaftlichen Reichtum angeht, der Umwelteffekt, und die armutsbedingten YLLs sind – auf einem insgesamt niedrigeren Niveau – für die radikalen Umweltstrategien am niedrigsten.

Seine normative Kraft entfaltet die ergänzte Forderung erst dann, wenn entweder im Rahmen der idealen Theorie⁵¹ davon ausgegangen werden kann, daß die reicheren Länder/Regionen ihre Verpflichtungen aus der Beistandspflicht vollständig erfüllen, oder wenn – im Rahmen der Theorie beschränkter Konformität – die Bereitschaft zu finanziellen Transfers zwar schon erschöpft ist, aber eine Bereitschaft zu einem Beitrag zum Umweltschutz noch vorhanden ist.⁵²

Trotz der Vorläufigkeit dieser Überlegungen zum Thema der gesundheitlichen Folgen des Treibhauseffekts in einer rawlsianischen Perspektive bleibt doch festzuhalten, daß die Frage der Auswirkungen des Klimawandels auf die menschliche Gesundheit entscheidend dafür sein wird, wie verschiedene Strategien des Umgangs mit dem Klimawandel zu beurteilen sind, daß in einer rawlsianischen Perspektive eine einfache Verrechnung des Gutes des menschlichen Lebens mit anderen Gütern kaum gerechtfertigt sein wird, und daß trotzdem in einer rawlsianischen Perspektive den Aufwendungen für Gesundheit sinnvolle Grenzen zu setzen sind.

Einer der Faktoren, die in der Diskussion um den Treibhauseffekt eine besondere Rolle spielen, ist die nicht auszuschließende Möglichkeit katastrophischer Entwicklungen. [7.3.8.2] Nordhaus u. Boyer [2000c] integrieren das Risiko katastrophischer Entwicklungen in ihren quantitativen Ansatz, indem sie ausgehend von einer Expertenbefragung sehr grob die Wahrscheinlichkeit eines Verlustes von 30 % des Weltbruttosozialprodukts für verschiedene globale Temperaturanstiege schätzen und den resultierenden Erwartungswert des Verlusts in ihre Schadensfunktion integrieren.

Zwar hat Rawls selbst meines Wissens nie die Frage schwer einschätzbarer, sehr großer globaler Umweltrisiken explizit behandelt. Doch handelt es sich hier – unter Voraussetzung seiner Kriterien [Rawls 2001, 98] – um einen fast paradigmatischen Fall für die Anwendung der Maximinregel der Entscheidung auf der Ebene des Rechts der Völker:

1. Die bedrohlichen „Singularitäten“ des Klimaprozesses sind engstens verbunden

⁵¹ Zum Begriff vgl. S. 132.

⁵² Eine durchaus als realistisch einzuschätzende Konstellation.

1 Einleitung

mit unserer Unwissenheit, Unwissenheit bezüglich vieler Details des Klimaprozesses, besonders wenn wir uns, was Intensität und Geschwindigkeit der Veränderungen angeht, außerhalb dessen begeben, was die Menschheit in „historischer“ Zeit bislang erlebt hat. Die menschenverursachte Klimaerwärmung ist ein „erstes“ Experiment, mit offenem Ausgang. In der Tat läßt sich also argumentieren, daß keine „verlässliche Basis“ für Wahrscheinlichkeitseinschätzungen vorhanden ist.

2. Wenn – wie wir im Rahmen der idealen Theorie erwarten dürfen – die Völkergemeinschaft ihre Beistandspflicht erfüllt und die Völker die Möglichkeit zur Errichtung einer wohlgeordneten Gesellschaft auch verwirklichen, dann ist *für die Völker* zumindest ein „recht befriedigendes“ Niveau erreicht.
3. Jedenfalls nicht ausschließbare „Albtraumszenarios“ , aber wohl auch das Szenario eines Einbruchs der wirtschaftlichen Produktion um 30% (wenn man in Rechnung stellt, was ein solcher Einbruch für die Stabilität sozialer, wirtschaftlicher und politischer Systeme bedeuten kann) können plausiblerweise als weit unter dem „recht befriedigenden“ Niveau einer verwirklichten wohlgeordneten Gesellschaft liegend und als „nicht tolerierbar“ angesehen werden.

Eine Anwendung der Maximinregel der Entscheidung würde dann verlangen, unter Voraussetzung der prioritären Realisierung der Bedingungen für die Verwirklichung wohlgeordneter Gesellschaften weltweit alles zu tun, um katastrophische Entwicklungen aufgrund des Treibhauseffekts zu verhindern.

2 Grenzen der Wohlfahrtsökonomie

2.1 Einleitung

Moderne Gesellschaften stehen immer wieder vor wirtschaftspolitischen Entscheidungen: Sollen bestimmte Steuern, Subventionen oder Zölle eingeführt, erhöht, erniedrigt oder abgeschafft werden? Soll man bestimmte Staatsausgaben erhöhen oder erniedrigen? Soll man bestimmten internationalen Verträgen beitreten oder nicht? Wie hoch dürfen die Schulden des Staates sind? Ist diese oder jene gesetzliche Regelung eines wirtschaftlichen Problems angemessener?

Dieser Liste eignet eine gewisse Beliebigkeit, und sie ließe sich natürlich fortsetzen oder präzisieren. Sie dient aber nur als Hintergrund einer vielleicht überraschenden Frage: Können die Wirtschaftswissenschaften helfen, solche Fragen zu beantworten?

Eine mögliche Antwort scheint mir unter Wirtschaftswissenschaftlern besonders verbreitet zu sein: Neben der positiven Ökonomik, die sich auf die Beschreibung und Erklärung wirtschaftlicher Tatsachen und Zusammenhänge beschränkt, gibt es einen zweiten, normativen Zweig der Wirtschaftswissenschaften (die Wohlfahrtsökonomie), der auf der Basis eines sehr harmlosen, von fast allen geteilten Werturteils (dem Paretoprinzip) normative Aussagen, Empfehlungen für die Wirtschaftspolitik ermöglicht. Diese normativen Aussagen, die auf dem Paretoprinzip aufbauen, betreffen die Allokation oder Effizienz und können zu anderen normativen Aussagen oder Empfehlungen, besonders solchen, die auf Theorien der Gerechtigkeit aufbauend Aussagen über die Verteilung machen, entweder in einem Verhältnis der Konkurrenz stehen (man spricht dann z. B. von einem Trade-off zwischen Effizienz und Gerechtigkeit), oder auch in einem harmonischeren Verhältnis: Dann wird etwa behauptet, jede beliebige Verteilung sei auf effiziente Weise erreichbar, deshalb solle man bestimmte Maßnahmen, die zu Effizienz führen, in jedem Fall verwirklichen.

Mit dieser Position möchte ich mich im ersten Teil meiner Arbeit auseinandersetzen und zeigen, daß auf der Basis eines „harmlos“ verstandenen Paretoprinzip sich so wenig sagen läßt, daß für praktische wirtschaftspolitische Entscheidungen nichts gewonnen ist (insbesondere gibt es keine wirtschaftspolitischen Maßnahmen, die man aus Effizienzgründen in jedem Fall ergreifen sollte), und daß nicht einmal sinnvoll von einem Trade-off zwischen auf der Basis des Paretoprinzips definierter Effizienz und Gerechtigkeit gesprochen werden kann.

Es seien aber zunächst noch ein paar Bemerkungen zu meiner Behauptung darüber erlaubt, was die zur Zeit vorherrschende Position zu normativen Fragen in der Ökonomie sei. Ganz klar ist, daß es eine ganze Reihe von Ökonomen gibt, die diese Position

nicht teilen. Sei es, daß sie im Sinn einer Weberschen Wertfreiheit¹ die Ökonomie von normativen Aussagen ganz frei sehen wollen, sei es auch, weil sie sich bewußt sind, daß für substantielle normative Aussagen erheblich mehr an Werturteilen erfordert ist, als sie das Paretoprinzip enthält. Kaum ein Lehrbuch, das das Paretoprinzip einführt, unterläßt es, dann auch gleich auf die Grenzen dieses Prinzips hinzuweisen, weil es z. B. keinen Vergleich zwischen zwei Zuständen zulasse, die von zwei Individuen umgekehrt geordnet werden. Angesichts dieser so klaren Hinweise auf die Grenzen des Paretoprinzips, sollte es eigentlich großes Erstaunen auslösen, zu welchem gewaltigem Gebäude sich die Wohlfahrtsökonomie inzwischen entwickelt hat, und vielleicht auch den Verdacht, daß da irgend etwas nicht stimmen kann. Meine Überzeugung ist, daß dieses Gebäude auf zwei Fundamenten ruht: zum einen auf weiteren, ergänzenden oder gar mit dem Paretoprinzip konkurrierenden Wertprämissen und zum anderen darauf, daß weite Teile dieses Gebäude keine klare Beziehung mehr zum Problem gesellschaftlicher, insbesondere wirtschaftspolitischer Entscheidungen haben. Natürlich lassen sich auf der Basis von irgendwelchen Definitionen, in denen Komparative wie „besser“, „überlegen“, oder Superlative wie „best“ oder „optimal“ vorkommen, beliebig viele Aussagen herleiten. Wenn man sich dabei aber nicht immer Rechenschaft darüber abgibt, was das noch mit Fragen der praktischen Entscheidung zu tun hat, d. h. der eigentlich normativen Frage, besteht die Gefahr, daß eine Maßnahme empfohlen wird, die zu Zuständen führt, die in irgendeinem Sinne „optimal“ sind, ohne daß man sich je Gedanken darüber gemacht hat, ob die Werturteile, die man zugibt, überhaupt ausreichen, um solche Zustände zu empfehlen, oder ob solche Zustände überhaupt erstrebenswert sind. Darum wird es also im folgenden immer wieder gehen: Sorgfältig zu bestimmen, welche normativen Urteile durch die eingestandenen normativen Voraussetzungen gedeckt sind.

Ökonomen, die erkannt haben, daß normative Aussagen in der Ökonomie nur möglich sind auf der Basis von Werturteilen, die über das Paretoprinzip weit hinausgehen, und die deshalb die Notwendigkeit erkannt haben, ethische Fragen zu klären, bevor man zu ökonomischen Empfehlungen gelangt, werden in den folgenden Ausführungen allenfalls noch das eine oder andere zusätzliche Argument dafür finden, wie wenig sich in der Tat in der Ökonomie ohne solche weitergehenden Werturteile normativ sagen läßt: für alle praktischen Zwecke der Wirtschaftspolitik nichts.

2.2 Das Paretoprinzip

2.2.1 Definitionen

Was ist das Paretoprinzip? Die meisten Ökonomen werden wahrscheinlich meinen, daß das eine sehr einfach zu beantwortende Frage ist. Tatsächlich herrscht aber in der Literatur z.T. eine gewisse Begriffsverwirrung, der ich dadurch begegnen möchte, daß ich drei Begriffe sorgfältig unterscheiden möchte: die Pareto relation, das Paretoprinzip und das Paretokriterium. Die Bezeichnungen, die ich für die drei Begriffe gewählt habe,

¹ Vgl. dazu Weber [1973b] und Weber [1973c].

sind dabei nicht so wichtig, obwohl ich meine, daß sie den jeweils angezielten Sachverhalt gut ausdrücken. Die sachliche Unterscheidung aber ist von größter Wichtigkeit.

Zunächst einmal ist es leicht und wenig kontrovers, die Pareto-Relation zu definieren:

Sei G die Menge aller möglichen, denkbaren gesellschaftlichen Zustände g . I sei die Menge der von diesen Zuständen betroffenen Individuen i . Wir nehmen weiterhin an, alle Individuen $i \in I$ besäßen eine vollständige, reflexive und transitive (Präferenz-) Ordnung \succeq_i über G , auf deren Grundlage die strikte Präferenz \succ_i und die Indifferenz in der herkömmlichen Weise definiert werden:

$$g \succ_i g' :\leftrightarrow g \succeq_i g' \wedge \neg g' \succeq_i g$$

$$g \sim_i g' :\leftrightarrow g \succeq_i g' \wedge g' \succeq_i g$$

Wir definieren dann die (schwache) Pareto-Relation \succeq_p :

$$g \succeq_p g' :\leftrightarrow \forall i \in I : g \succeq_i g'$$

Außerdem definieren wir die (strikte) Pareto-Relation \succ_p und die Paretoindifferenz \sim_p :

$$g \succ_p g' :\leftrightarrow g \succeq_p g' \wedge \neg g' \succeq_p g$$

$$g \sim_p g' :\leftrightarrow g \succeq_p g' \wedge g' \succeq_p g$$

Wichtig ist hierbei, daß damit wirklich nichts anderes als eine Relation definiert wurde, ohne daß damit irgendeine normative „Verpflichtung“ eingegangen wurde. Auch wenn wir uns darauf einigen, die Relation \succ_p als „... ist gegenüber ... pareto-besser“ zu lesen, ändert sich daran zunächst nichts: Wir hätten auch eine weniger wertende Sprache wählen können, und auch wer dieser Relation keinerlei normatives Gewicht beimißt, kann diese verwenden. (Auch wenn er eine weniger wertende Sprachregelung vorziehen würde.) Das ist der Grund, warum es mir wenig hilfreich erscheint, diese Relation selbst schon das „Paretoprinzip“, oder das „Paretokriterium“ zu nennen²: Sachlich wurde hier eine zunächst normativ völlig neutrale Relation definiert. Unter dem Paretoprinzip oder Paretokriterium dagegen verstehen die meisten Ökonomen schon ein normatives Prinzip, ein Werturteil, das aber auf weitgehendste Zustimmung hoffen könne.

Definitionen des Paretokriteriums oder Paretoprinzips sollten also versuchen, möglichst genau den Bezug dieser beiden auf das Problem der sozialen Entscheidung zu bestimmen. Es sollte klar sein, auf was sich jemand normativ verpflichtet, wenn er sich als einen Anhänger des Paretoprinzips oder des Paretokriteriums bekennt. Nur dann machen ja auch Aussagen wie die, daß diese beiden auf weitgehendste Zustimmung treffen, überhaupt Sinn.

Zwei weitere Aspekte sind bei dem Versuch der Definition von Paretokriterium und Paretoprinzip zu beachten. Das erste ist, daß die Pareto-Relation i.a. keine vollständige

² Wie z. B. Böventer u. Illing [1995, 256].

Ordnung über G darstellt. Immer dann, wenn zumindest zwei Individuen zwei gesellschaftliche Zustände g und g' verschieden ordnen, gilt weder $g \succeq_p g'$ noch $g' \succeq_p g$. Es wird deshalb auch allgemein angenommen, daß Paretokriterium und Paretoprinzip in entsprechenden Fällen nicht entscheidend sein können, also etwa keinen gesellschaftlichen Zustand als zu verwirklichenden herausgreifen können, wenn es gilt zwischen g und g' zu wählen. Allerdings sollte das Paretokriterium sehr wohl entscheidend sein, wenn z. B. über eine Menge E die Ordnungen aller Individuen übereinstimmen.

Das zweite ist, daß allgemein angenommen wird, daß Paretoprinzip und Paretokriterium „harmlos“ seien, und daß das darin ausgedrückte Werturteil von fast allen geteilt wird. Die Formulierung des Paretokriteriums und Paretoprinzips sollte deshalb diesen damit gemeinten Minimalkonsens (soweit es ihn jedenfalls unter Anhängern der paretianischen Wohlfahrtstheorie gibt) ausdrücken.

„Paretokriterium“ spielt schon in der Wortbildung darauf an, daß es sich hier um ein Entscheidungskriterium handeln soll. Ein Entscheidungskriterium liegt aber wohl sicher nur dann vor, wenn eine Beziehung angegeben wird zwischen z. B. einer Relation und dem Problem der (hier: sozialen) Wahl. Das versucht die folgende Definition eines Entscheidungskriteriums zu berücksichtigen.

Es sei gegeben eine Menge $E \subseteq G$ der erreichbaren gesellschaftlichen Zustände. Für eine beliebige zweistellige Relation \succeq über G definieren wir $B(\succeq, E)$, die Menge der bezogen auf \succeq besten Elemente aus E , wie folgt:

$$B(\succeq, E) := \{e \in E : \forall e' \in E : e \succeq e'\}$$

Wir nennen eine zweistellige Relation \succeq über G ein gesellschaftliches Entscheidungskriterium (und schreiben dann \succeq^e)³, wenn für alle E gilt:

Irgendein beliebiger der Zustände $b \in B(\succeq, E)$ ist zu verwirklichen.⁴

Unter dem Paretokriterium möchte ich im folgenden die Pareto-Relation \succeq_p verwendet oder akzeptiert als gesellschaftliches Entscheidungskriterium verstehen. Etwas formaler: \succeq_p^e , das Paretokriterium, ist eine zweistellige Relation für die gilt: (a) \succeq_p^e ist ein Entscheidungskriterium und (b) $\forall g, g' \in G : g \succeq_p^e g' \leftrightarrow g \succeq_p g'$.⁵

Ein paar Kommentare zu dieser Formulierung scheinen angebracht.

Ich habe mich hier auf zweistellige Relationen als Entscheidungskriterien beschränkt. Das ist eine klare Einschränkung, man kann sich z. B. auch nichtbinäre Entscheidungsregeln vorstellen.⁶ Diese Einschränkung ist aber in der vorliegenden Fragestellung in-

³ Diese Schreibweise könnte vielleicht etwas verwirrend sein, da ja die Relationen \succeq und \succeq^e jeweils dieselben sind. Im folgenden möge man deshalb das hochgestellte e immer einfach als eine Erinnerung daran nehmen, daß im entsprechenden Kontext die so indexierte Relation als Entscheidungskriterium betrachtet wird. (D. h. \succeq^e soll auch nicht als Abkürzung des Satzes „ \succeq ist ein Entscheidungskriterium“ verstanden werden.)

⁴ Das heißt natürlich „zu verwirklichen“ im Sinn der betrachteten normativen Auffassung.

Die Formulierung berücksichtigt die Möglichkeit, daß gemäß dem betrachteten Kriterium ein beliebiger von mehreren Zuständen verwirklicht werden soll.

Einige weitere Bemerkungen zur Formulierung siehe gleich im Folgenden.

⁵ Im Sinne der vorangehenden Fußnote drückt (b) aus, daß \succeq_p^e und \succeq_p dieselben Relationen sind. An (a) wird eben durch das hochgestellte e erinnert.

⁶ Einige Aspekte nicht-binäre Entscheidungsregeln werden diskutiert in Sen [1986, 1091 ff.].

sofern berechtigt, als die Pareto-Relation eben eine zweistellige Relation ist, und als in der Tradition der Wohlfahrtsökonomik, die uns hier beschäftigt, das Problem der sozialen Wahl im allgemeinen nach dem Vorbild der Haushaltstheorie im Rahmen der optimalen Entscheidung auf der Basis einer zweistelligen Präferenzrelation dargestellt wird.⁷ Weitere Forderungen werden an Entscheidungskriterien zunächst nicht gestellt: Insbesondere werden weder Transitivität noch Vollständigkeit verlangt.

Die von mir vorgeschlagene Definition des Paretokriteriums versucht vor allem seinen normativen Gehalt als ein Entscheidungskriterium möglichst präzise deutlich zu machen. Die für den normativen Handlungsbezug gewählte Formulierung („ist zu verwirklichen“) braucht dabei nicht auf die Goldwaage gelegt werden, hier wären auch andere Formulierungen (z. B. „ist anzustreben“, „ist zu wählen“) möglich. Deutlich werden sollte, zu was man sich normativ verpflichtet, wenn man das Paretokriterium akzeptiert. Wer die Pareto-Relation als Entscheidungskriterium akzeptiert, der akzeptiert dann auch (das ist die *Definition* eines Entscheidungskriteriums), daß ihn dieses Kriterium in bestimmten, genau definierten Situationen normativ, zum Handeln verpflichtet: dann nämlich, wenn die Menge B nicht leer ist.

Eine Alternative, die ich bewußt vermieden habe, da in einer solchen Formulierung der Handlungs- und Entscheidungsbezug zunächst wenig deutlich wird, wäre es, „ist gesellschaftlich wünschenswert“, „ist gesellschaftlich besser“ oder „ist gesellschaftlich vorzuziehen“ zu sagen. Sprachlich wären solche Formulierungen zu nahe an Formulierungen wie „ist pareto-besser“, die ja einfach das Vorliegen einer bestimmten Relation feststellen sollen. Es ist zunächst nicht selbstverständlich und eindeutig, welche normative Konsequenzen daraus gezogen werden sollen, daß ein Zustand g einem Zustand g' „gesellschaftlich vorzuziehen“ ist. Sachlich kann natürlich damit genau der normative Gehalt verbunden sein, wie in der hier gewählten Formulierung, doch ist dieser Gehalt eben nicht explizit gemacht.

Die Definition der Menge B stellt nicht sicher, daß diese Menge genau ein Element enthält. Die Menge kann leer sein (dazu gleich mehr), dann ist gemäß des gewählten Kriteriums nichts zu verwirklichen. Oder die Menge mag mehr als ein Element enthalten. Diese Elemente sind dann gemäß des gewählten Kriteriums indifferent und ein beliebiger der Zustände b ist zu verwirklichen. Das ist ganz parallel zu Formulierung individueller Entscheidungsprobleme, wo positive Indifferenz von Optima heißt, daß das Individuum einen beliebigen der indifferenten Zustände wählen soll bzw. wählen wird. Das ist klar zu unterscheiden von der Situation, daß einfach keine Ordnung zwischen bestimmten Elementen besteht (was dazu führen kann, daß B leer ist).

Zuletzt möchte ich noch eine Definition für das Paretoprinzip vorschlagen:

Schwaches Paretoprinzip (im folgenden P): Sei \succeq^e ein (oder das) gesellschaftliche Entscheidungskriterium, dann soll gelten:⁸

⁷ Außerdem sind viele der Argumente direkt auf Entscheidungsfunktionen übertragbar. Vgl. z. B. unten S. 107.

⁸ Neben den üblichen Regeln der Präzedenz folge ich der Konvention, daß Quantoren, wenn nicht Klammern anderes anzeigen, sich jeweils soweit erstrecken, bis die entsprechenden Variablen neu quantifiziert werden.

$$\forall g, g' \in G : (\forall i : g \succ_i g') \rightarrow g \succ^e g'$$

Starkes Paretoprinzip (im folgenden: PP): Sei \succ^e ein (oder das) gesellschaftliche Entscheidungskriterium, dann soll gelten:

$$\forall g, g' \in G : (g \succ_p g' \rightarrow g \succ^e g') \wedge (g \sim_p g' \rightarrow g \sim^e g')$$

Das Paretoprinzip in der gewählten Formulierung ist also eine Forderung an gesellschaftliche Entscheidungskriterien. Während das Paretokriterium ein solches Entscheidungskriterium ist (die Pareto-Relation als Entscheidungskriterium), verlangt das Paretoprinzip, daß Entscheidungskriterien bestimmten Anforderungen genügen, genauer daß sie die Pareto-Relation implementieren. Auch das Paretoprinzip hat also deutlich normativen Charakter, der dadurch hergestellt wird, daß es eine Forderung an Entscheidungskriterien stellt, also an Kriterien, die zu Handeln verpflichten können.

Die hier vorgeschlagenen Definitionen folgen sachlich im großen und ganzen der Tradition innerhalb der Sozialwahltheorie, wie sie etwa in Sen [1970] formuliert wird. Der hauptsächlichste Unterschied besteht in einer gewissen Pedanterie in dem Versuch, den normativen Bezug, den Handlungsbezug jeweils sehr klar zu machen. Daß diese Pedanterie gerechtfertigt ist, beweist ein Blick auf ein fast jedes Lehrbuch. Während in der Lehrbuchliteratur die Pareto-Relation (evtl. unter anderem Namen) im allgemeinen formal definiert wird, finden sich nur sehr vage Formulierungen dazu, welche Beziehung diese Relation zu gesellschaftlichen Entscheidungen haben sollte, also zu dem, was ich das Paretokriterium bzw. das Paretoprinzip genannt habe. Diese Vagheit ist besonders auch deshalb fragwürdig, weil sie fast immer von der Behauptung oder Insinuation begleitet wird, daß dieser normative Gehalt etwas sei, was man fast selbstverständlich teilen würde.

Um diese Behauptungen etwas konkreter zu machen, möchte ich als Beispiel die Ausführungen eines verbreiteten Lehrbuchs zitieren: „Consider two allocations x and x' . The allocation x' is said to **Pareto dominate** x if everyone prefers x' to x . If each individual prefers x' to x , it seems noncontroversial to assert that x' is ‘better’ than x and any projects that move us from x to x' should be undertaken. This is the **Pareto criterion**.“ [Varian 1992, 405] Zunächst wird in diesem Zitat (unproblematisch) eine Relation der strikten Pareto-Dominanz definiert. Im zweiten Satz werden dann zwei Dinge als nicht kontrovers behauptet: Daß man eine Allokation x , die eine Allokation x' im definierten Sinn strikt pareto-dominiert, „besser“ als die Allokation nennen könnte, und daß jedes Projekt, das von x zu x' führt, durchgeführt werden sollte. Diese letztere normative Aussage sei das Paretokriterium. Das Problem an dieser Formulierung ist, daß völlig unklar ist, was da nun (angeblich unkontrovers) normativ festgestellt wird: Daß eine Allokation x „besser“ ist als eine Allokation x' heißt solange normativ nichts, als nicht klar ist, wie aufgrund der Relation „besser“ Entscheidungen getroffen werden sollen. Vielleicht meint ja Varian mit „besser als“ etwas, was ich ein Entscheidungskriterium genannt habe. Aber jedenfalls macht er das nicht explizit, und die Fortsetzung des Satzes weist in eine etwas andere Richtung. Dieser zweite Teil des zweiten Satzes, in dem Varian ausführt, was das Paretokriterium sei, ist nämlich zumindest extrem ungenau.

Typisch ist dabei, daß er die normativen Konsequenzen der Relationen „besser“, bzw. „paretodominiert“ an einer binären Entscheidungssituation darstellt. Und man wird ihm dabei nur zustimmen können: Denn *wenn*⁹ das Paretokriterium wirklich irgendwo unproblematisch wäre, dann da, wo die gesellschaftliche Wahl stattfindet zwischen genau zwei Möglichkeiten (weitere Möglichkeiten gibt es nicht) und alle dieselbe dieser Möglichkeiten vorziehen. Aber wenn das das Paretokriterium *ist*, wenn also durch das Paretokriterium nur für eine solche Situation eine normative Empfehlung gegeben wird, dann ist die Aufgabe zu zeigen, daß dieses Kriterium normativ für alle praktischen Fragen vollkommen irrelevant ist, noch leichter zu erfüllen.¹⁰ Denn das wage ich mit Sicherheit zu sagen: In einer solchen Situation befindet sich eine Gesellschaft *nie*. Wenn das Paretokriterium und das Paretoprinzip überhaupt die Chance haben sollen, normativ relevant zu sein, dann müssen sie normative Konsequenzen der Pareto-Relation, bzw. Pareto-Dominanz für Entscheidungssituationen mit mehr als nur zwei Möglichkeiten formulieren.¹¹

Wenn aber mehr als zwei gesellschaftliche Zustände zur Wahl stehen, dann ist folgt aus der Tatsache, daß x besser ist als x' , noch lange nicht, daß man x' wählen soll. Die Umstände anzugeben, unter denen diese Folgerung gelten würde, wäre die Aufgabe einer sorgfältigeren Definition des Paretokriteriums bzw. des Paretoprinzips.

2.2.2 Die Irrelevanz des Paretoprinzips

So wie ich das Paretoprinzip und das Paretokriterium jetzt formuliert habe, schließt sich nahtlos und ohne alle Überraschung als Ergebnis eine Präzisierung der von mir in der Einleitung zum Kapitel formulierte These an: (1) I.d.R., genauer: für alle praktischen Zwecke der Wirtschaftspolitik immer, ist $\cap_i B(\succeq_i, E)$ leer. (2) Immer wenn $\cap_i B(\succeq_i, E)$ leer ist, (a) ist $B(\succeq_p^e, E)$ leer und (b) kann, solange wir über ein Entscheidungskriterium

⁹ Ob das Paretokriterium *tatsächlich* unproblematisch *ist* möchte ich hier nicht diskutieren. Aber siehe dazu Abschnitt 2.5.

¹⁰ Auch Varian gibt zu, daß Situationen in denen das Paretokriterium eine Entscheidung ermöglicht zumindest „selten“ sind. [Varian 1992, 405]

¹¹ Die Darstellung des Paretokriteriums im Rahmen einer rein binären Entscheidungssituation ist vor allem dann problematisch, wenn sie suggerieren sollte, daß das der Rahmen sei, in dem wirtschaftspolitische Maßnahmen ergriffen werden. Selbstverständlich sind im realen politischen Prozeß immer eine ganz große Anzahl verschiedener politischer Projekte möglich, und wann immer die Entscheidung im politischen Prozeß auf nur zwei Alternativen reduziert wird, stellt sich die Frage, woher diese zwei Alternativen eigentlich kommen. Man kann sich einmal am Ende eines Entscheidungsprozesses befinden, wo sich etwa im Parlament wirklich nur noch die Frage einer Zustimmung oder Ablehnung stellt (obwohl auch da der politische Prozeß immer weitergeht und neue Formen gesucht werden können, ein einmal abgelehnte Projekt wieder aufs Tapet zu bringen, oder ein beschlossenes Projekt faktisch zu modifizieren), aber dann sind in diesem Prozeß schon sehr viele Entscheidungen getroffen worden, hoffentlich mit wirtschaftspolitischem Sachverstand und nach ethischen Kriterien. Befindet man sich in einem früheren Zustand des Entscheidungsprozesses, dann kann es oft gerade die wichtigste und hilfreichste politische Aufgabe sein, weitere Alternativen überhaupt erst einmal bewußt zu machen. Deshalb mag die Darstellung der wirtschaftspolitischen Entscheidungssituation am Beispiel nur zweier Alternativen für eine lehrbuchmäßige Einführung hilfreich sein (ich halte sie allerdings eher für irreführend), als Basis von normativen Entscheidungen im realen wirtschaftspolitischen Fragen ist sie vollkommen ungeeignet.

2 Grenzen der Wohlfahrtsökonomie

\succeq^e nicht mehr wissen, als daß es das starke oder schwache Paretoprinzip erfüllt, i. d. R. nicht ausgeschlossen werden, daß $B(\succeq^e, E)$ leer ist.

(2)(a) ergibt sich einfach aus der Tatsache, daß $\cap_i B(\succeq_i, E) = B(\succeq_p^e, E)$ ¹², (2)(b) ergibt sich trivialerweise, da das Paretokriterium das Paretoprinzip erfüllt und $B(\succeq_p^e, E)$ i. d. R. leer sein wird.

Es ist natürlich unmöglich, die These (1) ähnlich formal und kurz zu begründen wie die These (2). These (1) ist eine empirisch-theoretische Behauptung, die prinzipiell durch ein Gegenbeispiel widerlegbar wäre: In der und der geschichtlichen Situation stand die und die Gesellschaft in folgender Entscheidungssituation, und unser bestes Modell von dieser Entscheidungssituation zeigt, daß da wirklich alle betroffenen Individuen denselben gesellschaftlichen Zustand als den besten vorgezogen hätten. Ich gehe davon aus, daß es ein solches Gegenbeispiel nicht gibt. Die Anforderung ist ja sehr hoch, daß *alle* Betroffenen Individuen unter *allen* vorliegenden Möglichkeiten dieselbe (oder dieselben) als die beste einordnen. Reale wirtschaftspolitische Entscheidungssituationen sind von Interessenkonflikten bestimmt, die nicht nur auf falsche/mangelnde Information der Betroffenen zurückgehen. Für den (nicht zu erwartenden) Sonderfall, daß für eine reale wirtschaftspolitische Entscheidung diese Bedingung erfüllt wäre, wäre eine normative Wirtschaftstheorie entbehrlich. Allenfalls bräuchte man eine positive Wirtschaftswissenschaft, die die Betroffenen über ihre „wahren“ Interessen aufklärt.¹³

Selbstverständlich ist es möglich, solche Situationen auf dem Papier zu konstruieren, und es mag auch Situationen mit einer sehr kleinen Zahl von Betroffenen geben, die aus irgendwelchen Gründen sich schon darauf geeinigt haben, nur eine kleine Zahl von Möglichkeiten zu betrachten. Aber das sind zum einen nicht Beispiele von wirtschaftspolitischen Entscheidungen (die regelmäßig viele Menschen betreffen), zum anderen verbergen sich hinter der Einigung darauf, nur wenige Alternativen zu betrachten, schon Entscheidungen, die dann nicht (wenigstens nicht sichtbar) aufgrund des Paretokriteriums getroffen worden wären: Daß es möglich ist, die Zahl der Alternativen realer Entscheidungssituationen aus irgendwelchen, nicht kontrollierten anderen Gründen soweit zu reduzieren, daß am Ende für das so reduzierte E gilt $B(\succeq_p^e, E) \neq \emptyset$, bezweifle ich nicht, ist aber kein Argument für die normative Relevanz des Paretoprinzips oder Paretokriteriums.

Ich bin nun allerdings zuversichtlich, daß diese meine Einschätzung realer wirtschaftspolitischer Entscheidungssituationen von der großen Mehrheit der Ökonomen geteilt wird, so daß der Einwand, es gebe doch relevante Entscheidungssituationen, in denen $B(\succeq_p^e, E)$ nicht leer ist, wohl kaum zu erwarten sein wird, und diese kurzen Bemerkungen zur These (1) genügen mögen.

¹² $g \in \cap_i B(\succeq_i, E) \leftrightarrow \forall i \in I \forall g' \in E: g \succeq_i g' \leftrightarrow \forall g' \in E: g \succeq_p g' \leftrightarrow \forall g' \in E: g \succeq_p^e g' \leftrightarrow g \in B(\succeq_p^e, E)$

¹³ Die wäre vor allem dann nötig, wenn die von einer Entscheidung Betroffenen von ihrer Interessenharmonie selbst nicht so ganz überzeugt wären. Solche Überlegungen passen allerdings kaum zu den Grundannahmen der heutigen dominierenden Wirtschaftswissenschaften, die ja gerade im Bereich der Wohlfahrtsökonomie noch fast immer von rationalen und voll informierten Individuen ausgeht. Wie die Instrumente der Wohlfahrtsökonomie auf Situationen mit nicht ganz rationale oder unzureichend informierte Individuen anzupassen wäre, ist noch immer eine wenig reflektierte Frage.

Die bisherigen Ausführungen haben im Grunde abschließend gezeigt, daß die Ausgangsthese in der Tat korrekt ist: Aufgrund des Paretoprinzips bzw. aufgrund des Paretokriteriums, also aufgrund der im allgemeinen eingestandenen und explizit vorausgesetzten Basis allokativer Aussagen, oder allgemeiner der modernen Wohlfahrtstheorie, lassen sich für alle praktischen Zwecke der Wirtschaftspolitik keine normativen Aussagen herleiten.

Wäre da nicht das gewaltige Gebäude der Wohlfahrtstheorie, wären da nicht immer wieder Behauptungen von Ökonomen, diese oder jene wirtschaftspolitische Maßnahme, sei aus Effizienzgründen empfohlen, könnte ich die Aufgabe, die ich mir vorgenommen habe, als erfüllt ansehen. Angesichts dieses Gebäudes, angesichts dieser Empfehlungen ist es aber notwendig, Rechenschaft darüber abzulegen, wie all das mit der behaupteten Irrelevanz des Paretokriteriums und -prinzips vereinbar ist. Im folgenden möchte ich einige Stücke aus der sogenannten Wohlfahrtstheorie herausgreifen, um entweder zu zeigen, daß sie keinen normativen Charakter tragen, oder daß sie weitere Wertprämissen voraussetzen. Zuvor werde ich noch kurz auf die Frage eingehen müssen, ob meine Definitionen (a) normativer Aussagen und (b) des Paretoprinzips, bzw. des Paretokriteriums nicht zu eng gefaßt sind.

Was die Frage (a) angeht, so sind ohne Zweifel weitere Definitionen des Begriffs einer „normativen Aussage“ möglich. Meine engere Definition ist ein Verwendungsvorschlag für den Zweck meiner Arbeit, mit dem einfach meine Fragestellung präzisiert wird: Wenn ich danach frage, ob und wie normative Aussagen in der Ökonomie möglich sind, frage ich danach, ob Aussagen möglich sind, die direkte Handlungsempfehlungen beinhalten.

Was die Frage (b) angeht, so scheinen mir meine Definitionen von Paretokriterium und Paretoprinzip deshalb nicht zu eng zu sein, weil sie die maximale normative Verpflichtung auf der Basis der Pareto-Relation ausdrücken, von der man noch behaupten könnte, sie sei ziemlich unumstritten. Mein Einwand gegen weitergehende Definitionen ist deshalb immer derselbe: Die mit ihnen verbundenen normativen Urteile sind nicht mehr derart, daß sie als „ziemlich“ unumstritten gelten könnten. Alle Definitionen des Paretoprinzips etwa, die gesellschaftliche Entscheidungen völlig unempfindlich gegen Verteilungsüberlegungen machen (indem sie etwa alle gemäß der Pareto-Relation nicht vergleichbaren Zustände als gleich gut deklarieren) oder die Umverteilungen sogar verbieten, scheiden deshalb aus, da sie keinesfalls als eine Art Minimalkonsens dienen können.

2.3 Paretianische Wohlfahrtstheorie

2.3.1 Die Irrelevanz der Paretoeffizienz

Dem Leser wird aufgefallen sein, daß ich bei meiner Darstellung des Paretoprinzips und des Paretokriteriums eine Menge nicht definiert haben, die sonst bei Einführungen in die Wohlfahrtstheorie gleich auf den ersten Seiten erscheint, und die auch Pareto selbst schon definiert hat: die Menge der paretooptimalen bzw. paretoeffizienten Zu-

stände, oder in der Sprache Paretos der Menge der Zustände maximaler Ophelimität.
In der Definition Paretos:

„We will say that the members of a collectivity enjoy maximum ophelimity in a certain position when it is impossible to find a way of moving from that position very slightly in such a manner that the ophelimity enjoyed by each of the individuals of that collectivity increases or decreases. That is to say, any small displacement in departing from that position necessarily has the effect of increasing the ophelimity which certain individuals enjoy, and decreasing that which others enjoy, of being agreeable to some, and disagreeable to others.“ [Pareto 1972, 261]

Oder formal:

$$P(E) := \{e \in E : \neg \exists e' \in E : e' \succ_p e\}$$

Ich habe diese Menge bislang nicht definiert, weil sie für normative Zwecke irrelevant ist, solange wir nicht über das Paretokriterium oder Paretoprinzip hinausgehen. Das ist eine einfache Konsequenz aus den bisherigen Überlegungen und ergibt sich insbesondere daraus, daß wir die Brücke von binären Relationen der gesellschaftlichen Präferenz (einer Präferenzordnung) zu normativen Aussagen über die Menge $B(\succeq^e, E)$ gebaut haben. Wenn wir die normative Frage beantworten wollen, dann suchen wir nach der Menge $B(\succeq^e, E)$, d. h. nach der Menge der besten Elemente aus E , der Menge der erreichbaren gesellschaftlichen Zustände.

Diese besondere Rolle der Menge der besten Elemente bei der Lösung von Entscheidungsproblemen ist in keiner Weise ein Sonderweg, den ich hier gehe. Es scheint einfach die natürliche Weise zu sein, um mit Hilfe einer Präferenzrelation zu Entscheidungen zu gelangen. Ganz selbstverständlich werden von Ökonomen deshalb auch individuelle Entscheidungsprobleme so gelöst. Und auch in der Sozialwahlliteratur ist das – soweit eine sorgfältige Beschäftigung mit Grundlagenfragen stattfindet – der gängige Weg.¹⁴

Wenn aber die normative Frage mit der Menge $B(\succeq^e, E)$ beantwortet wird, dann kann die Menge $P(E)$ also solche nichts dazu beitragen, das normative Problem zu lösen. Ohne weitere normative Annahmen führt die bloße Definition einer bestimmten Menge – wie wertbeladen ihre Bezeichnung auch immer sein mag – keinen Schritt weiter bei der Bestimmung dessen, was wir tun sollen. $P(E)$ kann also höchstens noch eine Hilfsfunktion bei der Suche nach $B(\succeq^e, E)$ haben.

Zum einen könnte $P(E)$ in konkreten Modellen bei der (mathematischen) Bestimmung von $B(\succeq^e, E)$ helfen. In vielen solchen Modellen läßt sich $P(E)$ relativ leicht und elegant mathematisch charakterisieren und deshalb kann unter Umständen die Bestimmung von $B(\succeq^e, E)$ erleichtert werden, wenn man von $P(E)$ statt von E direkt ausgeht. Dieser Weg über $P(E)$ ist möglich, weil unter Voraussetzung des Paretoprinzips (PP) $B(\succeq^e, E) \subseteq P(E)$.¹⁵

¹⁴ Vgl. Sen [1970], Sen [1986], Kern u. Nida-Rümelin [1994]

¹⁵ Siehe gleich unten.

Zum anderen kann unter bestimmten Umständen aufgrund der Beziehungen zwischen $P(E)$ und $B(\succeq^e, E)$ die Antwort auf die normative Frage allein mit Hilfe von $P(E)$ gegeben werden. Welches diese Umstände sind, möchte ich im folgenden genauer untersuchen. Die zusammenfassende Antwort sei aber schon vorausgeschickt: Dann, wenn aufgrund von These (1) die Antwort auf die Frage der sozialen Wahl sein wird: „Wir wissen nicht, was wir tun sollen.“

Genauer kann es im gewählten Rahmen auf die normative Frage zwei mögliche Antworten geben:

1. Aufgrund des gegebenen Kriteriums ist nicht entscheidbar, welcher Zustand zu verwirklichen ist.
2. Aufgrund des gegebenen Kriteriums ist ein (beliebiger) Zustand aus einer Menge X zu verwirklichen (anzustreben o.ä.).

Beide Antworten können wir jeweils mit Hilfe der Menge $B(\succeq^e, E)$ geben: 1) wenn die Menge $B(\succeq^e, E)$ leer ist, oder wenn wir bei den gegebenen Voraussetzungen ihre Elemente nicht bestimmen können. 2) Wenn die Menge $B(\succeq^e, E)$ nicht leer ist und wir ihre Elemente bestimmen können. Dann ist einfach $X = B(\succeq^e, E)$. Das gilt ganz allgemein, solange wir uns im Rahmen der Sozialwahl auf der Basis von zweistelligen Präferenzrelationen („Entscheidungskriterien“) bewegen, also auch dann, wenn wir das Paretoprinzip nicht akzeptieren.

Am deutlichsten wird die behauptete Irrelevanz von $P(E)$ für den Fall $\succeq^e \neq \succeq_p^e$.

Es bedarf zunächst wohl keiner weiteren Ausführungen, daß $P(E)$ dann normativ irrelevant ist, wenn $\succeq^e \neq \succeq_p^e$ und das Paretoprinzip nicht gilt.

Aber selbst wenn das (starke) Paretoprinzip gilt, können wir im allgemeinen mit $P(E)$ wenig anfangen. Intuitiv ist das klar: Wenn wir unsere Entscheidungen auf der Basis eines Kriteriums fällen, das über das Paretokriterium hinausgeht, dann können wir natürlich im allgemeinen weder die Frage, ob unser Kriterium ausreicht, zu einer Entscheidung zu kommen, also ob $B(\succeq^e, E) = \emptyset$, noch die Frage, wie denn $B(\succeq^e, E)$ aussieht, also welchen Zustand wir verwirklichen sollen, auf der Basis von $P(E)$ allein beantworten.

Formal wird die Frage nach der normativen Relevanz von $P(E)$ durch die Betrachtung der Beziehungen zwischen $P(E)$ und $B(\succeq^e, E)$ beantwortet. Unter Voraussetzung des Paretoprinzips (PP) gilt $B(\succeq^e, E) \subseteq P(E)$:

$$e \in B(\succeq^e, E) \wedge PP \rightarrow (\forall e' \in E : e \succeq^e e') \wedge PP \rightarrow (\neg \exists e' \in E : e' \succ^e e) \wedge PP \\ \rightarrow (\neg \exists e' \in E : e' \succ_p e) \rightarrow e \in P(E) \quad (2.1)$$

Das ist aber auch so ziemlich alles, was sich im allgemeinen über die Beziehung von $P(E)$ und $B(\succeq^e, E)$ sagen läßt. In der Regel (aber nicht notwendigerweise) wird dabei gelten, daß $B(\succeq^e, E) \neq P(E)$, d. h. $B(\succeq^e, E)$ wird eine echte Teilmenge von $P(E)$ sein.

2 Grenzen der Wohlfahrtsökonomie

Man kann versuchen, diese Beziehung soweit als möglich dafür auszuschlachten, $P(E)$ irgendeine Funktion bei der Suche nach $B(\succeq^e, E)$ zu geben. Zum einen gilt aufgrund dieser Beziehung: $P(E) = \emptyset \rightarrow B(\succeq^e, E) = \emptyset$. Wenn also $P(E)$ leer ist, ist a fortiori $B(\succeq^e, E)$ leer, oder mit anderen Worten: $P(E) = \emptyset$ ist eine hinreichende (nicht notwendige) Bedingung dafür, daß sich die normative Frage (unter Voraussetzung von PP) nicht beantworten läßt. Das ist nun kein besonders aufregendes Ergebnis. $P(E) = \emptyset$ dürfte vor allem bedeuten, daß man das gesellschaftliche Entscheidungsproblem auf einem recht fundamentalen Niveau falsch angesetzt hat, indem man etwa den entscheidungslogischen Kalkül auf eine Menge E anzuwenden versucht hat, die nicht kompakt ist.

Zum anderen gilt:

$$(\forall e \in E : (e \in P(E) \vee \exists e' \in E : e' \in P(E) \wedge e' \succ_p e) \wedge \forall e, e' \in P(E) : e \sim_p e') \rightarrow B(\succeq^e, E) = P(E) \quad (2.2)$$

Beweis. Für leere E gilt die Konklusion trivialerweise. Für nicht leere Mengen E folgt aus dem ersten Teil des Antezedens der Implikation, daß $P(E) \neq \emptyset$. Sei $e^* \in P(E)$. Aufgrund des Antezedens muß dann auch gelten: $\forall e \in E : (e^* \sim_p e \vee \exists e' \in E : e^* \sim_p e' \wedge e' \succ_p e)$, oder wegen der Transitivität von \sim_p und \succ_p : $\forall e \in E : (e^* \sim_p e \vee e^* \succ_p e)$, d. h. $\forall e \in E : (e^* \succeq_p e)$. Unter PP folgt daraus $\forall e \in E : (e^* \succeq^e e)$, bzw. $e^* \in B(\succeq^e, E)$. Zusammen mit $B(\succeq^e, E) \subseteq P(E)$ folgt daraus die behauptete Identität. \square

Auch das ist nun kein besonders aufregendes Ergebnis. Wie wir gleich sehen werden, ist das Antezedens für nicht leere Mengen E hinreichende und notwendige Bedingung dafür, daß $B(\succeq_p, E) = \bigcap_i B(\succeq_i, E) \neq \emptyset$, das heißt dafür, daß die Präferenzordnungen soweit übereinstimmen, wie wir das für alle praktischen Zwecke der Wirtschaftspolitik ausschließen können.

Es sollte keine Überraschung sein, daß $P(E)$ im allgemeinen nicht ausreicht, die normative Frage, die wir mit der Suche nach $B(\succeq^e, E)$ identifiziert haben, zu beantworten, wenn $\succeq^e \neq \succeq_p^e$. Interessanter könnte es sein, zu fragen, welche Rolle $P(E)$ spielt, wenn wir das Paretokriterium als Entscheidungskriterium benutzen wollen, d. h. wenn $\succeq^e = \succeq_p^e$. Selbstverständlich gilt auch hier wieder $B(\succeq_p^e, E) \subseteq P(E)$:

$$e \in B(\succeq_p^e, E) \rightarrow \forall e' \in E : e \succeq_p e' \rightarrow (\neg \exists e' \in E : e' \succ_p e) \rightarrow e \in P(E)$$

Außerdem gilt:

$$(\exists e, e^* \in P(E) : e \not\sim_p e^*) \rightarrow B(\succeq_p^e, E) = \emptyset \quad (2.3)$$

Beweis. Sei $e' \in B(\succeq_p^e, E)$ und $e, e^* \in P(E)$. Dann gilt wegen der Definition von $B(\succeq_p^e, E) : e' \succeq_p e \wedge e' \succeq_p e^*$. Da $e, e^* \in P(E)$ muß dann gelten: $e' \sim_p e \wedge e' \sim_p e^*$. Wegen der Transitivität von \sim_p folgt daraus: $e \sim_p e^*$ \square

Weiterhin gilt für nicht leere Mengen E :

$$\begin{aligned} \forall e \in E : (e \in P(E) \vee \exists e' \in E : e' \in P(E) \wedge e' \succ_p e) \wedge \forall e, e' \in P(E) : e \sim_p e' \\ \leftrightarrow B(\succeq_p^e, E) \neq \emptyset \quad (2.4) \end{aligned}$$

Beweis. Aus dem ersten Teil des Antezedens der Implikation folgt, daß $P(E) \neq \emptyset$. Da das Paretokriterium das Paretoprinzip erfüllt, gilt aufgrund des Antezedens, wie gezeigt, $B(\succeq_p^e, E) = P(E)$, zusammen also die Konklusion. Bleibt die Gegenrichtung zu zeigen. Sei $e^* \in B(\succeq_p^e, E)$, d. h. $\forall e \in E : e^* \succeq_p e$ oder $\forall e \in E : e^* \sim_p e \vee e^* \succ_p e$. Wegen der Transitivität von \succeq_p folgt aus $e^* \sim_p e : e \in P(E)$. Und da $e^* \in P(E)$ gilt also $\forall e \in E : (e \in P(E) \vee \exists e' \in E : e' \in P(E) \wedge e' \succ_p e)$. $B(\succeq_p^e, E) \neq \emptyset \rightarrow \forall e, e' \in P(E) : e \sim_p e'$ folgt unmittelbar aus 2.3. \square

Zuletzt gilt:

$$B(\succeq_p^e, E) \neq \emptyset \rightarrow B(\succeq_p^e, E) = P(E)$$

Beweis. Sei wieder $e^* \in B(\succeq_p^e, E)$, also $\forall e \in E : e^* \succeq_p e$. Sei $e' \in P(E)$. Dann gilt zusammengenommen $e^* \succeq_p e' \wedge \neg e^* \succ_p e'$, oder $e^* \sim_p e'$ und wegen der Transitivität von \sim_p und $\succeq_p : e' \in B(\succeq_p^e, E)$. \square

Für den Fall $\succeq^e = \succeq_p^e$ wäre es also möglich, die normative Frage allein aufgrund der Betrachtung von $P(E)$ zu beantworten. Allerdings wird wieder für alle praktischen Zwecke der Fall, wo in der Tat $B(\succeq_p^e, E) \neq \emptyset$, nie auftreten, so daß, wenn wir uns auf das Paretokriterium als Entscheidungskriterium beschränken, die Antwort regelmäßig sein wird, daß wir nicht wissen, was wir tun sollen.

2.3.2 Über Pareto hinaus?

Wenn in Einführungen in die Wohlfahrtstheorie dargelegt wird, wie wenig weit das Paretokriterium führt, wenn es darum geht auf seiner Grundlage wirtschaftspolitische Empfehlungen zu machen, so wird i. d. R. auf letzteren Sachverhalt angespielt: Wenn die Menge der paretoeffizienten Zustände mehr als ein Element besitzt, und diese Elemente nicht durch die Relation der Paretoindifferenz verbunden sind, dann gibt das Paretokriterium (genausowenig natürlich das Paretoprinzip) keine Hinweise darauf, welches der Elemente von $P(E)$ gewählt werden soll.¹⁶

Die Einsicht, daß $P(E)$ in aller Regel Elemente enthalten wird, die gemäß der Paretoindifferenz nicht vergleichbar sind, und daß deshalb die Kenntnis von $P(E)$ das normative Problem nicht löst, ist also in keiner Weise neu. Das wirklich Überraschende ist die Karriere, die $P(E)$ trotzdem in den Wirtschaftswissenschaften gemacht hat. Ein großer Teil

¹⁶ Diesen Fall gilt es klar von dem der Paretoindifferenz zu unterscheiden, wo das Paretokriterium zwischen den Elementen von $P(E)$ nicht diskriminieren könnte, aber da alle Elemente positiv als indifferent, gleich gut bestimmt würden, wäre es einfach gleichgültig, welches Element zu wählen wäre.

der sogenannten Wohlfahrtstheorie besteht ja in nichts anderem, als die Elemente dieser Menge $P(E)$ zu charakterisieren und eventuell zu anderen Mengen in Beziehung zu setzen. Wenn die bisherigen Überlegungen korrekt sind, insbesondere also eine positive Antwort der normativen Frage in der Angabe einer nichtleeren Menge $B(\succeq^e, E)$ besteht, $B(\succeq_p^e, E)$ aber für alle praktischen Fragen gleich der leeren Menge ist, dann kann uns die Kenntnis, das genauere Studium und die mathematische Charakterisierung von $P(E)$ *allein* (also nicht als Zwischenschritt zur Herleitung von $B(\succeq^e, E)$) für unsere normative Fragestellung immer wieder neu nur zeigen, was wir schon längst wissen: daß das Paretokriterium nicht ausreicht, das normative Problem zu lösen.

Meine bisherigen Überlegungen haben zwei Positionen vollkommen gleichberechtigt behandelt: Eine Position, die sich allein auf das Paretokriterium beschränken möchte und eine Position, die Entscheidungskriterien benutzt, die über das Paretokriterium hinausgehen, indem sie die Teilordnung der Pareto-Relation ergänzen. Diese Gleichbehandlung der genannten zwei Positionen ist kein Problem für den, der über das Paretoprinzip hinausgehende ethische Urteile für möglich und sinnvoll hält. Er zieht aus den obigen Überlegungen die einfache Konsequenz, solche weitergehende Urteile zu suchen und das Interesse an Aussagen über paretoeffiziente gesellschaftliche Zustände zu verlieren. Eine unter Ökonomen ziemlich weit verbreitete Überzeugung will es aber, daß das über das Paretokriterium hinausgehende normative Aussagen zumindest äußerst problematisch, wenn nicht überhaupt unwissenschaftlich und deshalb in der Wohlfahrtsökonomie zu unterlassen wäre.¹⁷

Wenn man sich deshalb auf den radikal skeptischen Standpunkt stellen wollte, der außer dem Paretokriterium überhaupt keine anderen Entscheidungskriterien auch nur *zulassen* möchte (weil diese nicht wissenschaftlich seien, unüberprüfbare Wertprämissen und den interpersonellen Vergleich von Nutzen verlangen, reine Geschmacksache seien etc.), dann kommt, das zeigen die obigen Überlegungen, das ganze Unternehmen der Wohlfahrtsökonomie in der Tat in gewisser Weise mit der Betrachtung der Menge $P(E)$ aus. Allerdings nur, um immer wieder erkennen, daß wir nicht wissen, was wir tun sollen, da nicht alle Elemente von $P(E)$ paretoindifferent sind: Eine Form des Eingeständnisses der praktischen Irrelevanz der Wohlfahrtsökonomie.

Als eine konsequente Weiterführung dieses skeptischen Standpunktes könnte es vielleicht erscheinen, den gordischen Knoten dadurch zu zerschlagen, daß man es nur noch auf die „Effizienz“ ankommen läßt. Das Paretokriterium, so wie es oben formuliert habe, führt i. d. R. zu keiner Entscheidung. Betrachten wir dann doch einfach alle Elemente von $P(E)$ als gleich gut und verwirklichen irgendeines dieser Elemente, im Wissen, damit eine (pareto-) „effiziente“ Lösung gefunden zu haben.¹⁸

¹⁷ Kritisch dazu im folgenden Abschnitt 2.5 und Kapitel 4.

¹⁸ Statt diese Intuition im Rahmen der hier zugrundegelegten Theorie kollektiver Entscheidung in der Form des „Paretoextensionskriteriums“ darzustellen, könnte man sie auch in der Form einer Alternative zu dieser Theorie darzustellen versuchen: Bei kollektiven Entscheidungen suchen wir nicht die Menge der besten gesellschaftlichen Zustände, sondern eben die Menge aller der Zustände, denen gegenüber nicht ein anderer vorgezogen wird, und aus der dann ein beliebiges Element zu verwirklichen ist. Im Falle des so neu interpretierten „Paretokriteriums“ also suchen wir $P(E)$. Es bedarf keiner weiteren Ausführungen, daß in dieser Interpretation das Paretokriterium sich keiner allgemeiner

Formal kann diese „Lösung“ des normativen Problems auf der Basis des Paretoextensionskriteriums dargestellt werden:¹⁹ [Vgl. Sen 1970, 74 ff.]

$$\forall g, g' \in G : g \succeq^e g' \leftrightarrow \neg g' \succ_p g$$

Daraus ergibt sich unmittelbar $B(\succeq^e, E) = P(E)$:

$$e \in B(\succeq^e, E) \leftrightarrow \forall e' \in E : e \succeq^e e' \leftrightarrow \forall e' \in E : \neg e' \succ_p e \leftrightarrow (\neg \exists e' e' \succ_p e) \leftrightarrow e \in P(E)$$

Dieses Paretoextensionskriterium ist aber in Wirklichkeit keine Weiterführung eines ethischen Skeptizismus. Es versucht einen gewaltigen Sprung vom normativen Nichtwissen hin zum Wissen – allein auf der Basis des Nichtwissens. Es ist ethisch eher maximalistisch, da es eine vollständige Präferenzrelation über jede Menge E definiert und deshalb z. B. bei endlichen Mengen immer zu nicht leeren Mengen $P(E)$ und $B(\succeq^e, E)$ führt. Das Paretoextensionskriterium beschränkt sich nicht auf das Paretokriterium, sondern ergänzt es, weitet es aus. Es ist eines der vielen möglichen paretoinklusiven Entscheidungskriterien, ein Kriterium, das über das Paretokriterium hinaus Wertprämissen verwendet.

Diese Wertprämissen unterliegen wie alle ethischen Überzeugungen der Möglichkeit der ethischen Kritik, die in diesem Fall sehr leicht fällt. Das Paretoextensionskriterium widerspricht den elementarsten ethischen Grundeinsichten vieler Menschen.²⁰ Jeder Mensch, für den Verteilungsgesichtspunkte auch nur eine minimalste Bedeutung haben, wird es ablehnen. Auch wenn es schwierig sein sollte, Verteilungsgesichtspunkte in überzeugender Weise in eine normative Wirtschaftswissenschaft zu integrieren: Ein Zustand in dem A im Überfluß lebt, und B verhungert, *ist* nicht gleich gut, wie ein Zustand in dem der Reichtum von A auf beide so verteilt ist, daß beide gut leben können. Es ist mir deshalb auch kein Ethiker und kein Ökonom bekannt, der dieses Kriterium als Entscheidungskriterium *explizit* verteidigen würde.

Umso beunruhigender ist es, daß der Eindruck entstehen kann, das Paretoextensionskriterium implizit einem großen Teil der modernen Wohlfahrtsökonomie zugrunde zu liegen scheint.²¹ [Sen 1970, 197] Die extensive Beschäftigung mit $P(E)$, mit allokativer Effizienz, den Bedingungen, die Elemente von $P(E)$ erfüllen, die Meinung, damit

Akzeptanz erfreuen dürfte. (Die kritischen Anmerkungen zum Paretoextensionskriterium gleich im folgenden gelten auch hier, da es sich ja um nichts anderes als das etwas anders formulierte Paretoextensionskriterium handeln würde.) Im Sinne der Schlußbemerkungen des Abschnitts 2.2.2 wäre das also auch keine akzeptable Interpretation des Paretokriteriums.

¹⁹ Das Paretoextensionskriterium erklärt nicht nur die paretoeffizienten Zustände für indifferent. Es stellt aber die einzige Möglichkeit dar, sicherzustellen, daß für jede mögliche Erreichbarkeitsmenge alle paretooptimalen Elemente indifferent sind, ohne die Bewertung eines Zustandes von der Erreichbarkeitsmenge abhängig zu machen.

²⁰ Diese Einsicht wurde von Sen [1970, 22] prägnant formuliert: „In short, a society or an economy can be Pareto-optimal and still be perfectly disgusting.“ Vgl. auch Sen [1970, 69, 197]. In welchem Sinn und in welchen Grenzen meine Aussage im Text ein überzeugendes ethisches *Argument* sein kann, vgl. Abschnitt 5.2.

²¹ Anders als Sen [1970] vermute ich allerdings, daß nicht das – konsistente – Paretoextensionskriterium, sondern vielmehr eine Idee, die die Nähe zur Paretogrenze als ein Maß für Effizienz sieht, intuitiv

etwas normativ relevantes zu tun, ist bei Zugrundelegung des Paretoextensionskriteriums vollkommen angemessen und korrekt. Die Einhelligkeit mit der ein großer Teil der ökonomische Profession „Effizienz“ zum Kriterium guter Wirtschaftspolitik erhoben hat, ohne evtl. „effizienzverbessernde“ Maßnahmen unter die notwendige Bedingung der Umverteilung zu stellen, erscheint sinnvoll, wenn man davon ausgeht, daß diese Ökonomen bei ihren wirtschaftspolitischen Empfehlungen das Paretoextensionskriterium benutzt. Unter den Voraussetzungen des Paretoextensionskriterium ist Paretoeffizienz hinreichender Grund, einen Zustand anzustreben, was unter den Voraussetzungen des Paretokriteriums oder -prinzips keineswegs der Fall ist.

Die Unterstellung, Ökonomen seien Anhänger des Paretoextensionskriteriums, kann ungerecht wirken.²² Und in der Tat kann die extensive Beschäftigung mit $P(E)$ auch dann sinnvoll erscheinen, wenn man eine modifizierte Form des Paretoextensionskriteriums unterstellt, die den oft recht vagen Erklärungen des Paretoprinzips zugrunde zu liegen scheint. Dort ist manchmal davon die Rede, daß nach dem Paretoprinzip bestimmte ökonomische Maßnahmen oder Projekte dann ergriffen werden sollte, wenn sie zu Paretoverbesserungen führen.²³ Hier wird also der verwirklichte gesellschaftliche Zustand g als Grundlage genommen und statt $P(E)$ wird $P(E, g) = \{e \in E : e \in P(E) \wedge e \succ_p g\} \subseteq P(E)$ betrachtet.

Selbstverständlich wird auch $P(E, g)$ i. d. R. mehrere Elemente enthalten, die nicht paretoindifferent zueinander sind, sodaß es notwendig wird, diese Elemente als indifferent zu erklären, wenn man zu normativen Ergebnissen der Art kommen möchte, daß ein beliebiges dieser Elemente zu verwirklichen sei. Auch hier werden dann also über das Paretoprinzip hinaus weitere Wertprämissen eingeführt, die wieder auf eine etwas andere Art Verteilungsüberlegungen ausschließen, indem zwar nicht alle Elemente $P(E)$, sondern nur die Elemente von $P(E, g)$ als gleich gut eingestuft werden. Vor allem aber werden bei Verwendung dieses Kriteriums Umverteilungen, d. h. die Verwirklichung eines Zustands, der von g aus für einige Individuen eine Verschlechterung bedeutet, von vornherein ausgeschlossen.

Dieses modifizierte Paretoextensionskriterium gibt dem verwirklichten sozialen Zustand g eine große normative Bedeutung, die zu rechtfertigen wäre. (Jemand, der den verwirklichten Zustand als extrem ungerecht empfindet, wird kaum bereit sein, diesem

der Effizienzvorstellung zugrunde liegt, da das Paretoextensionskriterium es z. B. nicht erlaubt, von Effizienzverlusten zu sprechen, wenn man sich zwar von der Paretogrenze entfernt, aber zumindest ein Individuum besser gestellt wird. Mehr dazu s. u. Abschnitt 3.4.

²² Außer mit der gleich folgenden Variante des Paretoextensionskriteriums könnten sich Ökonomen noch mit der Behauptung verteidigen wollen, daß die Paretoextensionsregel „ökonomisches Denken“ ausdrücke: Das was ein knallharter Ökonom sagen würde, und was dann eben durch andere Gesichtspunkte eventuell noch korrigiert werden müßte. Ein Ökonom als Ökonom wäre zwischen gemäß der ParetoRelation nicht vergleichbaren Zuständen (positiv!) indifferent. Als Staatsbürger, Privatmann etc. brächte er dann allerdings evtl. noch andere Kriterien mit ins Spiel. Das Paretoextensionskriterium wäre dann aber explizit *kein* Entscheidungskriterium und müßte durch weitere Überlegungen, wie z. B. Gerechtigkeitsüberlegungen ergänzt bzw. korrigiert werden. Bevor eine solche Ergänzung expliziert wird, sind die normativen Konsequenzen der Übernahme des Paretoextensionskriterium unbestimmt, diese deshalb problematisch. Die Frage ist auch, *weshalb* man „als“ Ökonom eine Wertordnung vertreten sollte, die man als Bürger etc. ablehnt.

²³ So etwa Varian [1992, 405]

Zustand eine so große normative Bedeutung beizumessen.)²⁴

Aber auch die Relevanz dieses modifizierten Paretoextensionskriteriums kann bezweifelt werden: Zwar wird man für Probleme, die eine kleine, überschaubare Personengruppe betreffen, vermuten können, daß Situationen, wie sie dieses Kriterium voraussetzt, immer wieder auftreten. Es handelt sich dabei um paradigmatische Situationen der Verhandlungstheorie. Für ganze Gesellschaften betreffende wirtschaftspolitische Fragen scheint es aber kaum Situationen zu geben, in denen tatsächlich Paretoverbesserungen möglich sind: Den Konflikt um wirtschaftspolitische Maßnahmen gibt es ja gerade auch deshalb, weil es regelmäßig Verlierer und Gewinner gibt.

Ich habe mit der Darstellung eines normativen Pessimismus begonnen, der meint, über das Paretokriterium hinaus gebe es überhaupt kein Kriterium, auf dessen Basis sich normative Entscheidungen fällen lassen. Ich habe dann zwei Positionen dargestellt, die diesen Skeptizismus konsequent weiterzudenken scheinen, in Wirklichkeit aber das Paretoprinzip, das Paretokriterium durch ethisch starke und auch sicher sehr umstrittene Wertprämissen ergänzen, die in keiner Weise selbst aus dem Paretoprinzip folgen.

Weit verbreitet unter Ökonomen ist aber vielleicht eine etwas weniger skeptische Position, die sich folgendermaßen darstellen ließe: Ethische Forderungen, Vorstellungen über eine gerechte Verteilung, allgemein Wertvorstellungen sind sehr lobenswert. Wir alle haben solche Vorstellungen und das gehört sich auch so für einen Staatsbürger. Allerdings sind diese Vorstellungen leider sehr verschieden, ja oft einander geradezu widersprechend und es besteht keine Hoffnung, daß wir uns je auf eine solche Wertvorstellung einigen werden. Werte sind ein bißchen Geschmackssache, jedenfalls ist ein wissenschaftlicher Diskurs über Werte (nicht nur beschreibend, sondern begründend) unmöglich. Aber es gibt eine Wertvorstellung, die wir (fast) alle teilen: das Paretoprinzip. Können wir also nicht einmal versuchen, wie weit wir mit diesem Prinzip kommen? Klammern wir einmal alles aus, worauf wir uns nicht einigen können (insbesondere Verteilungsvorstellungen) und beschränken uns auf normative Aussagen (allokative Aussagen, Effizienzaussagen), die sich schon allein aus dem Paretoprinzip/-kriterium ergeben.

In den ersten Absätzen dieses Abschnittes habe ich schon gezeigt, was sich i. d. R. aus dem Paretoprinzip/-kriterium allein normativ folgern läßt: nichts. Eine Gesellschaft wird sich praktisch nie in einer Situation befinden, wo die Menge der möglichen gesellschaftlichen Zustände so beschaffen ist, daß allen Mitgliedern dieser Gesellschaft ein und derselbe Zustand als der beste erscheint. Und nur unter dieser Bedingung würde aus dem Paretoprinzip bzw. -kriterium normative Konsequenzen folgen.

Diese Aussage wird auch intuitiv leicht einsichtig, wenn Ökonomen uns zugestehen, über das Paretokriterium hinaus (privat aber ehrenvoll) noch Wertvorstellungen zu haben. Wer als Ökonom (und sei es nur privat) irgendwelche Vorstellungen über

²⁴ Sowenig wie das Paretoextensionskriterium wird deshalb dieses modifizierte Paretoextensionskriterium als eine Interpretation des Paretoprinzips, auf das wir uns angeblich alle einigen können, gelten können. Das wird auch von Autoren bestätigt, die eine Einigung auf über das Paretokriterium hinausgehende Werturteile die Verteilung betreffend für unmöglich halten. Vgl. z. B. Rose u. Sauernheimer [1992, 520].

Verteilungsgerechtigkeit unterhält, dem wird es i. d. R. nicht gleichgültig sein, welcher der paretooptimalen Zustände verwirklicht wird. Nach seiner Wertskala müssen sich die sogenannten paretooptimalen Zustände in keiner Weise vor anderen Zuständen auszeichnen. Es kann paretooptimale Zustände geben, die gemäß eines paretoinklusi-ven Entscheidungskriteriums zu den schlechtestmöglichen sozialen Zuständen gehören. Damit, daß der betrachtete Ökonom weiß, daß ein sozialer Zustand paretooptimal ist, weiß er nichts darüber, ob dieser Zustand gemäß seiner Wertskala in irgendeiner Weise wünschenswert ist. Solange wir uns nicht über unsere über das Paretoprinzip hinausgehenden Vorstellungen einigen, oder einen Weg finden, wie im politischen Pro-zeß auf der Basis dieser Vorstellungen Entscheidungen getroffen werden sollen, hilft uns das Paretoprinzip zu nichts.

Unter Voraussetzung bestimmter Entscheidungskriterien (z. B. das des Nutzenrawlsianismus²⁵) und bestimmten Kontinuitätsannahmen kann die Menge der Paretooptimalen Zustände alle Qualitätsstufen von gesellschaftlichen Zuständen enthalten: Vom schlechtesten bis zum besten.²⁶ $B(\succeq^e, E) \subseteq P(E)$ stellt z.B. sicher, daß der beste Zustand sich unter den Elementen von $P(E)$ befindet. Trotzdem sollte auch der eiserne Verfechter der normativen Relevanz von $P(E)$ Verständnis dafür haben, daß ein Nutzenrawlsianer von einer wirtschaftspolitischen Maßnahme nicht sehr beeindruckt sein wird, die sicherstellt, daß keine Qualitätsstufe durch diese Maßnahme ausgeschlossen wurde.

2.3.3 Samuelson als Retter?

Kritisch gegen meine Behauptung der normativen Irrelevanz von $P(E)$ könnte nun aber eingewendet werden, daß ich „normative Relevanz“ zu eng definiert habe. Normative Relevanz hatte eine Aussage für mich nur, wenn sie bestimmte gesellschaftliche Zustände als zu verwirklichen aussondert. Dagegen könnte man Samuelson [1950] folgend auf die normative Relevanz von gesellschaftlichen „Situationen“ verstanden als geeignet definierte Mengen von gesellschaftlichen Zuständen und „Maßnahmen“, die jene verwirklichen, verweisen.²⁷

Samuelson definiert das Konzept der potentiellen Paretosuperiorität (bzw. des potentiellen Realeinkommens): Samuelson geht dabei vom Begriff einer „Situation“ aus. Der Begriff wird von Samuelson nicht definiert, aber es handelt sich bei einer Situation um nichts anderes als um eine unter bestimmten, jeweils beschriebenen Voraussetzun-

²⁵ Damit bezeichne ich eine Position, die in wirtschaftswissenschaftlichen Literatur mit dem Namen Rawls in Verbindung gebracht wird, und die die gesellschaftlichen Zustände nach dem *Nutzenniveau* des am meisten benachteiligten *Individuums* ordnet. Das ist *nicht* die Position Rawls'. Dazu mehr in Kapitel 5.

²⁶ Ähnlich, aber nicht bezogen auf den Nutzenrawlsianismus, Sen [1986, 1107].

²⁷ Wie wir gleich sehen werden, sieht Samuelson selbst keinen Zugewinn an normativer Relevanz durch die Einführung der „Situationen“. Einige Autoren scheinen aber doch in dieser Einführung einen Ausweg zu sehen, der normative Aussagen möglich macht. So etwa Rose u. Sauernheimer [1992, 519 f.].

Während ich innerhalb dieses Abschnitts die Begriffe (gesellschaftliche) „Situation“ und „Zustand“ strikt im gerade eingeführten Sinn verwenden werde, folge ich außerhalb dieses Abschnitts oft einfach dem normalsprachlichen Gebrauch dieser Begriffe.

gen gegebene Menge von erreichbaren sozialen Zuständen. Jede Situation wird jeweils durch eine mit ihr verbundene Nutzenmöglichkeitsmenge, jede Nutzenmöglichkeitsmenge durch die mit ihr verbundene Nutzenmöglichkeitsgrenze charakterisiert. Wenn die Nutzenmöglichkeitsmenge der Situation B echte Teilmenge der Nutzenmöglichkeitsmenge einer Situation A ist, so daß erreichbare soziale Zustände in A existieren, die nicht Element von B sind und zumindest einigen Zuständen von B gegenüber paretoüberlegen sind, dann ist A potentiell pareto besser als B .²⁸ Sei nun A die Menge der erreichbaren sozialen Zustände, und damit die Nutzenmöglichkeitsgrenze von A gleich der Menge der pareto optimalen Zustände. Angenommen nun, wir befinden uns in der Situation B und genauer in einem Zustand, der von zumindest einem Zustand der Nutzenmöglichkeitsgrenze von A pareto dominiert wird, und es gebe eine wirtschaftspolitische Maßnahme, die uns von der Nutzenmöglichkeitsgrenze von B , auf die Nutzenmöglichkeitsgrenze von A bringt. Sollte man dann nicht diese Maßnahme ergreifen?

Müßte man diese Frage mit „Ja“ beantworten, so wäre in der Tat meine Definition von normativer Relevanz zu eng gewesen, da sie ja den Fall von Maßnahmen, die uns von einer Nutzenmöglichkeitsgrenze auf eine potentiell paretoüberlegene Grenze führen, ohne daß man die jeweils erreichten Zustände beachtet, nicht betrachtet.

Aber auch Samuelson selbst beantwortet die Frage mit einem klaren Nein. Er weist darauf hin, daß das Konzept der „potentiellen Paretoverbesserung“, wie alle Wohlfahrtökonomie für sich genommen (d. h. ohne Ergänzung durch weitere ethische Annahmen), uns „virtually none of the way“ weiterführt, der zur Lösung der Probleme normativer Art führt.

„The inadequacy for actual policy decisions – even in the most idealized, simplified world – of all of the discussed measures of ‘real income’ can be illustrated by numerous examples. Consider the very best case where we can establish the fact that situation 2 is *potentially* better than 1 (in the sense of having a uniformly farther-out utility-possibility function). Would a good fairy given the chance to throw a switch from 1 to 2 be able to justify doing so? Upon reflection we must, I am afraid, answer *no*. Potentialities are not actualities – and unless she can give a justification of her act that will satisfy all reasonably defined social welfare functions, she cannot know whether or not to pull the switch.

A few negative remarks are possible: for any ethical system with the property that an increase in one individual’s well-being is, other’s being equal a good thing – for all such systems a final optimum position must necessarily be on 2 and not on 1. That we can certainly say. But without going

²⁸ Samuelson entwickelt seine Begrifflichkeit an Beispielen. Die von mir hier gegebenen (verbalen) Definitionen sind allgemeiner, als die von Samuelson gegebenen Beispiele nahelegen. Insbesondere ist zu beachten, daß Samuelson zwar in Teilen seines Aufsatzes den Begriff Situation durchaus so verwendet, daß er mit meiner allgemeinen Definition zusammenpaßt. Allerdings reserviert er ihn dann für „Situationen“, die (negativ formuliert) nicht nur durch die Umverteilung von bestehenden Güterbündel, sondern (positiv ausgedrückt) durch die Produktionsmöglichkeiten charakterisiert sind. Dieser Einengung des Begriffs folge ich hier nicht.

into the realm of (modern, streamlined) 'old' welfare economics, we cannot say more or get conclusive advice on this problem of policy. The attempt to divide the problem into two parts so that one can say 'a change from 1 to 2 is *economically* desirable in the sense of objectively increasing production or wealth, whether or not the actual resulting situation will be ethically superior', only gets into a semantic snarl and glosses over the intrinsic difficulties of the problem."²⁹ [Samuelson 1950, 11 f.]

Um die Vorstellung, man könne durch Einklammern weitergehender Gerechtigkeitsvorstellungen auf der Basis allein des Paretoprinzips doch noch irgendwelche normativen Folgerungen über „Maßnahmen“ herleiten, nachvollziehbar zu machen, könnte man sich die Sache vielleicht so vorstellen, daß mit Hilfe des Paretokriteriums zunächst einmal eine Art Vorauswahl getroffen wird. Man könnte das Entscheidungsverfahren in zwei Stufen aufteilen: Zunächst einmal suchen dafür kompetente Leute (also wahrscheinlich Ökonomen) aus den möglichen gesellschaftlichen Zuständen die Menge der paretooptimalen Zustände heraus, in einem zweiten Schritt einigt man sich dann darauf (gemäß der verschiedenen Verteilungsvorstellungen), welchen der paretooptimalen Zustände man verwirklicht.

Wer immer darauf Wert legt, daß seine über das Paretoprinzip hinausgehenden Wertvorstellungen in aktuelle Entscheidungen eingehen, wird sich hüten, sich blind auf ein solches Verfahren einzulassen. Zwar gilt für alle paretoinklusive Entscheidungskriterien $B(\succeq^e, E) \subseteq P(E)$, aber es ist keineswegs gesagt, daß unter der Voraussetzung verschiedener paretoinklusive Entscheidungskriterien und gegebener Entscheidungsverfahren, das beste was im politischen Prozeß erreichbar ist, ein Zustand aus $P(E)$ ist.

Betrachten wir ein einfaches konstruiertes Beispiel: Nehmen wir an, ein aus drei Experten bestehendes Gremium habe für zwei Individuen A und B, die nicht zu dieser Expertengruppe gehören, zwischen folgenden drei, jeweils durch die (voll vergleichbaren) Nutzenniveaus charakterisierten Zuständen zu wählen:

1. $u_A=9, u_B=9$
2. $u_A=10, u_B=10$
3. $u_A=8, u_B=15$

Sei 1. der gegenwärtig verwirklichte Zustand. Der Entscheidungsprozeß in dem gedachten Gremium laufe in zwei Stufen ab: Zunächst muß eine Entscheidung darüber getroffen werden, ob überhaupt eine Änderung des verwirklichten Zustands ins Auge

²⁹ Der letzte Satz antwortet auf eine andere Frage, als die danach, ob eine Maßnahme ergriffen werden sollte, die sicherstellt, daß wir uns auf der Paretogrenze befinden werden. Er antwortet auf die terminologische Frage, ob man eine solche Maßnahme „ökonomisch erwünscht“ nennen sollte. Aus einer solchen terminologischen Festlegung allein folgt natürlich nichts für die erste Frage. Wenn man eine solche Terminologie nicht mit einer Art von Handlungsanweisungen verbindet, dann hätte man – und da muß man Samuelson recht geben – die Terminologie angereichert, ohne sachlich einen Schritt weiterzukommen, was eher zur Verwirrung beitragen dürfte.

zu fassen ist. Dem konservativen Charakter des Gremiums entsprechend („Keine Experimente“) muß diese Entscheidung einstimmig fallen. In einem zweiten Schritt wird dann mit einfacher Mehrheit über den endgültig zu verwirklichenden Zustand entschieden. Das Gremium bestehe aus zwei Utilitaristen und einem Nutzenrawlsianer. Wie soll sich der Nutzenrawlsianer verhalten?

Es ist leicht zu sehen, daß $P(E)$ aus den Zuständen 2) und 3) besteht. Für einen Utilitaristen wird wegen $23 > 20 > 18$ gelten: $3) \succ 2) \succ 1)$. Für einen Nutzenrawlsianer dagegen wegen $10 > 9 > 8$: $2) \succ 1) \succ 3)$. Durch das dargestellte Verfahren hat der Nutzenrawlsianer die Möglichkeit zu verhindern, daß ein paretooptimaler Zustand verwirklicht wird. Und da, wenn er der Verwirklichung eines paretooptimalen Zustand zustimmt, die Mehrheitsverhältnisse zur Verwirklichung von 3) führen werden, wird er diese Möglichkeit auch ausnützen.

Man könnte dieses Beispiel auch in eine etwas andere Sprache kleiden. In einem gewissen Sinne ist ja für den Nutzenrawlsianer 2) bei den gegebenen Mehrheitsverhältnissen gar nicht „verwirklichbar“. In diesem Sinn bestünde für ihn $E = P(E)$ nur aus den Zuständen 1) und 3). Für ihn wäre also 1) paretooptimal. Sachlich hätte diese andere Sprachregelung an der Problematik nichts geändert. Es wäre dann aber sogar unmöglich sich schon allein auf die Menge $P(E)$ „wertfrei“ zu einigen, was schon allein sprachlich deutlich machen würde, welch schlechtes Geschäft es wäre, sich auf die Auswahl aus einer von Experten bestimmten Menge paretooptimaler Zustände zu beschränken.

An diesem Beispiel wird man vielleicht zwei Dinge kritisieren wollen: Zum einen ist es natürlich eingeständenermaßen konstruiert. Aber auch als konstruiertes zeigt es, daß politische Entscheidungsprozesse evtl. so organisiert sein können, daß es strategisch auch für jemanden, der dem Paretoprinzip zustimmt, ein großer Fehler sein könnte, der Bestimmung der Menge der paretoeffizienten Zustände eine zu große Rolle einzuräumen, oder sich gar auf Maßnahmen einzulassen, die die Verwirklichung eines paretooptimalen Zustandes sicherstellen. Allgemeiner geht es darum, daß immer wenn eine wirtschaftspolitische Maßnahme in der Realität ergriffen wird, damit gleichzeitig auch ein bestimmter gesellschaftlicher Zustand erreicht wird. Es gibt keinen gesellschaftlichen „Limbo“, Zwischenzustand, in den man eine Gesellschaft zunächst einmal versetzen könnte, und der aus mehreren (allen paretooptimalen) Zuständen gleichzeitig bestünde. In der Realität wird die Entscheidung für eine Maßnahme, die sicherstellt, daß ein paretooptimaler Zustand erreicht wird, immer auch zur Verwirklichung eines bestimmten Zustandes führen – der dann aber je nach paretoinklusive Entscheidungs-kriterium viel schlechter sein kann, als bestimmte nicht paretooptimale Zustände.

Eine weitere Kritik könnte der Vorwurf sein, daß ich hier normative mit strategischen Fragen vermischt habe. Diese Vermischung ist aber insofern gerechtfertigt, als wir ja von einer Vorstellung ausgehen, in der wir uns auf normative Urteile, die über das Paretoprinzip hinausgehen, unmöglich einigen können, es also angeblich keine sinnvollen, rationalen Kriterien geben kann, um bei einander widersprechenden normativer Urteilen zu einer Einigung zu kommen.

Statt der Vermischung von normativer und strategischer Fragestellung könnte man

natürlich die engen Grenzen einer Position, die nur das Paretokriterium als normatives Kriterium verwenden möchte, verlassen und versuchen, eine *normative* Theorie darüber zu entwickeln, wie Entscheidungen zu treffen sind, wenn die Entscheidungsträger verschiedene Entscheidungskriterien zugrunde legen. Für den Fall, daß alle diese Entscheidungskriterien das Paretoprinzip implementieren, gibt es wahrscheinliche keinen Grund, warum diese Theorie dann einen nicht paretooptimalen Zustand empfehlen sollte, da es dann ja annahmegemäß soziale Zustände geben würde, die alle Entscheidungsträger als besser empfinden würden. Man sollte sich aber keine Illusionen darüber machen, daß durch diese Möglichkeit die Menge $P(E)$ größere normative Relevanz bekäme: Eine solche Theorie wäre eine ausgereifte *normative* Theorie der sozialen Entscheidung, die auf einer höheren Ebene dasselbe Problem lösen müßte, das das Paretoprinzip auf einer tieferen Ebene nicht löst. Sie müßte Aussagen über die Auswahl der Entscheidungsträger treffen. Und ein Verfahren angeben, mit dessen Hilfe aus den verschiedenen Entscheidungskriterien eine Auswahlfunktion konstruiert würde. Eine solche Theorie würde alles andere bedeuten, als von den verschiedenen über das Paretokriterium hinausgehenden Gerechtigkeitsvorstellungen abzusehen – sie würde auf ihnen positiv aufbauen. Für den Fall, daß eine solche Theorie aufbauend auf den bei den Entscheidungsträgern anzutreffenden Entscheidungskriterien, selbst wieder ein Entscheidungskriterium (also eine Auswahlfunktion, die sich mit Hilfe eine zweistelligen Relation darstellen läßt) konstruieren würde, wären wir natürlich zurück bei dem Anfang dieser Ausführungen – gemäß dieses Kriteriums könnten in der Menge der paretoeffizienten Zustände die schlechtest möglichen dabei sein.³⁰

Die Überlegungen über die normative Irrelevanz der Menge paretooptimaler Zustände lassen sich direkt auf konkretere Fragestellungen der Wohlfahrtstheorie übertragen. Das ist möglich, weil moderne wohlfahrtstheoretische Arbeiten in der paretianischen Tradition regelmäßig darauf hinauslaufen, daß für eine bestimmte Fragestellung unter Vorgabe von Aussagen, die die Menge E festlegen, versucht wird, die Menge $P(E)$ der paretooptimalen Allokationen zu bestimmen. Diese Menge wird dann in Beziehung gesetzt zu einer Menge von Allokationen, die als Ergebnis einer bestimmten wirtschaftspolitischen Maßnahme erreicht werden können. Wirtschaftspolitische Maßnahmen können hier nun sehr vieles sein: Externalitäten korrigierende Steuern, Bereitstellung öffentlicher Güter, Steuern zur Finanzierung von Transfers, Öffnung eines Landes für Außenhandel, Zölle, Wahl eines ökonomischen Systems usw. Das Ziel dieser Übung ist es dann, zu zeigen, daß die beiden Mengen übereinstimmen, oder wenigstens, daß die eine Teilmenge der anderen ist. Da wir aber gesehen haben, daß das Paretoprin-

³⁰ Diese Bemerkungen sollen nicht bedeuten, daß ich eine solche normative Theorie empfehle. Sie hänge von den gewissermaßen „zufälligen“ ethischen Überzeugungen ab, die in einem Entscheidungsgremium vertreten wären. Abgesehen von dem Fall, in dem das Entscheidungsgremium sich gerade aus den Betroffenen einer Entscheidung zusammensetzt, wäre das ein nicht begründbares Element der Willkür. Wenn das Entscheidungskriterium sich aus den Betroffenen zusammensetzt, müssen wir näher erklären, was wir unter den Präferenzen der Betroffenen verstehen. Wenn die Präferenzen schon die ethischen Vorstellungen der Betroffenen mit ausdrücken, dann wäre die geforderte ethische Theorie der Entscheidung bei unterschiedlichen ethischen Überzeugungen nichts anderes als die Suche nach einer Wohlfahrtsfunktion. Trennen wir eigeninteressierte und ethische Präferenzen, dann stellt sich bei Berücksichtigung der letzteren wieder die Frage nach dem damit verbundenen Element der Willkür.

zip keineswegs zur Begründung von Maßnahmen ausreicht, die sicherstellen, daß ein paretooptimaler Zustand erreicht wird, ist selbst für den Fall, daß die Identität der erwähnten Mengen gezeigt werden kann, dies auf allein der Basis des Paretoprinzips nicht hinreichend, um irgendeine dieser konkreteren Maßnahmen zu empfehlen. Nur für Anhänger des Paretoextensionsprinzips wäre aus der Identität der entsprechenden Mengen unqualifiziert eine Empfehlung der untersuchten Maßnahmen abzuleiten (zunächst allerdings auch nur im Modell). Anhänger von anderen paretoinklusiven Wohlfahrtsfunktionen aber werden den untersuchten Maßnahmen nur dann in jedem Fall zustimmen, wenn sie das dieser Wohlfahrtsfunktion entsprechende Optimum verwirklichen.

2.3.4 Wohlfahrtstheoreme

Eine Untersuchung wohlfahrtstheoretischer Arbeiten zu den z. B. oben aufgezählten Themen, die die Frage zu beantworten sucht, ob die in diesen Arbeiten evtl. gemachten oder doch implizierten Empfehlungen auf der Basis des Paretoprinzips begründet werden können, wäre aus dem genannten Grund größtenteils eine Wiederholung. Es haben sich zwar in den verschiedenen Forschungsgebieten der Wohlfahrtsökonomie eigene Traditionen dazu entwickelt, wie die Beziehung zwischen den formalen Ergebnissen und normativen Empfehlungen ausgedrückt wird, und insofern könnte auch eine detaillierte Auseinandersetzung lohnend sein. Die Grundargumente für die normative Irrelevanz der entsprechenden wohlfahrtstheoretischen Ergebnisse wären aber immer wieder diesselben, so daß ein Beispiel genügen möge.

Von vielen Ökonomen werden sicher das erste und zweite Wohlfahrtstheorem als zentrale Ergebnisse der Wohlfahrtsökonomie betrachtet werden. Nun kann man einmal diese Theoreme einfach als mathematische Ergebnisse betrachten: Es werden bestimmte Mengen definiert und unter bestimmten Bedingungen wird die Identität dieser Mengen gezeigt. Ob diese beiden Theoreme vom mathematischen Standpunkt aus als bemerkenswert gelten können, darf dabei offen bleiben.

In Wirklichkeit aber werden diese beiden Theoreme aber natürlich nicht einfach als Beispiele angewandter Mathematik betrachtet. Sie erfahren auf einer ersten Stufe schon allein deshalb eine ökonomische Interpretation, weil die betrachteten Entitäten (Mengen, deren Elemente, die Beziehungen zwischen den Elementen) in einer ökonomischen Begrifflichkeit präsentiert werden: Es ist die Rede von Märkten, Haushalten, Wettbewerbsbedingungen, Preisen, Gütern, Haushalten, Unternehmen usw. Schon auf dieser Ebene stellen sich zwei interessante Fragen: Zum einen ist zu klären, wie die Beziehung zwischen den mathematischen Ergebnissen und unserer Vorstellung der ökonomischen Realität eigentlich gedacht werden. Ein solche Beziehung wird durch die verwendete Begrifflichkeit zunächst einmal in jedem Fall unterstellt. Die Antwort auf diese Frage ist alles andere als klar: Während für die Herleitung (evtl. unter den Beweis vereinfachenden Annahmen) meist relativ viel Raum verwendet wird, scheinen viele Autoren der Meinung zu sein, daß für die Interpretation einige cursorische Be-

2 Grenzen der Wohlfahrtsökonomie

merkungen genügen.³¹

Die Beziehung des den Theoremen der Wohlfahrtsökonomie zugrundeliegenden Modells zu der von uns angenommenen Wirklichkeit kann grundsätzlich in folgenden vier Weisen gedacht werden:

1) Als ein Verhältnis der vollkommenen Entsprechung: Die Wirklichkeit ist so, wie sie in unserem Modell beschrieben wird. Dabei ist zu beachten, daß das den Theoremen zugrundeliegende Modell *allgemein* sein soll (deshalb bezogen auf den Untersuchungszweck als vollständig spezifiziert anzusehen ist) und natürlich nicht beansprucht, das Ganze der Wirklichkeit zu beschreiben. Die vollkommene Entsprechung ist also so zu verstehen, daß es Präferenzordnungen, Produktionsfunktionen usw. mit den geforderten Eigenschaften gibt, auch wenn uns diese nicht bekannt sind, die das Modell zu einer genauen Beschreibung eines Teils der Wirklichkeit machen. Im Sinne einer vollkommenen Entsprechung würden die beiden Wohlfahrtstheoreme also behaupten, daß es zumindest *praktisch möglich* ist, die wirtschaftlichen Prozesse einer Gesellschaft in der empirischen Realität so in der Form „freier“ kompetitiver Märkte zu organisieren, daß die Voraussetzungen der beiden Theoreme erfüllt sind, und deshalb das Ergebnis der empirischen Marktprozesse so sein wird, daß demgegenüber keine Paretoverbesserungen möglich sein werden, und daß jeder Zustand, der bei (empirisch) gegebener Technologie, (empirisch) gegebenen Präferenzen, (empirisch) gegebenen Anfangsausstattungen unter der Voraussetzung der beliebigen Verteilbarkeit als paretoeffizient zu bezeichnen wäre, bei einer bestimmten (empirischen) Verteilung dieser Anfangsausstattung durch den empirischen Marktprozess erreicht würde.

2) Als ein Verhältnis der Annäherung: Die Wirklichkeit ist annähernd, ungefähr so, wie sie in unserem Modell beschrieben wird. Abweichungen sind vorhanden, aber für die vorliegenden Zwecke vernachlässigbar, d. h. die Annäherung ist hinreichend genau.

Diese Beschreibung nimmt Bezug auf die Zwecke, die mit Hilfe eines gegebenen Modells angezielt werden. Solche Zwecke könnten z. B. die Prognose sein oder wie in unse-

³¹ Das ist ein Beispiel dafür, daß man aus mathematischer Exaktheit und überhaupt der mathematischen Methode in der Ökonomie wirklich keinen Fetisch machen sollte. Gerade weil in der Ökonomie oft mit sehr stark vereinfachenden oder offen unrealistischen Annahmen gearbeitet wird, ist die Interpretation eine der wichtigsten Aufgaben. Eine Interpretation versucht ein Modell in Beziehung zu der von uns wahrgenommenen ökonomischen Realität zu setzen, und muß, da diese Realität vor allem erst einmal als hochkomplex, durch einfache Modelle nicht adäquat erfaßbar gedacht ist, selbst unpräzise, unformal sein. Exakt (mathematisch) können wir die Beziehung von einem einfachen Modell nur wieder zu einem vielleicht umfassenderen, komplizierteren Modell der Wirklichkeit darstellen, nicht zu unserer unpräzisen, vagen, höchste Komplexität annehmenden Vorstellung der ökonomischen Realität. Wenn als Empfehlung der mathematischen Methode immer wieder gesagt wird, daß da in jedem Fall alle gemachten Annahmen klar offengelegt werden müßten, so ist dagegen zu betonen, daß im Übergang von Modell zur „Realität“, d. h. in der Interpretation das einfach unmöglich ist; daß aber eine solche Interpretation in jedem Fall notwendig ist, wenn Ökonomie nicht Glasperlenspiel sein soll. Wen das nicht überzeugt, den möchte man fragen, ob er denn alle in seinen Modellen gemachten Annahmen in der Wirklichkeit für erfüllt ansieht. Vernünftigerweise wird die Antwort darauf „Nein“ sein. Dann kann man weiter danach fragen, welche Bedeutung denn ein Modell hat, dessen Annahmen nicht erfüllt sind. Worauf nur eine verbale, nicht in allen ihren Annahmen geklärte Erläuterung als Antwort folgen kann.

rem Zusammenhang, die Herleitung normativer Aussagen. In diesen Zwecken könnte man natürlich nun so ziemlich alles unterbringen, was man möchte und auf diese Weise jedes Modell als eine hinreichend genaue Beschreibung der Wirklichkeit verkaufen: Ein Modell könnte z. B. hinreichend genau die Wirklichkeit approximieren, um als Basis einer Hauptstudiumsvorlesung zu dienen. Die Geeignetheit für diesen Zweck sagt aber sicher wenig über die Nähe oder Ferne eines Modells zu der von ihm angeblich annähernd genau dargestellten Wirklichkeit. Selbst die ersten beiden Beispiele für Zwecke sind noch rein formal. Um zu wissen, ob ein Modell für Prognosezwecke genau genug ist, ist es nötig zu wissen, welchem Zweck die Prognosen dienen, und welche Genauigkeit von diesen Prognosen für diesen Zweck erfordert wird. Im Zusammenhang praktischer, normativer Fragestellungen werden diese Zwecke von Prognosen in letzter Analyse selbst praktische Zwecke sein. (Eine Prognose ist genau genug, um sicherzustellen, daß eine Brücke hält, daß der Patient mit einer gewissen Wahrscheinlichkeit nicht an der Behandlung stirbt usw.)

Die Frage nach der Genauigkeit von Modellen für Prognosezwecke führt zu praktischen Fragestellungen. Genauso führt die Frage nach der Genauigkeit von Modellen für normative Zwecke zur Frage nach der Prognosegenauigkeit. Normative Fragen haben mit der Frage von Handlungsanweisungen oder Handlungsempfehlungen in der von uns wahrgenommenen empirischen Wirklichkeit zu tun. Immer wenn Handlungen von ihren Konsequenzen her normativ bewertet werden, ist es notwendig zwischen Ziel und Mittel zu unterscheiden, und deshalb zu fragen, inwieweit eine bestimmte Handlung ein bestimmtes Ergebnis in der Wirklichkeit zur Folge haben wird. Das aber ist genau eine Prognose. Die geforderte Genauigkeit dieser Prognose hängt wiederum von dem angestrebten Ziel ab.

Wird das Verhältnis von Modell und Wirklichkeit als eines der Annäherung im eben beschriebenen Sinne verstanden, dann würden die beiden Wohlfahrtstheoreme also behaupten, daß es zumindest möglich ist, die wirtschaftlichen Prozesse einer Gesellschaft in der empirischen Realität so in der Form „freier“ kompetitiver Märkte zu organisieren, daß die Voraussetzungen der beiden Theoreme *annähernd* erfüllt sind, und deshalb das Ergebnis der empirischen Marktprozesse *mit hinreichend hoher Wahrscheinlichkeit* so sein wird, daß demgegenüber keine *wesentlichen* Paretoverbesserungen möglich sein werden (*weil diese nur sehr wenige Leute betreffen würden und/oder deren Los nur unwesentlich verbessern würden*), und daß jeder Zustand, der bei (empirisch) gegebener Technologie, (empirisch) gegebener Präferenzen, (empirisch) gegebenen Anfangsausstattungen unter der Voraussetzung der beliebigen Verteilbarkeit als paretoeffizient zu bezeichnen wäre, bei einer bestimmten (empirischen) Verteilung dieser Anfangsausstattung durch den empirischen Marktprozess *mit hinreichend hoher Wahrscheinlichkeit annähernd* erreicht würde.

3) Als die Isolierung eines Aspektes der Wirklichkeit, der, um zu einem Modell der Kategorien 1) oder 2) zu kommen, durch weitere Aspekte ergänzt werden muß: Bei dieser Verhältnisbestimmung von Modell und Wirklichkeit steht das Vorbild der Naturwissenschaften Pate. Ein Beispiel wären die physikalischen Modelle des freien Falles von Körpern im Vakuum und der Bewegung von Körpern auf eine reibungsfreien schiefen Ebene. Auch wenn diese modellmäßige Isolierung selbstverständlich eine

notwendige Vorgehensweise in allen Wissenschaften ist, sind doch zwei Hinweise zu beachten. Zum ersten ist im Falle der beiden physikalischen Beispiele die Isolierung des betrachteten Aspekts auch empirisch möglich: Die Modelle sind hinreichend genaue Beschreibungen von durchführbaren Experimenten. Zum zweiten ist im Falle der physikalischen Beispiele die Ergänzung durch weitere Aspekte zu einem volleren, die empirische Wirklichkeit beschreibenden Modell nicht bloß ein Desideratum: Reibung und Luftwiderstand lassen sich ohne weiteres in umfassendere Modelle integrieren. Für den Fall, daß weder das eine noch das andere möglich ist, ist die Behauptung, man habe einen Aspekt der Wirklichkeit isoliert, nichtssagend, da nicht nachprüfbar.³²

4) Als das Verhältnis einer explizit kontrafaktischen Konstruktion zur davon – in unserem Verständnis – wesentlich verschiedenen Wirklichkeit.

Aus den Fällen 3) und 4) lassen sich natürlich keine *real zu verwirklichenden* normativen Folgerungen im hier betrachteten Sinn herleiten. Denn eine Maßnahme, die in einem Modell, das einen (oder auch mehrere) Aspekte isoliert, nach irgendwelchen Kriterien empfehlenswert erscheinen mag, muß dies in einem Modell, das die vernachlässigten Aspekte mitberücksichtigt in keiner Weise mehr sein. Es sei denn, für die betrachtete Fragestellung bringen die vernachlässigten Aspekte keine wesentliche Änderung. Aber dann würden wir uns in Wirklichkeit (bezogen auf die betrachtete Fragestellung) im Fall 2) befinden.

Noch weniger können natürlich aus explizit kontrafaktischen Modellen gewonnene Empfehlungen normatives Gewicht beanspruchen. Auch hier wäre wieder der Fall zu unterscheiden, wo das Modell zwar in einigen Aspekten bewußt als kontrafaktisch, aber bezogen auf die vorliegende Fragestellung als gute Annäherung an die Realität gesehen würde. Aber schon diese Formulierung zeigt ja wieder, daß man damit faktisch auf den Fall zwei zurückgreift.

Nun wird gerade auch im Fall der Wohlfahrtstheoreme davon gesprochen, die Bedeutung dieser eingestandenermaßen kontrafaktische Annahmen machenden Theoreme bestünde darin, daß sie einen (normativen) Referenzpunkt oder „Maßstab“ bilden könnten.³³ Es ist aber nicht ganz klar, was mit einem solchen „Referenzpunkt“ gemeint ist. Denkbar wären zwei Interpretationen:

Die Wohlfahrtstheoreme und ihre Annahmen könnten einmal einen Referenzpunkt in dem Sinne bilden, daß sie so etwas wie den Hintergrund für Modelle bilden, die sich der Realität annähern möchten. Im Vergleich dieser „realistischeren“ Modelle mit dem Referenzpunkt, würde dann klar, in was sich diese *realistischeren* Modelle und deshalb auch die Wirklichkeit vom Basismodell der Wohlfahrtstheoreme unterscheidet. Der Referenzpunkt hilft uns dann gerade zu sehen, warum (u.a.) die normativen Folgerungen des Basismodells in der Realität *nicht* gelten. Diese Interpretation von Referenzpunkt ist unproblematisch und entspricht oft der Vorgehensweise wohlfahrtstheoretischer Einführungen. Nach der Darstellung der Wohlfahrtstheoreme werden dann etwa externe Effekte eingeführt und gezeigt, wie und warum die beiden Wohlfahrtstheoreme beim

³² Nichtssagend wäre diese Behauptung auch, wenn das Ergebnis des durch weitere Aspekte ergänzten vollständigen Modelles, durch die Hinzufügung/Weglassung des isolierten Aspekts nicht wesentlich, d. h. nicht prognoserelevant, beeinflusst würde.

³³ Vgl. etwa Jaeger [1981, 692] und Külp [1982, 485].

Vorliegen externe Effekte nicht mehr gelten. Diese Interpretation entspricht aber natürlich auch meiner These, daß auf einem als kontrafaktisch verstandenen Modell keine normativen Forderungen für die Wirklichkeit abgeleitet werden können.

Eine zweite Interpretation könnte die sein, daß das Basismodell einer kompetitiven Marktwirtschaft deshalb einen normativen Referenzpunkt darstelle, weil die Wirklichkeit zwar den Annahmen dieses Modelles in vielen Punkten nicht entspreche, aber wirtschaftspolitisches Handeln wegen der positiven normativen Eigenschaften der Modellwelt kompetitiver Märkte dem Ziel dienen sollte, die Wirklichkeit dem Modell anzunähern. Abgesehen davon, daß auf der Basis des Paretokriteriums aus den beiden Wohlfahrtstheoremen keinerlei normative Überlegenheit kompetitiver Märkte folgt, ist der dubiose Schluß von den „guten“ normativen Eigenschaften eines kontrafaktischen Modells auf die normativ „guten“ Eigenschaften politischer Maßnahmen, die die Wirklichkeit dem Modell annähern sollen, mit der Theorie des Zweitbesten eindeutig widerlegt. Wenn es wesentliche Abweichungen der Wirklichkeit von dem Basismodell der Wohlfahrtstheoreme gibt, die nicht in der Disposition der Wirtschaftspolitik liegen (wie z. B. die praktische Unmöglichkeit der Erhebung von nicht verzerrenden Steuern), dann kann die politische Eliminierung anderer Abweichungen die Verwirklichung eines zweitbesten Optimums verhindern. Wer aber die Verwirklichung eines Paretooptimums für wünschenswert hält, müßte unter den Bedingungen der Theorie des Zweitbesten ein zweitbestes Optimum anstreben, daß ja dann *unter den tatsächlich zur Wahl stehenden Zuständen* eben ein Paretooptimum wäre.³⁴ Das einzige Argument, das es dagegen gibt, ist es, diese nicht zur Disposition stehenden Abweichungen zu leugnen oder als unwichtig zu bewerten. Das aber führt uns wieder zum Fall 2) zurück.

Nur in den Fällen 1) und 2) lassen sich also aus Modellen überhaupt normative Folgerungen für die Wirklichkeit ableiten. Eingeständenermaßen aber erfüllen die Modelle, in denen die zwei Wohlfahrtstheoreme gelten, weder die Anforderungen an Modelle des Falles 1) noch des Falles 2). Mir ist überhaupt kein Ökonom bekannt, der eine vollkommene Entsprechung behaupten würde. Die meisten Ökonomen würden aber wohl auch zugeben, daß unsere ökonomische Wirklichkeit in wesentlichen Punkten so von den Annahmen des Modelles der Wohlfahrtstheoreme abweicht, daß dies Modelle nicht mehr als Annäherungen gelten können.³⁵ Neben der Tatsache, daß – wie wir gleich sehen werden – die normative Irrelevanz der Menge der paretooptimalen Zustände sich unmittelbar auf die beiden Wohlfahrtstheoreme überträgt, haben wir also hier einen zweiten, unabhängigen Grund für die normative Irrelevanz dieser Theoreme: Die ihnen zugrundeliegenden Modelle können nicht als hinreichend genaue Approximationen der Wirklichkeit gelten.

Die zweite Frage ist, ob diese Theoreme als solche eigentlich normativer Art sind oder nicht. Die Antwort auf diese zweite Frage scheint mir ein klares Nein zu sein. Das wird daran deutlich, daß selbst jemand, der das Paretoprinzip für falsch hält, diesen Theoremen zustimmen wird, sie für korrekt halten wird. Für ihn allerdings werden

³⁴ Zur Theorie des Zweitbesten vgl. etwa Schlieper [1982].

³⁵ Gründe dafür sind die Existenz von Externalitäten und öffentlichen Gütern, oder allgemeiner von fehlenden Märkten für bestimmte „Güter“ bzw. „Übel“, Steuern, Monopole, nicht geräumte Märkte, Informationsasymmetrien und nicht rationales Verhalten.

2 Grenzen der Wohlfahrtsökonomie

diese Theoreme normativ irrelevant sein, da er das Paretoprinzip ablehnt.

Auch wenn die Theoreme als solche also keinen normativen Charakter tragen, werden sie häufig als eine Empfehlung von marktwirtschaftlicher ökonomischer Organisation interpretiert.

Unter den Voraussetzungen der Fälle 1) und 2) oben ist eine solche Interpretation des *ersten* Wohlfahrtstheorems ist möglich, wenn man dem Erreichen eines paretoeffizienten Zustands als solchem normative Bedeutung beimißt. Wenn gemäß einer ethischen Überzeugung jeder paretoeffiziente Zustand besser oder gleichgut ist wie jeder paretoineffiziente, dann (und nur dann³⁶) kann aus der Tatsache, daß jedes kompetitive Marktgleichgewicht paretoeffizient ist (erstes Wohlfahrtstheorem) geschlossen werden, daß ein Wettbewerbmarkt zu verwirklichen ist. (Das wäre im Sinne der obigen Diskussion eine wirtschaftspolitische Maßnahme.)

Wenn man sich aber, wie es viele Ökonomen tun, von der Paretoextensionsregel distanzieren (mit dem Hinweis, daß paretooptimale Zustände unter ethischen Gesichtspunkten abstoßend sein können), dann folgt – und das war der Inhalt der bisherigen Diskussion – aus dem ersten Wohlfahrtstheorem normativ einfach nichts. Der kompetitive Marktprozess führt zu einem von der Ausgangsverteilung bestimmten paretooptimalen Zustand. Der aber kann eben einer dieser ethisch abstoßenden Zustände sein.³⁷

Unter bestimmten, gegenüber den Voraussetzungen des ersten Wohlfahrtstheorems noch anfordernderen Bedingungen ist jeder paretooptimale Zustand (und damit jeder

³⁶ Jedenfalls, solange die Entscheidung aufgrund einer Ordnung oder Teilordnung erfolgt.

³⁷ Feldman [1987, 890] führt zwei Argumente für die gegenteilige Behauptung an: Einmal behauptet er, eine normative Relevanz des „Ersten Fundamentalen Wohlfahrtstheorem“ damit, daß Paretooptimalität ein Gemeinwohlkonzept sei, das allgemeine Zustimmung finden könne: „No one would argue that society should settle for a situation that is not optimal, because if *A* is not optimal, there exists a *B* that all prefer.“ In der Tat habe ich aber zu zeigen versucht, daß aus der Tatsache, daß jede paretoinklusive Wohlfahrtfunktion einen paretooptimalen Zustand als den besten Zustand aussondern wird, in keiner Weise folgt, daß man sich in jedem Fall, d. h. unabhängig von dem tatsächlich erreichten Zustand, für eine Verwirklichung paretooptimaler Zustände einsetzen wird. Immer dann, wenn der bei Verwirklichung eines paretooptimalen Zustandes tatsächlich erreichte Zustand von einem nicht paretooptimalen Zustand übertroffen wird, ist es aber sicher sinnvoll, die Verwirklichung nicht paretooptimaler Zustände zu fordern bzw. durchzusetzen. (Das setzt wie schon erwähnt die Sprachregelung voraus, in der man Einigkeit über die Menge der paretooptimalen Zustände voraussetzt.)

Das zweite Argument Feldmans ist eine anschauliche Beschreibung der Tatsache, daß in den verbreiteten Modellen die Menge der paretooptimalen Zustände das Maß 0 besitzt. „Therefore, to say that the market mechanism leads an economy to an optimal outcome is to say a lot.“ Wenn aber diese Menge mit dem Maß 0 garantiert alle möglichen Wohlfahrtsniveaus abdeckt (was für viele paretoinklusive Wohlfahrtfunktionen der Fall sein wird), dann heißt zu sagen, daß der Marktmechanismus einen Zustand dieser Menge verwirklichen wird, normativ einfach nichts.

Als Vergleich betrachte man einen einzelnen Konsumenten mit einer Cobb-Douglas Nutzenfunktion und man biete ihm einen Zufallsmechanismus an, der garantiert, daß in jedem Fall ein Zustand seiner Budgetgeraden verwirklicht wird und zwar in der Form einer Gleichverteilung über die Budgetgerade. Alternativ einen Mechanismus der garantiert einen *bestimmten* Zustand verwirklicht, der *nicht* auf der Budgetgeraden liegt, dafür aber in der Nähe seines Optimum. Es gibt unendlich viele solcher Zustände, die ein rationaler Konsument bei weitem vorziehen würde.

Zustand, der von einer paretoinklusiven Wohlfahrtsfunktion als der beste ausgewählt werden könnte) von (zumindest) einer Ausgangsverteilung aus als Gleichgewicht eines Wettbewerbsmarktes zu erreichen. Dieses zweite Wohlfahrtstheorem soll nun das Problem lösen, daß paretooptimale Zustände und damit unter den Bedingungen des ersten Wohlfahrtstheorems die Zustände, die als Gleichgewicht eines Marktes erreicht werden können, z.T. ethisch inakzeptabel sein können. Um eine akzeptable oder gar die beste Allokation zu erreichen, muß man ja „nur“ die entsprechende Ausgangsverteilung verwirklichen und dann den Rest dem Markt überlassen.

Das verlangt aber bekanntermaßen die Möglichkeit der kostenlosen Umverteilung, wie sie theoretisch etwa durch lump sum Transfers verwirklicht werden könnte. Wenn es die Möglichkeit kostenloser Umverteilung nicht gibt (und man darf in der Realität davon ausgehen, daß es sie nicht gibt), dann sind die meisten der in den Modellen als paretooptimal ausgewiesene Allokationen durch die Kombination von Umverteilung und Marktprozeß *nicht* erreichbar, womit das auf dem zweiten Wohlfahrtstheorem beruhende Argument für die normative Überlegenheit des Marktprozesses zusammenbricht.

Man beachte, daß die unhandliche Formulierung, die von „in den Modellen als paretooptimal ausgewiesene Allokationen“ spricht, mit Bedacht gewählt wurde. Diese Allokationen sind dann, wenn keine Möglichkeit der nicht verzerrenden Umverteilung besteht, nicht erreichbar, und deshalb auch nicht paretooptimal. Ob die über tatsächlich zur Verfügung stehende Maßnahmen der Umverteilung zusammen mit dem Marktprozess erreichbaren Zustände mit der Menge der tatsächlich paretooptimalen Zustände zusammenfällt, hängt von den möglichen Alternativen zum Marktprozess ab und kann deshalb nur im Vergleich entschieden werden.

Das zweite fundamentale Wohlfahrtstheorem wird manchmal auch so interpretiert, daß daraus die Trennbarkeit von allokativen und distributiven Maßnahmen abgeleitet werden könne und damit die Berechtigung dafür, daß Ökonomen sich nur oder hauptsächlich mit allokativen Überlegungen beschäftigen und distributive Überlegungen ausklammern könnten, da diese ja nicht „wissenschaftlich“ entscheidbare ethische Aussagen voraussetzen würden.

Diese Interpretation ist in keiner Weise gerechtfertigt: Solange die Gesamtwohlfahrt auf der Basis von Endzuständen bestimmt wird, wie das beim Ansatz über paretoinklusive Wohlfahrtsfunktionen z. B. der Fall ist, zeigt das zweite fundamentale Wohlfahrtstheorem vielmehr die engste Verknüpfung zwischen allokativen und distributiven Überlegungen. Das zweite fundamentale Wohlfahrtstheorem gewinnt seine (sehr beschränkte) normative Bedeutung gerade aus der Tatsache, daß die rein allokativen Überlegungen des ersten Wohlfahrtstheorems nicht ausreichen, um die Überlegenheit von kompetitiven Marktprozessen zu zeigen. Die Ergebnisse des Marktprozesses werden evtl. nur bei Umverteilung normativ akzeptabel. Die Bestimmung einer akzeptablen Ausgangsverteilung setzt voraus, daß zunächst (auch verteilungsmäßig) akzeptable oder gar „beste“ paretooptimale Zustände bestimmt werden. Das geschieht auf der Basis einer paretoinklusiven Wohlfahrtsfunktion, die schon allokativ (Paretoprinzip) und distributive Gesichtspunkte integriert. Weiter muß dann für die verschiedenen möglichen Ausgangsverteilungen die Allokation bestimmbar sein, zu der von ih-

2 Grenzen der Wohlfahrtsökonomie

nen aus der Marktprozess jeweils führt. Wenn die „richtige“ Ausgangsverteilung nicht verwirklicht ist, kann der Marktprozess zu ethisch abstoßenden Ergebnissen führen. Um die „richtige“ Ausgangsverteilung zu bestimmen, müssen die allokativen Prozesse, die zum Marktgleichgewicht führen, bekannt sein.³⁸ Also: Um ein ethisch akzeptables Marktergebnis zu sichern, benutze ich ein Kriterium, das allokativen und distributiven Gesichtspunkte integriert, und brauche die Möglichkeit der Umverteilung; um „korrekt“ umverteilen zu können, brauche ich die Kenntnis der allokativen Prozesse. Verschränkter könnten Allokation und Distribution kaum gedacht werden.

Im Grunde geht es bei den beiden Wohlfahrtstheoremen gar nicht um Fragen der Trennbarkeit, oder allgemein des Verhältnisses von Allokation und Distribution. Vielmehr geht es darum zu fragen, inwieweit und unter welchen Voraussetzungen ein bestimmter Mechanismus (Markt) ein Element einer ohne Rücksicht auf diesen Mechanismus definierten Menge von Allokationen (die der paretooptimalen) erreicht.

In den beiden fundamentalen Wohlfahrtstheoremen wird eine Verbindung hergestellt zwischen zwei Mengen: Einmal die Menge der paretooptimalen Allokationen, die erreichbar sind, wenn die Anfangsausstattungen der Wirtschaftssubjekte beliebig umverteilt sind, und sonst nur die Gesamtanfangsausstattung der Ökonomie (und evtl. die vorhandene Technologie) als Beschränkungen auferlegt sind. Zum anderen haben wir die Menge der als Gleichgewicht eines kompetitiven Marktmechanismus erreichbaren Allokationen, wobei weiterhin die Anfangsausstattungen beliebig umverteilt sind, und die Gesamtausstattung der Ökonomie (und evtl. die vorhandene Technologie) gegeben sind. Das erste Wohlfahrtstheorem sagt, daß diese zweite Menge unter gewissen Voraussetzungen eine Teilmenge der ersten ist. Das zweite sagt, daß (unter engeren Voraussetzungen) die erste eine Teilmenge der zweiten ist. Sind die Voraussetzungen beider Theoreme erfüllt, sind die beiden Mengen also identisch.

Das zweite Wohlfahrtstheorem schließt aus, daß bestimmte paretooptimale Allokationen mit der Kombination von kostenloser Umverteilung und Marktmechanismus nicht erreichbar sind. Das erste Wohlfahrtstheorem sagt, daß jedes Gleichgewicht des Marktmechanismus eine paretooptimale Allokation ist (Markt -> Effizienz). Das zweite sagt, daß auch beliebige (paretoinklusive) Verteilungsziele sich über den Marktmechanismus erreichen lassen (Markt -> Verteilung). Zusammengefaßt heißt das also, daß wenn wir uns des Marktmechanismus bedienen, weil er (unter kaum verwirklich-

³⁸ Solche Überlegungen haben zu der Frage geführt, warum man dann überhaupt noch den Markt braucht und nicht einfach gleich das Marktergebnis verwirklicht, das im Falle einer reinen Tauschökonomie ja sogar eine der „richtigen“ Ausgangsverteilungen wäre. Diese Frage ist als Polemik gegen die Behauptung der Trennbarkeit von allokativen und distributiven Überlegungen auf der Basis des zweiten fundamentalen Wohlfahrtstheorem gerechtfertigt. Eine wirtschaftspolitische Empfehlung einer solchen Vorgehensweise kann daraus aber natürlich nicht abgeleitet werden: Das Argument bewegt sich in der gleichen unrealen Welt, wie die kritisierte Interpretation des zweiten Wohlfahrtstheorems. Wirtschaftspolitische Empfehlungen sollten auf Modellen aufbauen, die Maßnahmen beschreiben, von denen wir annehmen, daß sie so ungefähr auch in Wirklichkeit durchführbar sind, und Zustände beschreiben, von denen wir annehmen, daß sie so ungefähr verwirklicht sind. Ein Modell, das behauptet jeder als paretooptimal beschriebene Zustand eines traditionellen Gleichgewichtsmodells einer kompetitiven Marktwirtschaft sei durch administrative Maßnahmen erreichbar, gehört wohl nicht in diese Kategorie.

ten Bedingungen) Effizienz sicherstellt, dann müssen wir *wegen des Marktmechanismus* (nicht wegen des Effizienzziels) auf die Verwirklichung paretoinklusive Verteilungsziele nicht verzichten. Die zwei fundamentalen Wohlfahrtstheoreme sagen weniger etwas über das Verhältnis von Allokations- und Distributionszielen als über einen bestimmten Mechanismus. Deshalb ist es schon ganz und gar verfehlt, wenn aufgrund der angeblich in den fundamentalen Wohlfahrtstheoremen bewiesene Trennbarkeit von Allokation und Distribution, Ökonomen sich berechtigt fühlen, bei ganz anderen Fragen, bei der Betrachtung anderer Mechanismen, oder gar unter Verhältnissen, in denen die Voraussetzungen der beiden Theoreme nicht erfüllt sind, distributive Überlegungen zu vernachlässigen.

2.4 Kompensationskriterien

Wir sind bis jetzt davon ausgegangen, daß Effizienz mit der Verwirklichung eines paretooptimalen Zustands zu identifizieren sei, wie das ja auch der verbreiteten Sprachregelung unter Ökonomen entsprechen dürfte. Die Tatsache, daß aufgrund eines solchen Kriteriums für Effizienz sich in der Regel keine Handlungsanweisungen oder -empfehlungen ableiten lassen, hat dazu geführt, daß einige Ökonomen versucht haben, Effizienz auf der Grundlage von sogenannten *Kompensationskriterien* zu definieren.

Diese Kriterien gehen, um zu normativ relevanten Aussagen zu gelangen, z. T. bewußt über das Paretokriterium hinaus, und bestätigen so indirekterweise meine These, daß das Paretokriterium/Paretoprinzip im allgemeinen keine normativ relevanten Aussagen zuläßt. Allerdings wurden diese Kriterien zum Teil in der Meinung vorgeschlagen, damit „wertfreie“ Aussagen machen zu können, was natürlich eine Illusion gewesen ist. Diese Kriterien sind nicht einmal – wie vielleicht das Paretokriterium – in dem Sinne „relativ wertfrei“ zu nennen, daß sie auf quasi uneingeschränkte Zustimmung unter Ökonomen hoffen könnten: Auch viele Ökonomen finden die in ihnen ausgedrückten Werturteile alles andere als einleuchtend.³⁹ Im folgenden werde ich einige dieser Kriterien kritisch darstellen. Es soll dabei klarwerden, daß diese Kriterien – auch wenn sie mehr Werturteile voraussetzen als das Paretokriterium – keine überzeugende Basis für eine normative ökonomische Theorie liefern, sei es, weil sie mit internen Problemen (etwa der Konsistenz) behaftet sind, oder sei es, weil sie nicht ausreichen, um bestimmte wirtschaftliche Entscheidungen zu empfehlen, und damit ihren eigentlichen Zweck verfehlen. Selbst für eine klare Unterscheidung zwischen Gesichtspunkten der Effizienz und der Gerechtigkeit sind diese Kriterien nicht geeignet.

2.4.1 Das Kaldor-Hicks-Kriterium

Startpunkt der Karriere von Kompensationskriterien war der Vorschlag von Kaldor [1939], aufgegriffen von Hicks [1939]. Kaldor ging es darum zu zeigen, daß die Wohlfahrtsökonomie als Wissenschaft normative Empfehlungen geben kann. Ein Kriterium

³⁹ Und weisen z. B. darauf hin, daß die bloße *Möglichkeit* von Paretoverbesserungen kein hinreichendes Argument für die Behauptung ist, ein Zustand sei besser als der andere. So etwa Varian [1992, 405].

für Wissenschaftlichkeit war dabei, daß diese Empfehlungen ohne den interpersonellen Vergleich von Nutzen auskommen, da ein solcher Vergleich wissenschaftlich nicht möglich sei. Nach Kaldors Kompensationskriterium kann eine Politik dann auf wissenschaftlicher Basis empfohlen werden, wenn es nach Durchführung dieser Politik durch Kompensation, d. h. durch eine Form der Umverteilung möglich ist, einen Zustand zu erreichen, in dem es allen Individuen zumindest gleich gut geht, wie vor der Durchführung dieser Politik und einigen Individuen besser. Hicks [1939] akzeptierte dieses Kriterium und interpretierte es als Kriterium der ökonomischen Effizienz.

Bevor wir uns in technische Details dieses Vorschlages verlieren, möchte ich gleich zu Beginn zeigen, warum der Anspruch Kaldors, der sich insbesondere in seinen kurzen Ausführungen zu Pigou zeigt, und sein Kriterium nicht zusammenpassen.⁴⁰ Kaldor meint, daß sein Argument die Vorgehensweise Pigous rechtfertigen würde, Wohlfahrts-theorie in zwei Abschnitte zu teilen: einen, der es mit den Fragen des aggregierten Wachstums der Produktion zu tun hat und in dem der Ökonom mit nicht bezweifelbarem wissenschaftlichem Anspruch normative Empfehlungen geben könnte, und einem zweiten, der sich mit Fragen der Verteilung beschäftigt. Der Ökonom könne unmöglich entscheiden, welche Einkommensverteilung die soziale Wohlfahrt erhöhe, und so könne er in diesem zweiten Teil nur zeigen, wie eine gegebene gewünschte Einkommensverteilung auf beste Weise erreicht werden könne. Kaldor meint also ein Argument dafür zu liefern, was wir heute die Trennbarkeit von allokativen und distributiven Fragen nennen würden.

Betrachten wir eine einfache Welt von zwei Individuen. In Abb. 2.1 finden wir im Nutzenraum eingezeichnet: den Ausgangszustand *A* und die von ihm aus durch Umverteilung erreichbaren Zustände und einen Zustand *B*, der von *A* aus durch eine bestimmte politische Maßnahme („Freihandel“) erreicht würde, einschließlich der von ihm aus durch Umverteilung erreichbaren Zustände. Offensichtlich erfüllt *B* das von Kaldor vorgeschlagene Kriterium, und er würde die entsprechende Maßnahme deshalb empfehlen. Die Begründung dafür ist, daß, wenn unsere („nicht wissenschaftlichen“) Verteilungsvorstellungen verlangen sollten, daß Individuum 1 im Zustand *B* gegenüber dem Ausgangszustand entschädigt werden sollten, das ja ohne weiteres möglich ist, ohne daß dadurch die Politik, die uns zunächst nach *B* geführt hat, dadurch an ihrer Vorteilhaftigkeit einbüßt.⁴¹ Was Kaldor damit aber nicht berücksichtigt, ist die Möglichkeit, daß unsere Verteilungsvorstellungen einen Zustand wie *C*, der nur von *A* aus durch Umverteilung erreichbar ist, jedem Zustand, der von *B* aus erreichbar ist, vorziehen. In einem solchen Fall wäre die beste Weise, die gewünschte Einkommensverteilung zu erreichen, die vorgeschlagene Politikmaßnahme nicht zu unternehmen, oder wenn sie schon unternommen wurde, sie – wenn möglich – zurückzunehmen.

Dieses einfache Beispiel zeigt, daß es auf der Basis des Kriteriums Kaldor keineswegs möglich ist, eine klare Trennung von allokativen und distributiven Gesichtspunkten durchzuführen. Eine Politik, die nach seinem Kriterium empfehlenswert wäre, ist nur

⁴⁰ Das folgende Argument benutzt die von Scitovsky entdeckte Inkonsistenz des Kaldorkriteriums. Siehe dazu unten S. 64. Zu den dem von Scitovsky selbst vorgeschlagenen Kriterium siehe unten S. 71 ff.

⁴¹ Dieses Argument setzt klar voraus, daß es sich bei dieser Möglichkeit nicht nur um eine theoretische handelt.

für bestimmte Verteilungsvorstellungen akzeptabel.

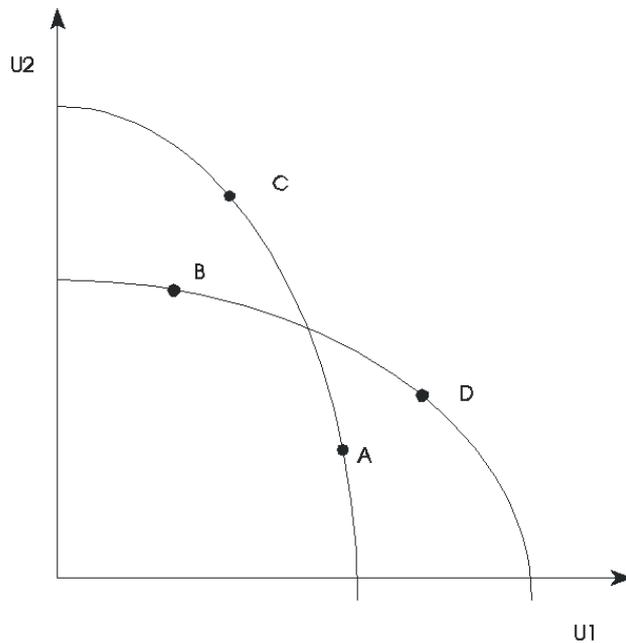


Abbildung 2.1: Kaldors Kriterium in einer einfachen Zweipersonenwelt

Ein Versuch, Kaldors Kriterium zu retten, könnte es sein, darauf hinzuweisen, daß der kritische Fall nur dann auftreten kann, wenn der Ausgangszustand nach der vorausgesetzten Verteilungsvorstellung nicht optimal war. Wäre Kaldors Argument nicht zu retten, wenn wir davon ausgehen könnten, daß im Ausgangszustand unsere (paretoinklusiven) Verteilungsvorstellungen schon verwirklicht sind?

Während es unter dieser Voraussetzung in der Tat korrekt wäre, daß die angenommene Verteilungsvorstellung in einem von *B* aus durch Umverteilung erreichbaren Zustand besser verwirklicht werden könnte als in einem von *A* aus erreichbaren, wäre damit für die Trennbarkeit von Allokation und Distribution nur wenig gewonnen. Denn wie will man beurteilen, ob diese Voraussetzung erfüllt ist, ohne die entsprechende Verteilungsvorstellung zu kennen und auf den Ausgangszustand anzuwenden? Diese Voraussetzung ist eine Annahme über die normativen Verteilungseigenschaften des Ausgangszustands, die im Sinne Kaldors nicht „wissenschaftlich“, d. h. von umstrittenen Werturteilen nicht frei sind. Außerdem gäbe es dann ja verwirklichtbare Zustände, die *A* gegenüber paretoüberlegen sind *und* die unseren Verteilungsvorstellungen besser entsprechen als *B*. Warum dann *B* und nicht einen dieser Zustände verwirklichen?

Kaldor hat sein Kriterium rein verbal und mit Hilfe eines Beispiels eingeführt. Deshalb ist es nicht ganz eindeutig, wie man sein Kriterium formal darstellen soll. Ich folge hier Sen [1970, 30-32], weil mir seine formale Rekonstruktion am ehesten den verbalen Ausführungen Kaldors⁴² und den in der Lehrbuchliteratur verbreiteten Darstellungen

⁴² Sen rekonstruiert das Kaldorkriterium als eine Relation zwischen gesellschaftlichen Zuständen. Es

des Kaldorkriteriums zu entsprechen scheint⁴³.

Sen folgend gelte als Definition der Relation „besser im Sinne des Kaldorkriteriums“ (\succ_k):

Sei $S(g)$ die Menge aller sozialen Zustände, die von g aus durch Umverteilung erreicht werden können. Dann gelte:

$$g \succ_k g' : \leftrightarrow \exists g'' : g'' \in S(g) \wedge \forall i : g'' \succeq_i g' \wedge \exists i : g'' \succ_i g'$$

Kaldor möchte die so definierte Relation als Entscheidungsregel verstanden wissen: Ein Ökonom sollte unzweideutig die Verwirklichung von g empfehlen, wenn $g \succ_k g'$.⁴⁴ Das wichtigste Problem dieser Entscheidungsregel ist bekannt. Schon Scitovsky [1941] hat gezeigt, daß für bestimmte Präferenzen diese Entscheidungsregel in dem Sinne inkonsistent ist, daß sie sowohl die Durchführung bestimmter Maßnahmen wie ihre Rücknahme, wenn einmal durchgeführt, empfehlen würde. D. h. für bestimmte Zustände g und g' kann bei bestimmten Präferenzen gelten.⁴⁵

$$g \succ_k g' \wedge g' \succ_k g$$

Das allein sollte eigentlich genügen, die von Kaldor vorgeschlagene Entscheidungsregel zu diskreditieren.⁴⁶ In der Tat hat die Entdeckung dieser Schwierigkeit zu einer Weiterentwicklung des Kriteriums geführt. Trotzdem gibt es überraschenderweise noch immer Ökonomen, die diese Regel als Entscheidungsregel empfehlen oder benutzen.⁴⁷ Deshalb möchte ich noch auf einige weitere Schwierigkeiten eingehen.

scheint mir der natürlichsten Interpretation von Kaldor [1939], Hicks [1939], Scitovsky [1941] und Kaldor [1946] zu entsprechen, daß diese Autoren den Vorschlag Kaldors genau so verstanden haben. Es gibt keinerlei Anzeichen in einem der genannten Aufsätze dafür, daß die entsprechenden Autoren an eine Relation von Paaren von gesellschaftlichen Zuständen und Mengen von aggregierten Güterbündeln gedacht haben, wie sie Chipman u. Moore [1978] in ihrer formalen Rekonstruktion des Kaldorkriteriums definieren. Sachlich würde der Chipman/Moore Vorschlag für die folgende Diskussion keinen Unterschied machen, solange man davon ausgeht, daß jedem gesellschaftlichem Zustand eindeutig ein Menge $S(g)$ der von ihm aus durch Umverteilung erreichbaren Zustände zugeordnet ist. Diese Voraussetzung wird von Kaldor, Hicks und Scitovsky ihrem Verständnis von Umverteilung entsprechend natürlicherweise gemacht. Sie gilt aber nicht mehr für die durch Samuelson inspirierten „Situationen“. Vgl. unten S. 74 f.

⁴³ So mit kleinen Variationen Varian [1992, 405], Böventer u. Illing [1995, 259], Rose u. Sauernheimer [1992, 516 f.]. Auch der Name, mit dem das Kriterium bezeichnet wird, ist nicht einheitlich. Im deutschen Sprachraum scheint „Kaldor-Hicks-Kriterium“ am verbreitetsten zu sein. Vgl. auch Kulp [1982, 476-479].

⁴⁴ Kaldor möchte gegen Harrod und Robbins zeigen, daß Ökonomen auf wissenschaftlicher Basis wirtschaftspolitische „Empfehlungen“ machen können. (Das englische „prescriptions“ ist sogar noch stärker als „Empfehlung“.) Maßnahmen, die seinem Kriterium entsprechen, sind (aus wirtschaftswissenschaftlicher Sicht) durchzuführen. Ob danach oder daneben noch Verteilungsmaßnahmen ergriffen werden sollen, ist für diese „Vorschrift“ irrelevant.

⁴⁵ Das wäre etwa im Beispiel der Abb. 2.1 der Fall. Vgl. Scitovsky [1941] und Sen [1970].

⁴⁶ Kaldor [1946] hat als Reaktion auf Scitovsky [1941] sein ursprüngliches Kriterium aufgegeben und das Scitovsky-Kriterium empfohlen. Dieses Kriterium werde ich im folgenden Unterabschnitt darstellen.

⁴⁷ Nur als Beispiel: Caplan [1999, 835]. Stringham [2001, 41] geht soweit, zu behaupten: „Kaldor-Hicks efficiency is widely accepted by the economics profession and is gaining ground in the area of law. Modern law and economics is almost entirely based on Kaldor-Hicks, and judges such as Richard

Die Relation \succ_k ist binär und Kaldor stellte sie im Rahmen von binären Entscheidungssituationen vor: Es gibt einen gegebenen Ausgangszustand, in dem sich eine Ökonomie befindet und es steht *eine* wirtschaftspolitische Maßnahme zur Auswahl. Die von Kaldor vorgeschlagene Entscheidungsregel ist es dann einfach, die Durchführung der wirtschaftspolitischen Maßnahme zu empfehlen, wenn der dadurch erreichte Zustand in Sinne seines Kriteriums besser ist als der Ausgangszustand.

Ohne die Definition einer Relation der Kaldorindifferenz läßt sich das Kaldorkriterium auf Entscheidungssituationen mit mehr als zwei Alternativen nur dann übertragen, wenn alle betrachteten gesellschaftlichen Zustände durch die Relation \succ_k eindeutig geordnet sind.

Dieser extrem eingeschränkte Anwendungsbereich des Kaldorkriteriums gerät in Konflikt mit der Formulierung des Kriteriums und dem Anspruch mit ihm über das Pareto-kriterium hinausgehen zu können.

Betrachten wir das Beispiel, das Kaldor [1939] selbst verwendet, die Aufhebung der englischen Korngesetze als Beispiel für die Einführung von Freihandel. Kaldor behauptet (was für die Einführung von Freihandel keineswegs allgemein der Fall sein muß), daß die Geschädigten einer solchen Maßnahme für ihre Einkommensverluste voll entschädigt werden könnten, indem die Regierung einen aus den Einkommenszuwächsen der Nutznießer finanzierten Transfer vornimmt, so daß die Situation aller sich bessert. Eine solche Umverteilung sei einer Regierung immer möglich. Dabei geht Kaldor ganz offensichtlich von der Vorstellung aus, daß eine solche Umverteilung eine reale (nicht nur auf dem Papier stehende) Möglichkeit ist, über die man (allerdings nicht nach ökonomischen Kriterien) entscheiden kann.⁴⁸

Nehmen wir für einen Augenblick an, es gäbe genau eine Form, in der diese Umverteilung möglich wäre, dann würde also gelten, daß wir drei mögliche, erreichbare gesellschaftliche Zustände zu betrachten haben: g der Ausgangszustand vor Freihandel, g' der Zustand nach Freihandel ohne Umverteilung und der Zustand g'' mit Umverteilung. Dabei würde gelten $g' \succ_k g$, $g'' \succ_p g$ und $g'' \succ_k g$. Außerdem wäre $P(E) = \{g', g''\}$. Zwischen g' und g'' kann das Kaldorkriterium nicht entscheiden. Trotzdem empfiehlt Kaldor die Verwirklichung von g' . Die sinnvollste Interpretation davon wäre, daß für ihn *als* Ökonomen die beiden Zustände indifferent sind, *als* Ökonom hätte er also auch keinerlei Einwände gegen die Verwirklichung von g'' . Dann stellt sich aber unmittelbar die Frage, unter welchen Bedingungen zwei Zustände als indifferent gelten sollen: Paretoindifferenz reicht hier nicht aus, da ja die Zustände g' und g'' gemäß der Pareto-Relation nicht vergleichbar sind. Alle gemäß der Pareto-Relation nicht vergleichbaren Zustände einfach als indifferent zu erklären (wie in der Paretoextentionsregel), führt nicht weiter, da ja dann g und g' als indifferent zu betrachten wären. Aufgrund der schon erwähnten Inkonsistenz des Kaldorkriteriums wird es wahrscheinlich einfach unmöglich sein, eine konsistente und plausible Definition für die Kaldorindifferenz zu finden.

Eine andere Möglichkeit, die Entscheidung für g' zu rechtfertigen, wäre es, g'' irgend-

Posner advocate using the standard to decide cases ..."

⁴⁸ Darauf beruht die Kraft seines Arguments.

wie von einer Verwirklichung auszuschließen. Aus irgendwelchen, den ökonomischen Überlegungen vorausgehenden Entscheidungen oder Tatsachen steht g'' gar nicht zu Wahl. Aber was für Gründe könnte es dafür geben?

g'' könnte nicht verwirklicht sein. Das würde aber dem Argument Kaldors alle Überzeugungskraft nehmen. Er meinte ja gerade, der Ökonom könne g' empfehlen, weil etwa ein Zustand wie g'' verwirklicht sei. Wenn g' aus irgendwelchen Verteilungsüberlegungen heraus nicht akzeptabel wäre, wäre es problemlos möglich, von da aus das aus Verteilungsgründen akzeptable g'' zu erreichen.

g'' könnte aus Verteilungsgründen nicht erwünscht sein. Damit würde aber die Empfehlung des Ökonomen wieder von Verteilungsüberlegungen abhängen. Und natürlich käme man in Probleme mit einer Situation wie in Abb. 2.1: Denn dort wäre dann aufgrund dieser Information über die Verteilungsvorstellungen C zu verwirklichen (von A aus zu erreichen) und nicht B .⁴⁹

Eine letzte Möglichkeit wäre es, daß der Ökonom überhaupt erst zu Rate gezogen wird, nachdem die Entscheidung gegen g'' schon gefallen ist. Aber warum sollte er dann aus *ökonomischen Gründen* g' empfehlen? In einem Fall der Inkonsistenz wie in Abb. 2.1 müßte er ja dann vielmehr (gegen das Kaldorkriterium) C empfehlen, solange er selbst dem Paretokriterium verpflichtet ist. Aber selbst wenn wir von einem solchen Fall absehen, wie könnte Kaldor – wissend das g'' nicht verwirklicht werden wird – g' empfehlen, wenn er gleichzeitig (nicht aufgrund ökonomischer Überlegungen) wüßte, daß g' ungerecht ist? Wie könnte er eine solche Empfehlung gegenüber den durch die Verwirklichung von g' Geschädigten rechtfertigen? Zumindest müßte er also voraussetzen, daß g' aus Verteilungsüberlegungen heraus unproblematisch ist, womit wir wieder bei Verteilungsüberlegungen zurück wären. Die Begründung Kaldors für sein Kriterium funktioniert in diesem Fall jedenfalls nicht mehr. Wenn er weiß, daß g'' nicht verwirklicht werden wird, kann er nicht g' damit empfehlen, daß – falls g'' ungerecht wäre – man ja immer noch g' verwirklichen könnte.⁵⁰

Die bisher an einem Beispiel durchgeführten Überlegungen möchte ich noch einmal in einer größeren Allgemeinheit darstellen. Unter Ökonomen ist die Vorstellung wahrscheinlich recht verbreitet, das Kaldorkriterium führe – einmal abgesehen von den Fällen der Inkonsistenz – irgendwie weiter als das Paretokriterium. Das Paretokriterium führt ja meist zu Situationen der Unentschiedenheit, weil i. d. R. die Menge der paretooptimalen Zustände mehrere Elemente enthält, die nicht paretoindifferent sind. Es gibt nun in der Tat Mengen E von erreichbaren Zuständen, für die das Paretokriterium nicht zu einer Entscheidung führt, wohl aber das Kaldorkriterium. Dies gilt aber nur, solange man entweder Erreichbarkeitsmengen betrachtet, die nicht alle die Zustände enthalten, die im Rahmen des Kriteriums selbst als erreichbar vorausgesetzt werden oder indem man die Relation der Kaldorüberlegenheit durch eine Relation Kaldordifferenz ergänzt.

Nehmen wir einmal an, für jeden Zustand g enthalte die Menge E alle (beliebig de-

⁴⁹ Die Verteilungsvorstellung als paretoinklusiv vorausgesetzt.

⁵⁰ Die letzten Überlegungen tragen selbst schon ethischen Charakter. Es wird – hier – von einem Ökonomen verlangt, seine Entscheidungen und Empfehlungen gegenüber den davon Betroffenen (insbesondere davon Benachteiligten) rechtfertigen zu können.

finierten!) Zustände $S(g)$, wie sie in der Anwendung des Kaldorkriteriums gebraucht werden. Wir betrachten dann zwei beliebige paretooptimale Zustände g und g' . Selbstverständlich kann das Kaldorkriterium unter den gegebenen Bedingungen keine Ordnung zwischen diesen Zuständen herstellen, da es keinen Zustand $g'' \in E$ geben kann (und deshalb auch keinen Zustand $g'' \in S(g) \subseteq E$ oder $g'' \in S(g') \subseteq E$) für den gilt $g'' \succ_p g$ oder $g'' \succ_p g'$.

Der vermeintlich über das Paretokriterium hinausführende Charakter beruht also *notwendig* auf der zweideutigen Verwendung von „Erreichbarkeit“. Immer dann, wenn es gilt, das Kaldorkriterium auf eine Menge E erreichbarer gesellschaftlicher Zustände anzuwenden, die alle Zustände enthält, die in der Definition von $S(g)$ Kaldorkriteriums als erreichbar erklärt werden, führt das Kaldorkriterium keinen Schritt weiter als das Paretokriterium.

Für die Begründung dieser Aussage wurden bis jetzt aber nur paretooptimale Zustände betrachtet. Muß diese Aussage aber nicht modifiziert werden, wenn wir auch Zustände betrachten, die nicht paretooptimal sind? Zunächst einmal ist es klar, daß es für jeden nicht paretooptimalen Zustand mindestens einen Zustand gibt, der diesem Zustand paretoüberlegen und deshalb auch „besser im Sinne von Kaldor“ ist, sonst wäre der Zustand paretooptimal. Unter geeigneten Annahmen über die Topologie von E bezüglich \succeq_p gibt es dann auch solche Zustände in $P(E)$. Wenn wir uns in einem nicht paretooptimalen Zustand g befinden, dann würde das Paretokriterium zu einer Entscheidung nicht ausreichen, wenn die Mächtigkeit von $P(E)$ größer als eins ist. Dagegen würden das Paretoextensionskriterium und das („geschichtlich“) modifizierte Paretoextensionskriterium (s. o. S. 46) zu einer Entscheidung führen, indem sie entweder alle Elemente von $P(E)$ oder alle Elemente von $P(E)$, die gegenüber g' pareto-besser sind als indifferent erklären. Das ist aber genau das, was das Kaldorkriterium tun muß, um eine Entscheidung in diesem Fall zu ermöglichen. Man kann annehmen, daß eine halbwegs plausible Form, die Kaldorrelation durch eine Kaldorindifferenz zu ergänzen, nicht alle Elemente von $P(E)$, aber mehr als die geschichtlich modifizierte Paretoextensionsregel als indifferent erklären würde.⁵¹

Ein weiteres oft erwähntes Problem ist die Tatsache, daß der Kaldortest, auch wenn er explizit auf den interpersonellen Vergleich von Nutzen verzichtet, deshalb natürlich nicht „wertfrei“ genannt werden kann. Liest man den Beitrag von Kaldor [1939], dann kann kein Zweifel daran bestehen, daß für Kaldor die *Möglichkeit* der Kompensation hinreichende Bedingung für eine ökonomische Politikempfehlung war, unabhängig davon, ob die entsprechende Kompensation durchgeführt wird oder nicht. Es ist ohne weiteres möglich, einfache Beispiele zu konstruieren, wo diese Empfehlung mit paretoinklusive Wohlfahrtsfunktionen konfligiert, für die die Maßnahme nur durchzuführen ist, wenn die Kompensationen tatsächlich gezahlt werden. Diese unterschiedliche Bewertung einzelner Maßnahmen ist Ausdruck der unterschiedlichen *Wertmaßstäbe*, die zu ihrer Beurteilung benutzt wurden, hinreichender Hinweis auf die über das

⁵¹ Gegen eine entsprechende Entscheidungsregel wären dann ähnliche Einwände zu erheben wie gegen die beiden Versionen des Paretoextensionskriteriums. Diese Überlegungen werde ich weiter unten noch einmal kurz aufgreifen, wenn ich das Scitovsky-Kriterium betrachte, das die mögliche Inkonsistenz des Kaldorkriteriums vermeidet, aber von den letzten Einwänden ebenfalls betroffen ist.

Paretoprinzip hinausgehende Wertgeladenheit des Kaldorkriteriums.⁵²

Nun ist es keineswegs das Ziel dieser Überlegungen, ein Kriterium deshalb zu diskreditieren, weil es über das Paretokriterium hinaus Werturteile voraussetzt. Ganz im Gegenteil, geht es gerade darum, daß nur auf der Grundlage solcher weitergehender ethischer Urteile überhaupt wirtschaftspolitische Empfehlungen möglich sind. Allerdings stellt sich dann direkt die Frage der ethischen *Attraktivität* der entsprechenden Werturteile und des darauf aufbauenden Kriteriums. Die aber ist sicher gering, wenn man die Fülle von Problemen betrachtet, mit denen das Kaldorkriterium behaftet ist.

Wenn das Kaldorkriterium auch keinen expliziten interpersonellen Vergleich von Nutzen voraussetzt, so findet doch implizit ein solcher Vergleich insofern statt, als ganz real bei der Verwirklichung einer Maßnahme ohne Kompensation die Verbesserung der Situation einiger Individuen mit der Verschlechterung anderer Individuen einhergeht. Offensichtlich wiegt die Verbesserung der einen die Verschlechterung der anderen auf, wie sonst könnte die entsprechende Maßnahme empfohlen werden? Für diejenigen, die Opfer der Maßnahme sind, ohne kompensiert zu werden, ist es ein schwacher Trost, daß der empfehlende Ökonom darauf verzichtet hat, explizit den Gewinn der einen gegen den Verlust der anderen abzuwiegen. Gegen dieses Argument könnte ein Verteidiger des Kaldorkriterium allenfalls noch darauf verweisen, daß die Sprache eines Vergleichs von Verbesserung der einen und Verschlechterung der anderen überhaupt unverständlich und unsinnig ist. Das aber liefe auf einen generellen Ausschluß von Verteilungsüberlegungen hinaus – eine sehr starke, von nur wenigen Ökonomen offen vertretene Position.

Wie schafft es nun das Kaldorkriterium, zu wirtschaftspolitischen Empfehlungen zu kommen, ohne dazu einen (expliziten) interpersonellen Vergleich von Nutzen zu benötigen? Die Antwort ist offensichtlich und doch überraschend: indem das Prinzip verletzt wird, daß die Bewertung eines sozialen Zustandes allein von den Bewertungen dieses Zustandes durch die betroffenen Individuen und nicht von irgendwelchen zusätzlichen, vom Ökonomen selbst an die Beurteilung herangetragenen Wertmaßstäben. Wenn man das Kaldorkriterium benutzt, tut man gerade das, was Hicks so beredt verurteilt:

Man ersetzt „the given thermometers (the scales of preferences of the individuals) by a new thermometer of one’s own. The investigator himself decides what he thinks to be good for society, and praises or condemns the system he is studying by that test. This is the method which is rightly condemned as unscientific. It is the way of the prophet and the social reformer, not of the economist.“ [Hicks 1939; zitiert nach dem Nachdruck in Hicks 1981, 63]

Dieser Punkt ist wichtig, weil er immer gilt, wenn Kompensationskriterien (auch ausgereifere als das inkonsistente Kaldorkriterium) dazu verwandt werden, einzelne gesellschaftliche Zustände zu „ordnen“.

Um die hier aufgestellte Behauptung zu begründen, ist es zunächst einmal notwen-

⁵² Kaldor [1939] selbst weist auf *eine* Wertannahme hin: „... the scientific status of his prescription is unquestionable, provided that the basic postulate of economics, that each individual prefers more to less, a greater satisfaction to a lesser one, is granted.“ Allerdings hat er in keiner Weise gezeigt, wie man ohne zusätzliche Wertannahmen von hier zu seinen Empfehlungen kommt.

dig, sie zu präzisieren. Eine Lesart des erwähnten Prinzips ist es, es so deuten, daß darin verlangt wird, daß die gesellschaftliche Ordnung zweier sozialer Zustände allein davon abhängt, wie die betroffenen Individuen die beiden Zustände ordnen,⁵³ oder (wenn man irgendeinen Grad der Vergleichbarkeit von Nutzen annimmt, was Kaldor aber ausdrücklich als unwissenschaftlich ablehnt) allein von den Nutzen der betroffenen Individuen.

Ein Blick auf die Definition von \succ_k zeigt aber, daß diese Relation der eben formulierten Forderung nicht entspricht: Die Ordnung zweier gesellschaftlicher Zustände g und g' hängt nicht allein von der Ordnung dieser *beiden* Zustände durch die betroffenen Individuen ab, sondern von der Ordnung der Elemente von $S(g)$ und g' . (Sie hängt unter anderem *auch* von der Ordnung von g und g' ab, weil trivialerweise $g \in S(g)$.) Deshalb kann sich die gesellschaftliche Ordnung zweier Zustände g und g' ändern, ohne daß sich die individuelle Ordnung dieser Zustände ändert, wenn entweder $S(g)$ sich ändert oder wenn sich die Ordnung von von g verschiedener Elemente von $S(g)$ und g' ändert.

Die Tatsache, daß das Kaldorkriterium die hier verwendete Lesart des Prinzips, daß die gesellschaftliche Bewertung sozialer Zustände allein von der Bewertung dieser Zustände durch die betroffenen Individuen abhängen soll, widerspricht, desavouiert dieses Kriterium nicht automatisch.

Zum einen kann man ja diese spezielle Lesart des Prinzips ablehnen. Allerdings sollte man sich dabei bewußt sein, daß, wer mit diesem Argument das Kaldorkriterium benutzt, auf antipaternalistisches Pathos vielleicht eher verzichten sollte. Schließlich haben die in den Modellen, in denen das Kaldorkriterium Verwendung findet, betrachteten Individuen nun einmal keine Präferenzen über Mengen $S(g)$ im Vergleich mit Zuständen g' , sondern über Zustände.

Diese Lesart des erwähnten Prinzips scheint mir also deshalb angemessen, weil es dem hinter diesem Prinzip liegenden Anspruch entspricht, bei ökonomischen normativen Urteilen die Kriterien der betroffenen Individuen zugrunde zu legen. In den Modellen in denen das Kaldorkriterium Anwendung finden soll, haben die Individuen eben eine Präferenzordnung über soziale Zustände. Das Kaldorkriterium ordnet soziale Zustände, weil sie in einer bestimmten Relation zu anderen Zuständen stehen: ein Gesichtspunkt, den die Individuen selbst nicht verwenden.

In jedem Fall sollte man sehen, daß das Kaldorkriterium das „Thermometer“ der betroffenen Individuen nicht weniger durch ein eigenes ersetzt als etwa eine Wohlfahrtsfunktion. Unter den Bedingungen des Arrow-Theorems kann es gar kein Kriterium geben, das die Forderung erfüllt, gesellschaftliche Zustände allein auf der Basis der Ordnung dieser Zustände zu ordnen.⁵⁴ Individualistische Wohlfahrtsfunktionen erfüllen die Forderung, gesellschaftliche Zustände allein auf der Basis der Nutzen dieser Zustände für die betroffenen Individuen zu ordnen. Sie benützen also Informationen über die Vergleichbarkeit von Nutzenniveaus oder Nutzenintensitäten. Wer solche Ver-

⁵³ Sie soll also „neutral“ sein. Vgl. Sen [1970, 72].

⁵⁴ D. h. unter den Bedingungen des unbeschränkten Definitionsbereichs, des Paretoprinzips, der Unabhängigkeit von irrelevanten Alternativen und der Nichtexistenz eines Diktators. Vgl. Sen [1970, 41-46].

gleiche für möglich hält, könnte sogar den Anspruch erheben, sich mehr auf das Thermometer der Betroffenen zu stützen, als ein Anhänger des Kaldorkriteriums. Denn das Kaldorkriterium benützt, um zu (k)einer (!) Ordnung der gesellschaftlichen Zustände zu kommen, direkt Informationen über die Ordnung von gesellschaftlichen Zuständen, die nicht zur Wahl stehen. Das erscheint, solange man eine Vergleichbarkeit von Nutzen annimmt, eher als willkürlich, als die Benutzung von Nutzeninformationen. Nimmt man eine solche Vergleichbarkeit nicht an, so sind vom Standpunkt der betroffenen Individuen beide Verfahren wohl als gleich willkürlich zu beurteilen.

Zum ändern kann man natürlich das Prinzip überhaupt und die mit ihm verbundene Rhetorik aufgeben. Dann ist die Verletzung einer bestimmten Lesart des Prinzips nicht mehr schlimm, allerdings steht dann das Kaldorkriterium in Konkurrenz zu anderen Kriterien, die das Prinzip auf andere Weise verletzen und es wird eine ethische Reflexion zu unternehmen sein, um sich für eines zu entscheiden.

Wenn man das Kaldorkriterium als eine Relation von sozialen Zuständen versteht, so wie es Sen [1970] getan hat, wie es in den meisten Darstellungen geschieht, wie es der natürlichen Lesart der Aufsätze von Kaldor [1939] und Hicks [1939] entspricht, und was auch letztlich Voraussetzung dafür ist, daß das Kaldorkriterium überhaupt zu normativen Aussagen führen kann, dann ergibt sich als Problem die Abhängigkeit von $S(g)$. Unterschiedliche $S(g)$, d. h. unterschiedliche Annahmen darüber, was von g aus durch Umverteilung zu erreichen ist, führt auch zu unterschiedlichen Bewertungen von g .

$S(g)$ ist die Menge der Zustände, die von g aus durch Umverteilung erreichbar sind. In den meisten Modelle, in denen Kompensationskriterien verwandt werden, wird diese Menge dann einfach mit der Menge gleichgesetzt, die bei „kostenloser“ Umverteilung erreichbar wäre. Eine der sicher korrekten Überzeugungen vieler Ökonomen ist es aber, daß solche kostenlose Umverteilung in Wirklichkeit nicht möglich ist. Wenn man nun von dieser Überzeugung ausgeht, dann wird ein Werturteil, das sowieso schon verlangt hat, möglichen und nicht tatsächlichen Zuständen ein normatives Gewicht zu geben, noch fragwürdiger, da nun diese möglichen Zustände nicht mehr real sondern nur noch auf dem Papier „möglich“ sind. Warum soll es irgendeine normative Bedeutung haben, daß in einer fiktiven Welt ein Zustand möglich wäre?

Ein Ausweg wäre, $S(g)$ irgendwie realistischer zu verstehen, in dem Sinne wie dies von Samuelson [1950] vorgeschlagen wurde. Auf ein damit verbundenes Problem hat schon Samuelson selbst hingewiesen: Während es sehr einfach ist, die Idee einer realistischeren Sicht von Umverteilung auf einem Blatt Papier mit Hilfe eines frei gezeichneten Bleistiftstrichs zu verdeutlichen, wäre es natürlich nötig, um zu normativen Ergebnissen auf der Basis von Kompensationskriterien zu kommen, daß man diese Menge $S(g)$ wirklich (jedenfalls in einer sinnvollen Annäherung) bestimmen kann. Das dürfte sich oft als großer Problem herausstellen, auch weil sich positive Ökonomie vielleicht zu wenig mit Möglichkeiten und Formen der Umverteilung beschäftigt hat. Am weitesten in die gewünschte Richtung geht noch die sogenannte Literatur des Zweitbesten, in der ja z. B. versucht wird, auf der Grundlage von einfachen Annahmen über zur Verfügung stehende Instrumente der Umverteilung modellmäßig Aussagen über dadurch erreichbare Zustände herzuleiten.

Die normale ökonomische Interpretation eines gesellschaftlichen Zustandes ist die, daß es sich dabei um eine bestimmte Allokation handelt. Umverteilung ausgehend von einer Allokation heißt dann nichts anderes, als daß die gegebenen aufsummierten Gütermengen (unter eventueller Berücksichtigung von „allokativen Verzerrungen“ oder sonstiger Umverteilungskosten) auf andere Weise unter den betroffenen Individuen verteilt werden. Das sich daraus ergebende Verständnis von $S(g)$ hat den Vorteil, daß sich Ambivalenzen bei der Bestimmung von $S(g)$ nur aufgrund der erwähnten eventuell nötigen Berücksichtigung von Umverteilungskosten ergeben. Ein Nachteil dieses Verständnisses von $S(g)$ ist es, daß auf dieser Basis sich etwa die Überlegenheit von Freihandel gegenüber dem Autarkiezustand nicht zeigen läßt. Olson [1958] hat darauf hingewiesen und Samuelson [1962] hat bestätigt, daß Fälle denkbar sind, in denen sich durch Umverteilung der bei Freihandel sich ergebenden Allokation nicht alle Individuen besser stellen lassen. Ein Ausweg daraus ist es, Samuelson folgend eine weitere Definition von $S(g)$ zuzulassen, bzw. diese Menge überhaupt nicht mehr ausgehend von einer gegebenen Allokation her zu definieren. Daraus ergibt sich dann allerdings eine zweite Form der Ambivalenz der Kompensationskriterien, da dadurch die Definition von $S(g)$ oder der $S(g)$ entsprechenden Menge ziemlich beliebig wird.

2.4.2 Das Scitovsky-Kriterium

Scitovsky [1941] hat nicht nur auf die mögliche Inkonsistenz des Kaldorkriteriums hingewiesen, sondern auch gleich eine Möglichkeit der Korrektur vorgeschlagen, das später so genannte Scitovsky-Doppeltkriterium:⁵⁵

Sei $S(g)$ definiert wie oben. Die Relation „besser im Sinne Scitovskys“ \succ_s sei dann definiert:

$$g \succ_s g' :\leftrightarrow g \succ_k g' \wedge \neg g' \succ_k g$$

Diese Definition schließt die erwähnte Inkonsistenz trivialerweise aus, indem sie eben direkt verlangt, nur dann davon zu sprechen, daß ein Zustand einem anderen gegenüber „besser im Sinne Scitovskys“ ist, wenn die erwähnte Inkonsistenz nicht vorliegt.

Gegenüber der Kaldorkriterium ist das Scitovsky-Kriterium also insofern ein Fortschritt, als sie nicht in derselben Weise inkonsistent sein kann. Der Preis der dafür gezahlt wird, ist zwar, daß das Scitovskykriterium in weniger Fällen zu einer wirtschaftspolitischen Empfehlung führen wird. Das ist aber insofern kein großer Verlust, als diese Fälle, in denen das Scitovsky-Kriterium zu keiner Entscheidung mehr führt, genau die sind, in denen das Kaldorkriterium zu widersprüchlichen Empfehlungen führte.

⁵⁵ Scitovsky hat übrigens nicht nur ein Kriterium vorgeschlagen, das seiner Überzeugung nach „Effizienz“ konsistent definieren kann, sondern hat zugleich eine wesentlich skeptischer Position darüber vertreten, ob aufgrund seines Kriteriums als Effizienzverbesserungen ausgewiesene Reformen tatsächlich unternommen werden sollten: Nur nach einer nicht weniger wichtigen Einigung darüber, daß dabei auch Gerechtigkeits- und Verteilungskriterien genügend berücksichtigt sind, wären für Scitovsky solche Reformen durchzuführen. Scitovsky teilt allerdings mit Kaldor die Überzeugung, daß sein Effizienzkriterium „objektiv“ sei, was für Verteilungskriterien nicht gelte.

Es dauerte nicht lange und ein eigenes (zunächst etwas weniger gravierendes) Problem des Scitovsky-Kriteriums wurde entdeckt: Bei bestimmten Präferenzen der Individuen ist die Scitovsky-Relation intransitiv.⁵⁶ Das ist insofern nicht ganz so gravierend, als es nicht ganz eindeutig ist, daß Transitivität eine notwendige Bedingung eines rationalen Entscheidungskriteriums ist. Zwar verbinden wir mit dem Komparativ „besser“ normalsprachlich sicher die Eigenschaft der Transitivität und würden mit einigem Erstaunen reagieren, wenn jemand behaupten würde, er fände Tee besser als Kaffee und Kaffee besser als Kakao aber keineswegs Tee besser als Kakao. Die natürlichste Interpretation wäre wohl anzunehmen, daß, wer solches behauptet, einfach nicht versteht, was es heißt, irgend etwas als „besser“ als etwas anderes zu bezeichnen. Aber Transitivität einer Relation \succeq ist nicht Voraussetzung für die Existenz einer konsistenten Entscheidungsfunktion $B(\succeq, E)$, auch wenn diese Entscheidungsfunktion dann allerdings wenig intuitiv wäre.⁵⁷

Chipman u. Moore [1978] haben nun aber ein Beispiel konstruiert, mit drei Individuen und drei Gütern und Nutzenfunktionen, die alle „schönen“ Eigenschaften, die man sich so wünschen kann, erfüllen, wo aber Allokationen (die dazu noch Positionen eines allgemeinen Gleichgewichts sein können) existieren, zwischen denen gemäß dem Scitovsky-Kriterium ein Zyklus auftritt, d. h. $X \succ_s Y \succ_s Z \succ_s X$.

Azyklizität ist nun aber eine Mindestvoraussetzung für die rationale Wahl auf der Basis von binären Relationen, auf die kaum verzichtet werden kann. Azyklizität ist z. B. notwendige und hinreichende Bedingung dafür, daß für eine über S vollständige und reflexive Relation \succeq die Menge der besten Elemente $B(\succeq, E)$ für jede endliche Menge $E \subseteq S$ nicht leer ist. Für unendliche Menge kombinieren bislang vorgelegte hinreichende Bedingungen Azyklizität mit weiteren Anforderungen. [Sen 1986, 1079]

Unabhängig von solchen mathematischen Überlegungen ist das Vorliegen von Zyklen in einer binären Relation mit unserer Vorstellung von rationaler Entscheidung und davon, was es heißt, daß ein Zustand „besser“ ist als ein anderer, beim besten Willen nicht mehr vereinbar: Eine Gesellschaft, die einem solchen Entscheidungskriterium folgen wollte, könnte dann auf den Weg einer ewigen Wanderschaft zwischen zyklisch verbundenen Zuständen geschickt werden, und würde sich dabei angeblich immer besser stellen, obwohl sie immer wieder die genau gleichen Zustände durchläuft.

Wie im Fall des Kaldorkriteriums die Möglichkeit der Symmetrie einer als strikt verstandenen (da hinreichender Grund für die Empfehlung einer Veränderung) Relation,

⁵⁶ Gorman [1955]. Für Beispiele sei auf diesen Text oder auch auf Sen [1970, 31] verwiesen.

⁵⁷ Entscheidungsfunktionen die auf intransitiven Relationen strikter Präferenz aufbauen können – um in dem Beispiel von Entscheidungen am Frühstückstisch zu bleiben – zu folgenden Entscheidungen führen: Vor die Entscheidung gestellt zwischen Tee, Kaffee und Kakao zu wählen, wählt unser Frühstückler immer Tee. Bei der Wahl zwischen Tee und Kaffee entscheidet er sich immer für Tee. Bei der Wahl zwischen Kaffee und Kakao immer für Kakao. Bei der Wahl zwischen Tee und Kakao wählt er manchmal Tee manchmal Kakao. Normalerweise würde man solches *Verhalten* wahrscheinlich so zu erklären versuchen, daß für die Entscheidungen des betrachteten Frühstückenden Mitmenschen nicht nur die Präferenzen über die möglichen Getränke, sondern auch situative Elemente (wie z. B. die spezifische Duftmischung, die entsteht, wenn alle drei Getränke auf dem Tisch stehen) eine Rolle spielen. Das Verhalten könnte aber auch das Resultat intransitiver binärer Präferenzen über Getränke sein.

so sollte eigentlich auch die mögliche Zyklizität einer als strikt verstandenen „besser als“ Relation ausreichen, das Scitovsky-Doppelkriterium als Kriterium einer rationalen gesellschaftlichen Entscheidung aufzugeben. Zumal die anderen oben aufgeführten Probleme sich mutatis mutandis wiederfinden. (Was kein Wunder ist, da das Kriterium in einfacher Weise direkt auf dem Kaldorkriterium aufbaut).

Nicht anders als das Kaldorkriterium wird das Kriterium normalerweise als ein Kriterium der binären Wahl verstanden und beruht dann (und immer dann, wenn nicht für alle verglichenen g gilt: $S(g) \subseteq E$) auf einer doppeldeutigen Verwendung des Begriffs der Erreichbarkeit. Gilt dagegen für alle verglichenen g : $S(g) \subseteq E$, dann kann das Kriterium zwischen paretooptimalen g nicht entscheiden und führt deshalb keinen Schritt weiter als das Paretokriterium selbst. Wieder wäre es notwendig, eine Relation der Scitovsky-Indifferenz zu definieren, die dann wohl mehr Elemente aus $P(E)$ als die geschichtlich modifizierte Paretoextensionsregel und weniger als die Paretoextensionsregel selbst als indifferent erklären würde.⁵⁸

Auch das Scitovsky-Kriterium ordnet gesellschaftliche Zustände nicht allein auf der Basis der Ordnung dieser Zustände durch die betroffenen Individuen und ist abhängig von der Definition der Mengen $S(g)$. Die Menge $S(g)$ ist als Menge der durch Umverteilung der Allokation g zu erreichenden Allokationen verstanden, deshalb ist das Kriterium entgegen der Vorstellung z. B. Scitovskys nicht geeignet, das Theorem des Handelsgewinns zu beweisen. Die Relation „besser im Sinne Scitovskys“ ist als strikte Relation der Präferenz zu verstehen und wäre deshalb, wie schon gesagt, evtl. durch eine Relation der Scitovsky-Indifferenz zu ergänzen.

2.4.3 Nochmals Samuelson

Die nächste wichtige Etappe der Entwicklung sogenannter Kompensationskriterien war der Vorschlag von Samuelson [1950]. Ein wichtiger Punkt ist gleich zu Beginn festzuhalten: Für Samuelson folgt aus der Tatsache, daß ein Zustand in seiner Terminologie einem anderen potentiell paretoüberlegen ist, ausdrücklich *nicht*, daß dieser Zustand auch verwirklicht werden soll.⁵⁹

Formal kann die Relation der potentiellen Paretoüberlegenheit (\succeq_{pp}) wie folgt definiert werden:

⁵⁸ Gegen eine entsprechende Entscheidungsregel wären dann ähnliche Einwände zu erheben wie gegen die beiden Versionen der Paretoextensionsregel.

⁵⁹ Welche normative Bedeutung die potentielle Paretoüberlegenheit hat, ist deshalb alles andere als klar. Für sich alleine genommen hat sie in jedem Fall *keine*. Samuelson [1956] entwickelt (am Beispiel einer Familie) die Idee einer „guten Gesellschaft“ (so benannt in Anspielung an die Schlußbemerkung von Samuelson [1956, 22] durch Chipman [1987, 530]), in der faktisch immer so umverteilt wird, daß eine gerechte Verteilung sicher gestellt ist. In einer solchen Gesellschaft wären dann alle Reformen, die zu potentiell paretoüberlegenen Situationen führen, zu unternehmen. Das umgeht natürlich einfach das Problem, da für konkrete wirtschaftspolitische Maßnahmen die Tatsache einer begleitenden gerechten Umverteilung nicht einfach vorausgesetzt werden kann, und wenn man eine solche Umverteilung durchführen möchte, man natürlich wissen muß, wie eine solche Umverteilung aussehen soll. Das aber setzt in der „guten Gesellschaft“ Samuelsons einfach voraus, daß man sich auf eine bestimmte „Wohlfahrtsfunktion“ geeinigt hat.

Seien Y und Y' zwei Mengen von aggregierten Güterbündel, und sei allgemein $A(Y)$ die Menge der durch Verteilung der Güterbündel in Y unter den betroffenen Individuen erreichbaren Allokationen X . Sei $X \in P(A(Y))$ und $X' \in P(A(Y'))$. Dann gelte:

$$X \succeq_{pp} X' : \leftrightarrow \forall X'' \in P(A(Y')) : \exists X^* \in P(A(Y)) : X^* \succeq_p X''$$

Auch hier wieder eine (diesmal schwache) Relation der gesellschaftlichen Präferenz über gesellschaftliche Zustände, genauer über Allokationen definiert. Das habe ich zunächst so belassen, weil Samuelson [1950, 6 f.] zunächst genau das getan hat, zum anderen aber auch, weil die normative Relevanz einer binären gesellschaftlichen Präferenzrelation über gesellschaftliche Zustände verstanden als Entscheidungskriterium klar ist. (Dagegen ist die normative Relevanz einer Präferenzrelation über Mengen von gesellschaftlichen Zuständen alles andere als klar.)

Die Relation \succeq_{pp} ist des weiteren nur für solche Allokationen definiert, die jeweils zur Menge der paretooptimalen Allokationen irgendwelcher vorgegebener Mengen von Allokationen $A(Y)$ gehören. Auch das ist etwas, was faktisch schon so für Samuelsons Einführung dieser Relation in seinem Aufsatz 1950 gilt.

Das eigentliche Problem ist nun dabei die Rolle der $A(Y)$, bzw. noch tieferliegender, der Mengen von aggregierten Güterbündel Y . Es fällt ja auf, daß zwar eine Relation von Allokationen definiert werden soll, diese Allokationen aber im Definiens explizit überhaupt nicht auftauchen. Es ist dabei wichtig zu sehen, daß Samuelson in seinem Aufsatz nur mit Beispielen solcher Mengen arbeitet. Genauer: Er arbeitet in dem Aufsatz von 1950 nur mit zwei solchen Beispielen: Im ersten Beispiel sind die Mengen $A(Y)$ und $A(Y')$ gleich den Mengen $S(X)$ bzw. $S(X')$. Würden wir uns auf solche Fälle beschränken (und wäre es irgendwie klargestellt, welche Allokationen von einer gegebenen Allokation aus durch Umverteilung erreichbar sind), dann würden auch im Definiens die Mengen, über die eine Relation definiert werden soll, explizit auftreten, und das erwähnte Problem würde nicht bestehen. Allerdings verwendet Samuelson seine Definition dann auch schon anhand eines weiteren Beispiels, wo jetzt die Mengen $A(Y)$ und $A(Y')$ auf der Basis von Produktionsmöglichkeitenmengen Y und Y' definiert werden. In späteren Aufsätzen wird Samuelson die Mengen Y zusätzlich noch als (aggregierte) Konsummöglichkeitenmengen bei Freihandel oder bei Handel mit optimalen Zollsätzen verstehen. Das führt zu einem großen formalen Problem: Die Mengen $S(X)$ waren noch von X aus (unter gegebenen Vorstellungen über die Möglichkeiten der Umverteilung) *eindeutig* definiert. Das ist aber schon bei der zusätzlichen Betrachtung von Produktionsmöglichkeitenmengen keineswegs mehr der Fall. Und je mehr ich die Möglichkeiten der Definition der Mengen Y ausweite (und was sollte da eine im voraus festlegbare Grenze sein?), umso weniger ist sichergestellt, daß in der Tat jeder Allokation genau eine Menge $A(Y)$ zugeordnet ist. Damit fällt der Versuch, eine Relation zwischen Allokationen (allgemeiner über gesellschaftliche Zustände) über eine Beziehung von Mengen, zu denen sie gehören, zu definieren, aufgrund schon allein formaler Probleme in sich zusammen.⁶⁰

⁶⁰ In einem präzisen Sinn bemühen sich Ökonomen nicht besonders um die Konsistenz ihrer Theorien: Ökonomische Theorien sind in Wirklichkeit oft Sammlungen von nach mehr oder minder einheitlichen

Daraus wurde die Konsequenz gezogen, das Samuelson-Kriterium nicht mehr als eine Relation zwischen Allokationen (gesellschaftlichen Zuständen) zu verstehen, sondern z. B. als Relation zwischen Mengen von Allokationen oder Mengen von aggregierten Güterbündeln. Ohne formale Probleme ist es dann auch möglich, darauf aufbauend z. B. entsprechende Relationen zwischen geordneten Paaren (X, Y) von Allokationen *und* Mengen von Allokationen und oder von geordneten Tripeln (X, Y, p) von Allokationen, Mengen von Allokationen *und* Preisvektoren („allgemeinen Gleichgewichten“, vgl. Chipman u. Moore [1978]) zu definieren. Allerdings ist es dabei wichtig, daß nicht durch die Hintertür daraus wieder eine Ordnung von Allokationen gemacht wird. Im allgemeinen folgt aus einer Ordnung von Paaren oder von Tripeln keine eindeutige Ordnung der einzelnen Elemente.

In Anlehnung an Chipman u. Moore [1978] ließe sich die Relation potentieller Pareto Superiorität als Relation zwischen Mengen Y von aggregierten Güterbündeln in Parallelität zu dem obigen Definitionsversuch definieren:

$$Y \succeq_{pp} Y' :\leftrightarrow \forall X' \in P(A(Y')) : \exists X \in P(A(Y)) : X \succeq_p X'$$

Wenn wir die Kritikpunkte, die wir gegen das Kaldorkriterium, gegen das Scitovsky-Doppeltkriterium oder gegen das Samuelson-Kriterium vorgebracht haben noch einmal durchgehen, sehen wir, daß keiner dieser Kritikpunkte diese Definition trifft.

Die Definition ist formal korrekt, da in ihr eine Relation zwischen zwei Mengen eben auf Basis einer Beziehung zwischen diesen Mengen definiert wird. Es wird hier eine schwache gesellschaftliche Präferenzrelation definiert. Eine auf der Basis einer schwachen Präferenzrelation definierte strikte Präferenzrelation schließt definitorisch eine Inkonsistenz im Sinne einer Symmetrie aus. Aufgrund der Transitivität der Pareto Relation ist die Relation \succeq_{pp} transitiv und daraus folgend azyklisch. Die Definition einer Relation der „potentiellen Paretoindifferenz“ ergibt sich auf die herkömmliche Weise direkt und problemlos.

Prinzipien gestalteten Modellen. Die Annahmen zwischen einzelnen Modellen können dann erheblich differieren. (Was gehen mich meine Annahmen von gestern an?)

Deshalb ist es vielleicht sinnvoll, darauf hinzuweisen, daß die Strategie, sich bei der Anwendung des Samuelson-Kriteriums als Relation zwischen Allokationen auf solche Modellbeispiele zu beschränken, in denen jeder betrachteten Allokation „zufällig“ gerade genau eine Menge $A(Y)$ entspricht, zwar das Problem in der Tat nicht auftreten läßt. In dem Moment aber, wo daraus irgendwelche wirtschaftspolitischen Empfehlungen abgeleitet werden, also in dem Moment, wo die anhand dieser Modelle hergeleiteten Aussagen normative Bedeutung gewinnen sollen, wird die Frage erlaubt sein, ob man denn in der Tat annehme, daß in der Realität nur solche Mengen von Allokationen zu unterscheiden wären. (Was immer das heißen mag – aber wie man voneinander zu unterscheidende „Situationen“ oder Mengen von Allokationen in der Realität „entdeckt“ oder isoliert, ist vor allem eine Problem für denjenigen, der auf solchen „Situationen“ aufbauend ein Kriterium für die Bewertung von Allokationen herleiten möchte.) Mit dem geschickten Nichtbeachten von „störenden“ Optionen könnte man das Ergebnis der Wohlfahrtsanalyse beeinflussen.

Allokationen, die nicht eindeutig einer Menge $A(Y)$ zugeordnet sind, kommen nur dann sicher nicht vor, wenn alle $A(Y)$ im Verhältnis echter Teilmengen zueinander stehen. Und dann lassen sich anhand des Samuelson-Kriteriums auch eindeutige Wohlfahrtsaussagen machen. Aber für welche Frage sollte ein solches Modell noch relevant sein?

2 Grenzen der Wohlfahrtsökonomie

Die Definition ist (wenn wir an der eingangs gegebenen Definition der $A(Y)$ festhalten) eindeutig, da bei gegebenen individuellen Präferenzordnungen eindeutig festgelegt ist, ob zwei Mengen Y und Y' in dieser Relation stehen oder nicht. D. h. die Zweideutigkeiten, die sich bei den anderen betrachteten Kriterien dadurch ergaben, daß das Vorliegen der entsprechenden Relation zwischen Allokationen X von den wechselnden Definitionen der benutzten $S(g)$ oder Y (und in Folge $A(Y)$) abhing, ist vermieden.

Das wirkliche Problem der so definierten Relation ist, daß völlig unklar ist, welche normative Relevanz sie besitzen soll. Die Antwort von Samuelson [1950] haben wir schon zitiert: keine. Wenn die Argumente, die wir angeführt haben, um zu begründen, warum die Menge der paretooptimalen Allokationen keine normative Relevanz besitzt, korrekt sind, dann gelten sie ohne weiteres auch hier: Die Menge der paretooptimalen Allokationen ist ja nur ein Sonderfall, für den die potentielle Paretoüberlegenheit über alle Teilmengen der erreichbaren aggregierten Güterbündel bei Voraussetzung der entsprechenden topologischen Eigenschaften von E immer gelten muß. Verwirklicht werden immer gesellschaftliche Zustände die gleichzeitig „allokative“ Eigenschaften *und* Verteilungseigenschaften besitzen und nicht eine Situation (eine Menge von gesellschaftlichen Zuständen). Selbst wenn aus dem Samuelson-Kriterium eine Ordnung von gesellschaftlichen Zuständen nach Effizienz Gesichtspunkten ableitbar wäre (was nur geht, wenn willkürlich gesellschaftliche Situationen so definiert werden, daß sie alle im Verhältnis der strikten potentiellen Paretosuperiorität stehen) würde aus der Tatsache, daß ein Aspekt eines gesellschaftlichen Zustands gut ist (der allokative), nicht folgen, daß der Zustand als ganzer gut ist.

Samuelson [1950, 11] meinte mit seinem Kriterium die einzige „zufriedenstellende und konsistente Definition“ eines potentiellen Zuwachses an realem Einkommen „within the sphere of the 'new' (relatively *wert-frei*) welfare economics“ geliefert zu haben. Um diesen Anspruch für die in der Tat konsistente zuletzt betrachtete Version seines Kriterium zu überprüfen, fragen wir einmal, ob Kaldors ursprüngliches Argument für sein eigenes Kriterium hier noch anwendbar ist. Das wäre nur der Fall, wenn wir uns auf Situationen beschränken würden, die mit den Mengen $S(g)$, also mit den Mengen der von einem bestimmten Zustand aus durch Umverteilung erreichbaren Zustände, beschreibbar sind. Denn hier wäre es dann in der Tat möglich – im Sinne des Arguments von Kaldor – nach der Verwirklichung eines potentiell paretoüberlegenen Zustandes von diesem aus, wenn gewünscht, einen auch nach Verteilungsgesichtspunkten dem ursprünglichen Zustand überlegenen zu erreichen. Die Beschränkung auf die Betrachtung wäre gegenüber der von Samuelson intendierten Anwendung seines Kriteriums eine erhebliche Einschränkung. Trotzdem wäre selbst in diesem für Kompensationskriterien sicher besten Fall eine Verwirklichung des entsprechenden Zustands allein aus ökonomischen Gründen nicht zu empfehlen. Solange nicht (hinreichend) sichergestellt ist, daß ein unter Gerechtigkeitsgesichtspunkten akzeptabler Zustand auch in der Tat verwirklicht werden wird, gibt es für den, der solche Gesichtspunkte berücksichtigt haben möchte, keinen Grund der entsprechenden Veränderung zuzustimmen. Das zeigt deutlich, daß eine Empfehlung einer solchen Veränderung weitgehende Werturteile im-

pliziert.⁶¹

Einen grundsätzlichen Denkfehler haben alle Kompensationskriterien, und die normative Verwendung der Zugehörigkeit/Nichtzugehörigkeit zur Menge der paretooptimalen Zustände, oder gar der Nähe/Ferne zu Zuständen dieser Menge gemeinsam: die Vermischung von Kriterien der Bewertung und der Beschreibung der Menge, aus der gewählt werden soll.

Die Unterscheidung ist Ökonomen doch vollkommen klar, wenn sie es mit Haushalten oder Unternehmen zu tun haben: Da haben wir auf der einen Seite die Präferenzordnung bzw. Nutzenfunktion und auf der anderen die Budgetmenge, oder wir haben auf der einen Seite die Gewinnfunktion und auf der anderen die Produktionsmöglichkeitenmenge. Wie Haushalte oder wie Unternehmen verschiedene Zustände bewerten, ist vollständig durch die Präferenzordnung, bzw. durch die Gewinnfunktion bestimmt. Budgetmenge und Produktionsmöglichkeitenmenge geben an, was zu bewerten ist und tragen nicht selbst noch einmal zur Bewertung bei. Kompensationskriterien sind nun der explizite Versuch und die Empfehlung der Verwirklichung von Effizienz zumindest der implizite Versuch, aus der Kombination von Bewertungskriterien und Beschreibung der Menge, aus der gewählt wird, weitere Bewertungskriterien abzuleiten. Das muß daneben gehen.

Niemand käme auf die Idee, aus der Tatsache, daß bei monotonen Nutzenfunktionen Haushalte ihr Optimum auf der Budgetgeraden (Budgetgrenze) finden werden, abzuleiten, daß alle Punkte auf der Budgetgrenze im Sinne des betrachteten Haushalts besser seien, als Punkte innerhalb der Budgetmenge. Punkte auf der Budgetgrenze können so schlecht sein wie nur irgendeiner. Die Intuition, die hinter den Kompensationskriterien liegt, könnte nun die sein, daß man doch immerhin sagen kann, daß ein Haushalt bei größerem Einkommen in jedem Fall sich besser stehen wird, d. h. die Verschiebung nach außen der Budgetgrenze bedeutet für einen Haushalt in jedem Fall eine Verbesserung. Das ist korrekt, weil wir annehmen, daß der betroffene Haushalt „frei“ unter den Elementen der Budgetgrenze wählen kann, und den für ihn optimalen Zustand wäh-

⁶¹ Zwei Anmerkungen sind hier vielleicht noch hilfreich. Die erste betrifft einen wesentlichen Unterschied in der Intuition Kaldors und in der Samuelsons. Kaldor betrachtet in seinem Argument ex post Umverteilungen, was seinem Argument für die Trennbarkeit von allokativen und distributiven Aspekten seine Stärke gibt, aber sein Kriterium (in der im Text betrachteten konsistenten Form) sehr anfordernd macht. Samuelson geht – wie für einen neoklassischen Ökonomen natürlich – von ex ante Umverteilungen aus. Das gibt seinem Kriterium einen weiteren Anwendungsbereich (so läßt sich etwa eine potentielle Paretoüberlegenheit von Freihandel zeigen), auf der anderen Seite kann man von einer Trennung allokativer und distributiver Aspekte nicht mehr sprechen, da nur die Kenntnis der allokativen Prozesse die Bestimmung der notwendigen Ausgangsverteilung zuläßt.

Samuelson [1956, 12 ff.] läßt vielleicht daran zweifeln, daß Samuelson in der Tat von Umverteilungen der Ausgangsverteilungen ausgeht. Man beachte aber, daß Samuelson dort in einem dynamischen Kontext argumentiert. Wenn er verlangt, daß die Umverteilung immer unmittelbar vor dem Zustandekommen des „endgültigen“ Gleichgewichts stattfinden soll, möchte er nichts anderes, als sicherstellen, daß die für dieses Gleichgewicht relevante Ausgangsverteilung korrekt bestimmt wird.

Die zweite Anmerkung betrifft eine sehr begrenzte, wirklich relativ wertfreie negative Anwendbarkeit der konsistenten Form des Samuelson-Kriteriums. Angenommen aus explizit wertgeladenen normativen Kriterien heraus wird eine Reform vorgeschlagen, die die Nutzenmöglichkeitenmenge nach außen verschiebt. Dann macht es sicher keinen Sinn, diese Reform aus rein ökonomischen Effizienzargumenten heraus *abzulehnen*.

len wird. Das würde der Vision der „guten Gesellschaft“ Samuelsons entsprechen, wo angenommen wird, daß die Gesellschaft „frei“ aus der Menge der Nutzenmöglichkeitsmenge den im Sinne der Gesamtwohlfahrt besten Zustand wählen wird. Daß das eine sinnvolle Annahme als Grundlage wirtschaftspolitischer Maßnahmen sei, wird aber sicher weder von denen angenommen werden, die meinen, man könne sinnvoll von so etwas wie Gesamtwohlfahrt reden (denn die haben dann ein Kriterium, nach dem wirtschaftspolitische Maßnahmen inklusive der sie begleitenden Umverteilungsmaßnahmen mal die Situation verbessern, mal verschlechtern), noch von denen, die die Rede von einer Gesamtwohlfahrt grundsätzlich ablehnen.

Von einer skeptischen Position aus mag man nun aber einwenden, daß aus der „Tatsache“, daß wir überhaupt nur wissen können, ob Individuen sich verbessern oder nicht, keinerlei Vergleich, keinerlei Abwägen zwischen Individuen dagegen möglich sei, folge, daß der obige Vergleich hinkt. Man müsse vielmehr von einer Situation ausgehen, wo ein bestimmtes Individuum *nur* wisse, daß es eine monotone Nutzenfunktion habe. Dann solle man ihm doch in jedem Fall empfehlen, irgendeinen Punkt auf der Budgetgrenze zu wählen. *Wenn* ein Konsument wirklich nur weiß, daß er eine monotone Nutzenfunktion besitzt, dann würde in der Tat das Argument vom unzureichenden Grund dafür sprechen, alle Punkte der Nutzenfunktion als indifferent zu betrachten. Ein bißchen realistischer wäre es aber doch anzunehmen, daß ein vernünftiger Konsument aus der Tatsache, daß er verschiedene Güter überhaupt attraktiv findet, folgern wird, daß „wahrscheinlich“ ein ganz einseitiges Güterbündel auf der Budgetgrenze ihm nicht sehr behagen wird. Er würde deshalb ein irgendwie „ausgeglichenes“ Güterbündel wählen. D. h. die Vorstellung, ein Konsument könne überhaupt nur wissen, daß er eine monoton steigende Nutzenfunktion habe, ist *sehr* unrealistisch. Wie verhält es sich im Fall einer Gesellschaft? Die Vorstellung, wir könnten überhaupt nur wissen, daß die Wohlfahrt einer Gesellschaft sich verbessere, wenn einzelne Individuen sich verbessern, ist genauso unrealistisch. Das Werturteil, daß halbwegs ausgewogene Verteilungen besser sind, als kraß ungleiche, daß etwa ein Zustand, in der es eine beträchtliche Zahl von Reichen, eine Mehrheit von gut versorgten Menschen, eine kleine Gruppe von Armen und praktisch keine Verhungerten gibt, besser ist als ein Zustand einer kleinen Minderheit von Superreichen und einer Masse von Hungernden, ist sicher genauso verbreitet, oder sogar weiter verbreitet, als das Werturteil, daß Paretoverbesserungen die Welt verbessern. Werturteile sind beide, und welche Position auch immer man in der Frage der Begründbarkeit von Werturteilen man vertritt, die Behauptung, das Paretoprinzip ließe sich begründen, das erwähnte „Verteilungsurteil“ dagegen nicht, ist inkonsequent.

Noch schwieriger ist die Sache im Fall der Kompensationskriterien. Für die Beurteilung von Situationen oder gar von Zuständen ist ja die Abgrenzung von „Situationen“ entscheidend. „Situationen“ liegen aber nicht einfach in der Realität vor, auch nicht einfach in „positiven“ Modellen, sondern sie sind Ergebnis der klassifikatorischen Arbeit von Ökonomen. Damit haftet ihnen ein gewisser Grad von Willkür an. Einen Zustand besser zu nennen, weil er auf der Nutzengrenze einer bestimmten Situation liegt, mißt einer bestimmten Einteilung, Beschreibung von Realität normative Bedeutung bei. Am Beispiel von Samuelson [1950] und Samuelson [1962] sei dies noch einmal gezeigt:

Samuelson beginnt den ersten Artikel mit Situationen, die durch die Umverteilung gegebener Güterbündel definiert sind, und schließt daraus, daß, mehr Güter zu haben, „potentiell pareto besser“ ist. Dann verwirft er diese Situationen als die relevanten, mit der Begründung, daß Güterbündel nicht vom Himmel fallen, sondern produziert werden. Auf Situationen definiert auf der Basis von Grenzen von Produktionsmöglichkeitenmengen komme es also an: Damit sind mehr Güter nicht mehr notwendig besser als weniger Güter. [Samuelson 1950, 6 f. u. 12 f.] In seinem Beitrag 1962 vergleicht er vier Kategorien von Situationen: wieder die durch Umverteilung von unter Autarkie, bei Freihandel und bei Handel unter Optimalzöllen erreichbaren festen Güterbündel definierten, die er kurzerhand für irrelevant erklärt; dann die den Nutzenmöglichkeitenmengen bei Autarkie, bei Freihandel und bei Handel unter Optimalzöllen entsprechenden. Was ist da nun der normativ relevante Vergleich? Die Nutzenmöglichkeitsgrenze bei Optimalzöllen liegt am weitesten außerhalb. Hätte Samuelson den Aufsatz also besser mit „The Gains from Optimal Tariffs Again“ überschreiben sollen? Warum halten viele Ökonomen, die Tatsache, daß die Nutzengrenze bei Freihandel überall außerhalb der Nutzengrenze bei Autarkie liegt, für ein gutes Argument für Freihandel, daß die Nutzengrenze bei Optimalzöllen überall außerhalb der Nutzengrenze bei Freihandel liegt, aber für kein überzeugendes Argument für Optimalzölle? Ein Grund könnte sein, daß das Erreichen der Nutzengrenze bei Optimalzöllen verlange, daß für die gegebene Verteilungssituation auch der richtige Zoll erhoben werden muß, eine Bewegung entlang der Nutzengrenze verlange also nicht nur eine Veränderung der Ausgangsverteilung sondern zusätzlich die Bestimmung des richtigen Zollsatzes. Nun ist das eine inhärent keine komplexere Aufgabe, was die erforderlichen Informationen und die notwendigen Berechnungen betrifft, als das andere. Auch die Implementierung eines Satzes optimaler Zollsätze dürfte inhärent keine schwierigere staatliche Aufgabe sein als die Durchführung einer optimalen Umverteilung. Aber selbst wenn es schwerer wäre, optimale Zollsätze zu implementieren als eine optimale Umverteilung, wäre das dann ein Argument für Freihandel ohne Zölle? Statt eines Zollsatzes von 0 für alle Güter könnte eine Regierung irgendeinen Zollsatz erheben. Und da es einen Bereich von Zollsätzen gibt, für die die damit verbundenen Nutzenmöglichkeitsgrenzen, die Nutzenmöglichkeitsgrenze bei Freihandel schneiden würde, *könnte* damit evtl. sogar ein paretoüberlegener Zustand erreicht werden. Warum also nicht irgendeinen Zollsatz erheben? Oder genauer: Warum für einen Abbau von faktisch bestehenden Zöllen eintreten, wenn also in keiner Weise sicher ist, daß das zu einer Paretoverbesserung führen wird? Mit welchen Argumenten Verteidiger des Freihandels auf diese Frage reagieren werden, soll hier nicht weiter verfolgt werden.⁶² Wenn man mit potentieller Paretoüberlegenheit argumentieren möchte, dann muß man in jedem Fall sowohl die Situation der Optimalzölle als auch alle Situationen nicht optimaler Zölle, deren Nutzengrenze die Nutzengrenze bei Freihandel schneidet, mit irgendwelchen Gründen als für den Vergleich irrelevant ausschließen. Ein Zollsatz von Null auf alle Güter ist ana-

⁶² Die Ausführungen im Text verstehen sich nicht als ein Beitrag zur Diskussion über die beste Außenhandelspolitik. Sie sollen nur an einem Beispiel zeigen, daß die Frage der Auswahl der zum Vergleich herangezogenen Situationen nicht nur theoretisch relevant ist, sondern verbreitete normative Überzeugungen von Ökonomen betrifft.

lytisch von einem anderen Zollsatz nicht prinzipiell verschieden, nur außerhalb der Argumentation über die potentiell Paretoüberlegenheit oder Effizienz liegende Gründe, lassen diesen Zollsatz von 0% als etwas besonderes erscheinen.⁶³

2.5 Unproblematisch?

Die Darstellung der Konzepte des Paretokriteriums, Paretoprinzips oder der Paretoeffizienz werden häufig von Qualifikationen begleitet, die diese Konzepte in irgendeiner Weise als unproblematisch oder gar selbstverständlich erscheinen lassen sollen. Dabei lassen sich zwei Grundpositionen unterscheiden.

Vor allem in der frühen Diskussion dieser Konzepte wurde der Anspruch erhoben, daß diese Konzepte Platz hätten innerhalb der Ökonomie als einer positiven Wissenschaft, weil es sich bei ihnen um wertfreie Konzepte handeln würde. Diesen Anspruch scheint zumindest implizit schon Pareto selbst für seine Konzept der „maximalen sozialen Ophelimität“ erhoben zu haben, das mit dem heutigen Konzept der Paretoeffizienz übereinstimmt.⁶⁴

⁶³ Dies gilt jedenfalls, solange Handelsgewinne aus der Perspektive eines einzelnen Landes betrachtet werden.

⁶⁴ Pareto [1972] kündigt in einem einleitenden Kapitel an, daß er sich in diesem Werk ausschließlich mit ausschließlich wissenschaftlichen Fragen beschäftigen werde (I,1). Ausdrücklich schließt er dann alles, was den Charakter einer Vorschrift besitzt, also alles was normativen Charakter hat, aus dem Bereich der Wissenschaft aus: „Anything which has the appearance of a precept is not scientific [...]“ (I, 39) In einem der Zentralkapitel seines Werkes („Economic Equilibrium“) gibt er dann unter der Überschrift „Properties of equilibrium“ (VI, 32) seine Version dessen, was wir heute das Erste Wohlfahrtstheorem nennen: „For phenomena of Type I, when equilibrium takes place at a point where the indifference curves of the contracting parties are tangent, the members of the collectivity under consideration enjoy maximum ophelimity.“ (VI, 34) Die erläuternden Ausführungen zu Gleichgewichten mit Produktion und zu Gleichgewichten in kollektivistischen Gesellschaften (VI, 48-61) zeigen dann aber deutlich, daß für Pareto „maximale Ophelimität“ ein erstrebenswertes Ziel ist.

Noch deutlicher ist der normative Charakter des Begriffs der „maximalen Ophelimität“ im „Traité de sociologie générale“, wenn Pareto den *ökonomischen* Begriff der Ophelimität für eine Gesellschaft darstellt: „Quand la collectivité se trouve en un point Q dont elle peut s'éloigner à l'avantage de tous les individus, en leur procurant à tous de plus grandes jouissances, il est manifeste qu'au point de vue économique et si l'on ne recherche que l'avantage de tous les individus qui composent la collectivité, il convient de ne pas s'arrêter en un tel point, mais de continuer à s'en éloigner tant que c'est à l'avantage de tous.“ [Pareto 1917, § 2129]

Diese Situation läßt verschiedene Interpretationen zu: Die im Text adoptierte nimmt die Ankündigung Paretos ernst, und unterstellt ihm deshalb, daß er „maximale Ophelimität“ als ein (in seinem Sinne) rein wissenschaftliches Konzept betrachtet hat. Eine mögliche Erklärung dafür könnte es sein, daß er für dieses Konzept die Möglichkeit in Anspruch nehmen würde, die er in I, 40 beschrieben hat. Bei einer präskriptiven Aussage wie „Der Staat soll kein schlechtes Geld emittieren.“ könnte es sich danach um eine elliptische Aussage halten, „and, in this case, it could become scientific by removing the ellipsis. For example, if it were said that the state *should not* issue bad money if obtaining maximum utility for society is desired, and if what is meant by this *maximum utility* were defined by means of facts, the proposition would become amenable to experimental verification and consequently would become a scientific proposition.“ Maximale Ophelimität wäre demnach eine Eigenschaft einer Gesellschaft, die zumindest prinzipiell einer experimentellen Überprüfung fähig wäre, und damit würden auch Aussagen, die etwa Bedingungen der Erreichung maximaler Ophelimität ausdrücken, wissenschaftlichen Charakter tragen. „Maximale Ophelimität“ gehörte zu den Worten, „which clearly correspond to well

Auch Autoren wie Kaldor, Hicks und Scitovsky waren noch davon überzeugt, daß die von ihnen vorgeschlagenen Kriterien und a fortiori das Paretokriterium in den Rahmen objektiver Wissenschaft gehören würden.⁶⁵

Offensichtlich standen diese Autoren unter dem Eindruck der Äußerungen Robbins, der den „unwissenschaftlichen“, nicht objektiven Charakter von Aussagen der Wohlfahrtsökonomie unter anderem damit begründet hatte, daß solche Aussagen den Vergleich der Wohlfahrt verschiedener Individuen voraussetzen würden. Ein solcher Vergleich aber sei wissenschaftlich nicht möglich [Robbins 1984, 136-141]. Da ihre Kriterien und a fortiori das Paretokriterium einen solchen Vergleich nicht (explizit) voraussetzen, schien ihnen deren wissenschaftliche Charakter gesichert.⁶⁶ [Kaldor 1939 u. Hicks 1939; im Nachdruck Arrow u. Scitovsky 1969, 388 u. Hicks 1981, 63 f.]

Während also diese Autoren eine scharfe Grenze gezogen hätten, zwischen Urteilen auf der Basis ihrer Kriterien und des Paretokriteriums, die sie als objektive Urteile in den Bereich der Ökonomie aufnehmen würden, und anderen weitergehenden Wohlfahrtsurteile, die etwa auch Verteilungsfragen berücksichtigen, sind sich modernere Autoren wohl i. d. R. bewußt, daß auch das Paretokriterium ein Werturteil ist (oder auf Werturteilen basiert). Was das Paretokriterium in den Augen modernerer Autoren von anderen Werturteilen unterscheidet (und was es als Basis eines ganzen Zweiges der Ökonomie verwendbar erscheinen läßt), ist seine behauptete Akzeptanz. Immer wieder liest man, daß dieses Kriterium nicht kontrovers, nicht umstritten, allgemein akzeptiert oder akzeptierbar sei.⁶⁷ Für viele Ökonomen scheint das Paretokriterium so

defined real things“ (I, 1). Wenn man diese Bemerkungen Paretos so versteht, daß er damit die Untersuchung von Mittel-Zweck Beziehungen in der Bereich wissenschaftlicher Arbeit aufnehmen möchte, ohne daß damit eine Wertung der entsprechenden Zwecke verbunden wäre, wären aber damit zumindest zwei Probleme verbunden. Zum einen muß man die Ausführungen VI, 48 ff. doch wohl so verstehen, daß Pareto selbst einen Zweck (die Erreichung maximaler Ophelimität) als erstrebenswert voraussetzt, was auch unter Berücksichtigung von I, 40 in einer wissenschaftlichen Arbeit nach seinen Ansprüchen keinen Platz hätte. Dem Anspruch „never to use words which may bias the reader’s mind“ (I, 1), ist er mit diesen Absätzen wohl nicht gerecht geworden. Zum anderen lassen auch andere Zielvorstellungen eine „naturalistische“ Interpretation zu. Es wäre dann zu fragen, warum Pareto und in seiner Folge die moderne Wohlfahrtstheorie sich so ausschließlich mit maximaler Ophelimität bzw. Effizienz beschäftigten. Warum hat sich keine ähnlich bedeutsame Literatur entwickelt, die etwa nach den Bedingungen einer Einkommensverteilung nach dem Rawlsschen Differenzprinzip fragt?

⁶⁵ Kaldor [1939] (im Nachdruck: Arrow u. Scitovsky [1969, 389]) hält den wissenschaftlichen Status von Empfehlungen, die sich auf allokativen Fragen beziehen, für unbezweifelbar. Der Aufsatz ist eine Antwort auf Robbins [1984] (ursprünglich 1932 erschienen), dessen Leugnung der Objektivität von Aussagen der Wohlfahrtsökonomie zitiert wird. Ähnlich Hicks [1939] (im Nachdruck: Hicks [1981, 75]): „By adopting the line of analysis set out in this paper, it is possible to put welfare economics on a secure basis, and to render it immune from positivist criticism.“ Scitovsky [1993, 73] erklärt direkt: „[...] we can make objective statements about efficiency.“

⁶⁶ Damit „übersahen“ sie aber Robbins (a.a.O.) Ausführungen 142 ff., in denen dieser (in der Tradition Max Webers) zu zeigen versuchte, daß selbst wenn ein wissenschaftlicher Vergleich von Nutzen möglich wäre, der „logische Abgrund“ (148) zwischen normativen und positiven Urteilen weiterbestünde.

⁶⁷ Sohmén [1976, 30 f.]: „Auch bei diesem Kriterium [dem Paretokriterium] handelt es sich nicht um eine objektive, wissenschaftlich begründbare Aussage, sondern um ein Werturteil, das jeweils nur die subjektive Ansicht einzelner (gegebenenfalls auch aller) Betrachter verkörpern kann. Wegen seiner vergleichsweise unverfänglichen Aussage dürfte es sich allerdings um ein Werturteil handeln, das relative gute Chancen hat, von der überwiegenden Mehrzahl akzeptiert zu werden.“ Harsanyi [1978,

selbstverständlich zu sein, daß es nicht notwendig erscheint, das Kriterium ausführlicher zu begründen, oder genauer zu bestimmen, was mit der Akzeptanz des Pareto-kriteriums gemeint ist. Da auf dem Paretokriterium aber doch – dem Anspruch nach – ein beachtliches Teilgebiet der modernen Ökonomie aufbaut, ist es vielleicht doch sinnvoll, eine größere Klarheit in diesen Fragen zu suchen. Eine größere Klarheit kann dabei in zwei Aspekten angestrebt werden. Man kann einmal fragen, für was eigentlich so große Akzeptanz beansprucht wird. Der Beantwortung dieser Frage galt vor allem der Abschnitt 2.2.1, wo ich versucht habe, klar zwischen dem Paretokriterium, Paretoprinzip und dem Paretoextensionskriterium zu unterscheiden. Das Paretokriterium in der von mir vorgeschlagenen Fassung wäre noch am ehesten etwas, für was man ziemlich allgemeine Akzeptanz behaupten könnte. Zum anderen könnte man fragen, in welchem Sinn man von allgemeiner Akzeptanz sprechen kann: Wer müßte das Kriterium dafür akzeptieren und aus welchen Gründen? Dieser Frage dienen vor allem die jetzt folgenden Überlegungen.⁶⁸

Eine erste Möglichkeit wäre es, für das Paretokriterium Evidenz zu beanspruchen. Das setzt eine epistemologische Position voraus, die zumindest die Möglichkeit solcher evidenten Einsichten nicht ausschließt, die dann keiner weiteren Begründung mehr bedürfen (und ihrer vielleicht auch gar nicht fähig sind). Als evident sollen dabei Aussagen gelten, die von allen Verständigen nach ausreichender Beschäftigung mit der Materie als wahr und als keiner eigentlichen Begründung bedürftig akzeptiert würden. Eine evidente Aussage ist eine Aussage, die jeder vernünftige Mensch, wenn er sie versteht, auch für wahr hält.⁶⁹

Der Anspruch der Evidenz für irgendeine Aussage ist aber immer dann problematisch, wenn es Leute gibt, denen die entsprechende Aussage keineswegs als evident erscheint, sondern vielmehr als der Begründung bedürftig oder gar als falsch. Denn was bedeutet es, den Anspruch der Evidenz zu erheben, um eine Aussage gegen ihre

226] vom Paretoprinzip (bei ihm „Pareto optimality“): It „is a moral principle – but it is surely a rather noncontroversial moral principle.“ Varian [1992, 405]: „If each individual prefers x' to x , it seems noncontroversial to assert that x' is 'better' than x and any projects that move us from x to x' should be undertaken. This is the **Pareto criterion**.“ Böventer u. Illing [1995, 256 f.]: „Um überhaupt einen konsensfähigen Wohlfahrtsmaßstab zu erhalten, haben sich die Ökonomen auf ein Kriterium geeinigt, das Konflikte in der Beurteilung von verschiedenen Zuständen ausschließt. Die Norm, auf der ein Großteil der Wohlfahrtstheorie fußt, das **Pareto-Kriterium**, ist eine Art Minimalkonsens. [...] Der große Vorteil eines solchen **Minimalkonsenses** ist, daß er zu jedem Zeitpunkt und bezüglich jeder Fragestellung für alle Beteiligte akzeptabel ist. Auch Nicht-Ökonomen werden am Pareto-Kriterium wenig auszusetzen finden. [...] Alles in allem ist das Pareto-Kriterium und mithin auch das Konzept des Pareto-Optimums [der Paretoeffizienz in der hier verwendeten Terminologie A.G.] wohl in der Tat konsensfähig.“

⁶⁸ Die Fragen hängen allerdings eng zusammen.

⁶⁹ Von diesem erkenntnistheoretischem Begriff der Evidenz sollte man sorgfältig subjektive, gruppenspezifische, kulturelle etc. „Evidenzen“ unterscheiden, die aber eben gerade nicht den Anspruch erheben können, allgemein akzeptiert zu werden. „Evidenz“ in diesem zweiten Sinn sagt etwas über die *faktische* Akzeptanz einer Aussage in einer bestimmten Gruppe. „Evidenz“ im eigentlich erkenntnistheoretischen Sinn ist mit dem *Anspruch* verbunden, daß *jeder Verständige* nach ausreichender Beschäftigung mit der Materie ein entsprechendes Urteil akzeptieren würde. Daß es durchaus legitim ist, von solchen subjektiven Evidenzen in der Argumentation auszugehen, ist Teil der hier vertretenen epistemologischen Position. (S. u. Kapitel 4.) Allerdings reicht die Überzeugungskraft einer solchen Argumentation nicht weiter als die Verbreitung der entsprechenden subjektiven „Evidenz“.

Leugner zu verteidigen, oder Evidenz dann zu behaupten, wenn sie von anderen in Frage gestellt wird? Der Anspruch der Evidenz kann dann nur aufrecht erhalten werden, wenn man entweder den Leugnern der Evidenz unterstellt, sie hätten die umstrittene Aussage noch nicht verstanden (was natürlich wahr sein *kann*), oder wenn man sie aus der zu berücksichtigenden Kommunikationsgemeinschaft vernünftiger Menschen ausschließt. Evidenz für eine Aussage zu beanspruchen, die geleugnet wird, läuft deshalb auf den Abbruch eines gleichberechtigten Dialogs hinaus: Sei es, daß der Leugner aus der zu berücksichtigenden Kommunikationsgemeinschaft überhaupt ausgeschlossen wird, sei es daß er nur noch zum Adressaten von Belehrungen wird, die ihm helfen sollen, das was er (angeblich) nicht versteht, zu verstehen, das was er nicht sieht, zu sehen.

Ist das Paretokriterium evident? Es scheint mir, daß keine Fassung des Paretokriteriums als evident gelten kann. Das ist deshalb der Fall, weil in jeder Fassung des Paretokriteriums sowohl die Menge der zu berücksichtigenden Individuen als auch der zu berücksichtigenden Präferenzen bestimmt werden muß. Ohne die entsprechenden Festlegungen liegt noch überhaupt kein Entscheidungskriterium vor. Ich sehe aber nicht, wie diese Festlegungen auf eine evidente Art getroffen werden könnten. Es gibt doch eine ganze Reihe von Möglichkeiten, unter denen man weniger aufgrund einer Evidenz und eher auf der Basis der abwägenden Vernunft wählen wird. Evidenz für eine der Fassungen des Paretokriteriums kann man deshalb nur beanspruchen, wenn man bereit ist, in der beschriebenen Weise Vertreter verbreiteter und/oder allgemein als diskussionswürdig betrachteter ethischer Positionen aus der Gemeinschaft gleichberechtigter Diskussionspartner auszuschließen.⁷⁰

Nehmen wir als Beispiel eine Fassung des Paretokriteriums, in der die in der Terminologie Harsanyis persönlichen Präferenzen aller jetzt lebenden Menschen zu berücksichtigen seien unter ausdrücklichem Ausschluß von Präferenzen, die sich in anderer Leute Angelegenheiten einmischen (um das liberale Paradox zu verhindern). Es ist durchaus plausibel, daß in dieser Interpretation das Paretokriterium eine zu wirkwirkende Paretoverbesserung anzeigt, wenn Kinderprostitution unter bestimmten Bedingungen erlaubt würde.⁷¹ Es ist aber eine sicher weit verbreitete (und wie ich meine korrekte) Überzeugung, daß Kinderprostitution eine nicht akzeptierbare Form der Ausbeutung von Kindern ist, die deshalb mit den Mitteln der staatlichen Gewalt soweit als möglich verhindert werden sollte. Die hier betrachtete Fassung des Paretokriteriums ist deshalb wohl nicht nur nicht evident, sondern falsch.

Man kann nun in verschiedenen Weisen versuchen, das vorgeschlagene Kriterium zu verbessern, so daß es nicht mehr Aufgabe der strafrechtlichen Verfolgung von Kinderprostitution verhindert. Zu welcher Fassung des Paretokriteriums man auch immer von hier her kommt, diese Fassung kann nicht als evident gelten, wenn sie (unter anderem) damit *begründet* wird, daß sie nicht die Erlaubnis von Kinderprostitution verlangt. Aber wie könnte eine solche Fassung des Paretokriteriums aussehen? Eine Möglichkeit

⁷⁰ Das gilt a fortiori für das Paretoprinzip und für Einstimmigkeitsforderungen bei politischen Entscheidungen.

⁷¹ Angesichts der Notsituation, in der sie sich befinden, kann für manche Jugendliche Prostitution subjektiv wie ein (zeitweiliger) Ausweg aus dem Elend erscheinen.

wäre sicher die von mir unten dargestellte, in der die Präferenzen (z.B.) aller lebenden Menschen, so wie sie sind, d.h. auch wenn sie ethische Präferenzen sind und auch wenn sie sich in die Angelegenheiten anderer einmischen. Da viele Menschen Kinderprostitution für verabscheuenswert halten, wird aufgrund dieses Kriteriums Kinderprostitution sicher nicht erlaubt werden. Aber auch diese Form der Bestimmung der zu berücksichtigenden Individuen (hier: alle jetzt lebenden Menschen), ist nicht evident: Denn eine andere Lösung des Problems besteht gerade darin, mit dem Hinweis auf die fehlende Volljährigkeit die Anwendbarkeit des Paretokriteriums auf das Problem der Kinderprostitution zu leugnen, d. h. das Paretokriterium nur für Erwachsene gelten zu lassen.⁷²

Die Wortwahl der Autoren, die für das Paretokriterium behaupten, es sei nicht kontrovers, läßt meistens nicht vermuten, daß diese Autoren für das Paretokriterium im strengen Sinn Evidenz beanspruchen wollen. Drei mögliche Interpretationen liegen wesentlich näher: 1. Das Paretokriterium wird faktisch von der Mehrheit (oder gar allen Mitgliedern) einer bestimmten Gruppe (der Ökonomen) akzeptiert. 2. Das Paretokriterium würde faktisch von der Mehrheit (oder gar allen) einer bestimmten Gruppe (z.B. der stimmberechtigten Staatsbürger) akzeptiert, wenn es diesen vorgestellt würde. 3. Das Paretokriterium wäre theoretisch in einem bestimmten Sinn allgemeiner Zustimmung fähig.

Die Interpretationen 1. und 2. passen am ehesten zu der unten Ökonomen weitverbreiteten Überzeugung, daß normative Aussagen keiner kognitiven Begründung fähig sind. Das beste, was sich dann noch erreichen läßt, ist eine Art pragmatischer Lösung, die sich aus der faktischen Übereinstimmung von Werturteilen ergibt. *Mehr* läßt sich sowieso nicht erreichen. Deshalb ist es auch kein Problem, daß natürlich kein Urteil (auch kein Werturteil) deshalb *wahr* wird, weil es von einer Mehrheit irgendeiner Gruppe akzeptiert wird. Allerdings ist es selbst eine ethische Position, daß ein solches pragmatisches Kriterium, wie die faktische oder gar nur potentielle faktische Zustimmung einer Mehrheit, hinreichender Grund für die Adoption eines Entscheidungskriteriums für politische Maßnahmen sein kann. Wenn wir die Möglichkeit einer faktischen wirklich allgemeinen Zustimmung zu irgendeiner Fassung des Paretokriteriums einmal als völlig unrealistisch ausschließen, und wenn wir einmal mit einigem Realismus annehmen, daß die faktische Zustimmung auch nur einer Mehrheit der Stimmbürger für das Paretokriterium deshalb nicht vorliegt, weil dieses Kriterium den meisten Stimmbürgern einfach unbekannt ist, können wir uns für die Argumentation auf zwei Möglichkeiten beschränken: 1) Die Mehrheit der Ökonomen stimmt *einer* Fassung des Paretokriteriums zu. 2) Die Mehrheit der Stimmbürger, würde dem Paretokriterium zustimmen, wenn es ihnen vorgelegt würde.⁷³

⁷² Allerdings lassen sich auch Beispiele anführen, die unter der Voraussetzung von Präferenzen, die sich nicht einmischen, das Paretokriterium für die Anwendung unter Erwachsenen ähnlich zweifelhaft erscheinen läßt. Ein (relevantes) Beispiel wäre die Frage der Schuldknechtschaft. Ein weiteres (besonders abstoßendes) Beispiel der Kanibalismus im gegenseitigen Einverständnis.

⁷³ Keine dieser empirischen Behauptungen wurden meines Wissens bisher mit empirischen wissenschaftlichen Methoden überprüft. Ich nehme an, daß in der Tat die Mehrheit der Ökonomen irgendeiner (aber in der Regel verschiedenen) Fassungen des Paretokriteriums faktisch zustimmen, aber nicht *einer*

Im Fall der Möglichkeit 1) wäre also zu fragen, ob die Tatsache, daß die Mehrheit der Ökonomen einer bestimmten Fassung des Paretoprinzips zustimmt, hinreichender Grund dafür ist, daß Maßnahmen, die von dieser Fassung empfohlen werden, auch durchgeführt werden. Das ist doch wenigstens dann eher zweifelhaft, wenn diese Maßnahmen von nach eigener Einschätzung davon Betroffenen abgelehnt werden. Mit einer solchen Möglichkeit ist aber bei allen Fassungen des Paretokriteriums zu rechnen, die nicht auf ein *positives* Einstimmigkeitskriterium hinauslaufen.⁷⁴ Vor allem ist die Möglichkeit nicht auszuschließen, daß in der Berufsgruppe der Ökonomen noch andere normativen Überzeugungen von einer Mehrheit geteilt werden. Es ist nicht einzusehen, wie diese bloße Mehrheit irgendeine Maßnahme gegenüber von ihr nach eigener Einschätzung negativ Betroffenen rechtfertigen sollte.

Etwas komplizierter ist die zweite Möglichkeit. Ganz offensichtlich ist die Praxis der parlamentarischen Demokratie mit der Überzeugung verbunden, daß die faktische Mehrheit der Stimmbürger in bestimmten Fällen ausreicht, auch Entscheidungen zu rechtfertigen, die von einer Minderheit als ihr schädlich empfunden werden. Allerdings gilt das eben nur in bestimmten Fällen (solange etwa Grundrechte nicht dadurch eingeschränkt werden). Eine faktische Mehrheit ist deshalb als solche in vielen Demokratien *nicht hinreichend*, solche Entscheidungen zu rechtfertigen. Außerdem ist die Idee und Praxis der parlamentarischen Demokratie wesentlich mit der Überzeugung verbunden, daß nicht ein stipulierter und *potentieller*, sondern nur ein faktischer und in demokratischen Verfahren ausgedrückter Mehrheitswille, einer Minderheit Lasten auferlegen kann. Vor allem aber gilt folgendes: Das Entscheidungsverfahren der Mehrheitsentscheidung unter bestimmten Grenzen ist ein wesentlich weitergehendes Entscheidungsverfahren als das Paretokriterium. Wenn eine Mehrheitsentscheidung aber ein Entscheidungskriterium rechtfertigen kann, ist überhaupt nicht mehr einzusehen, warum man sich auf irgendeine Form des Paretokriteriums beschränken und nicht gleich das Verfahren der Mehrheitsentscheidung (evtl. unter gewissen Einschränkungen) als Entscheidungskriterium adoptieren sollte.

Es gibt nun aber einen sehr präzisen Sinn, in dem das *Paretokriterium* in einer bestimmten Perspektive in der Tat unproblematisch, und von allen akzeptierbar ist, weil es im Sinne der dritten Interpretation allgemein zustimmungsfähig ist.

Nehmen wir einmal (kontrafaktisch) an, es gebe eine Möglichkeit, Entscheidungskriterien auf eine Weise zu beschließen, daß sie, nachdem sie einmal approbiert wurden, immer auch implementiert würden. Immer also, wenn sie Verwirklichung eines bestimmten gesellschaftlichen Zustands oder einer Maßnahme empfehlen, würde diese Empfehlung auch umgesetzt. Man beachte, daß ein entsprechender Beschluß eines Entscheidungskriteriums gleichzeitig eine teilweise Entmachtung aller anderer gesellschaftlicher Entscheidungsprozesse bedeuten würde. Gibt es nun Entscheidungskriterien, die in dem Sinne unproblematisch wären, daß niemand einen Grund hätte, sich

übereinstimmenden. Ob alle Stimmbürger z.B. der Bundesrepublik Deutschland irgendeiner Fassung des Paretoprinzips zustimmen würden, wenn es ihnen vorgelegt und erklärt würde, scheint mir vor entsprechenden empirischen Untersuchungen völlig offen.

⁷⁴ D. h. ein Kriterium, das Einstimmigkeit als hinreichende aber nicht notwendig notwendige Bedingung für die Durchführung einer Maßnahme erklärt.

einem solchermaßen unwiderruflichem und effektiven Beschluß zu widersetzen?

Es gibt ein solches Kriterium und es handelt sich dabei gerade um das Paretokriterium (wie oben definiert) auf der Basis als vorhanden angenommener faktischer, unbereinigter, möglicherweise auch ethischer und sich einmischender Präferenzen. Es ist klar, warum niemand etwas gegen den unwiderruflichen und effektiven Beschluß des Paretokriteriums haben könnte: Eine Situation, in der das Paretokriterium anwendbar ist und eine Maßnahme vorschlägt, war gekennzeichnet dadurch, daß alle Beteiligten in ihrer Einschätzung des besten gesellschaftlichen Zustandes übereinstimmen.⁷⁵ Gäbe es also eine Abstimmung darüber, ob im gegebenen Fall dieser Zustand verwirklicht werden sollte, dann würden alle in einem minimalen Sinn rationalen⁷⁶ Abstimmungsberechtigte für die Verwirklichung stimmen. Hätten wir also das Paretokriterium unwiderruflich und effektiv beschlossen, so würde *aufgrund* des Paretokriteriums nie ein Zustand verwirklicht oder eine Maßnahme unternommen, die nicht jeder Betroffene selbst durchgeführt hätte, oder die zumindest von jedem Betroffenen als gleich gut eingestuft wird, wie die verwirklichte. Es gäbe niemanden, der je sich über den Beschluß des Paretokriteriums beklagen könnte.⁷⁷

Betrachten wir dagegen eine Situation, in der das Paretokriterium zu keiner Entscheidung führt. Es gibt dann mindestens zwei Individuen, die in ihrer Einschätzung des besten gesellschaftlichen Zustands nicht übereinstimmen. Welcher Zustand auch immer dann verwirklicht wird, es gibt zumindest ein Individuum, das einen anderen vorgezogen hätte. Würde also ein Entscheidungskriterium unwiderruflich und effektiv beschlossen, das in zumindest einer möglichen Situation, in der das Paretokriterium zu keiner Entscheidung führt, eine solche ermöglicht, kann ein Individuum die *Möglichkeit* nicht ausschließen, daß einmal eine Situation auftritt, in der ohne diesen Beschluß ein von ihm vorgezogener Zustand verwirklicht worden wäre. Es ist dann nur noch bei besonderen Annahmen über die Präferenzen der Individuen und über die möglichen Situationen möglich, zu sagen, daß kein betroffenes Individuum einen Grund hätte, einem solchen Beschluß zu widersprechen.

Wenn wir das eben geschilderte Gedankenspiel als Maßstab dafür akzeptieren, daß eine Entscheidungsregel wirklich unproblematisch und allgemein zustimmungsfähig ist, dann legt die Affirmation der allgemeinen Zustimmungsfähigkeit des Paretokriteriums gleichzeitig einige Aspekte der Interpretation des Paretokriteriums fest.

Es ist z.B. umstritten, welche Art von Präferenzen bei der Verwendung des Paretokriteriums zugrunde gelegt werden sollte. So wurde etwa in der Diskussion um das Liberale Paradox vorgeschlagen, bei der Anwendung des Paretokriteriums keine Präferenzen zuzulassen, die sich auf bestimmte Weise in die Angelegenheiten anderer ein-

⁷⁵ Genauer: Es gibt mindestens einen Zustand, der Element der Menge der besten gesellschaftlichen Zustände aller betroffenen Individuen ist.

⁷⁶ Rational heißt hier nicht mehr, als daß ein Wähler in einer Abstimmung für die Alternative stimmt, die er vorzieht.

⁷⁷ Das Kriterium für die allgemeine Zustimmungsfähigkeit eines Entscheidungskriteriums ist gewissermaßen ein Maximalkriterium, da es fragt, ob ein Entscheidungskriterium akzeptiert werden könnte, wenn es immer implementiert würde. Das Paretokriterium würde diesen Test aber auch dann passieren, wenn wir etwas realistischer davon ausgingen, daß es nur manchmal angewandt würde, z.B. weil die Bedingungen seiner Anwendbarkeit nicht immer erkannt würden.

mischen. Eine ganze Reihe von Regeln kollektiver Entscheidungen sind nur dann paretoinklusiv, wenn das Paretokriterium dieselben Präferenzen unberücksichtigt läßt, wie die entsprechende Entscheidungsregel. So wenn etwa Harsanyi [1978] als Basis der utilitaristischen Wohlfahrtsfunktion nur die persönlichen und nicht die moralischen Präferenzen verwendet, oder ebenfalls Harsanyi [1977a] „assoziale“ Präferenzen ausschließt. In der Wohlfahrtsökonomie findet eine faktische Beschränkung der beachteten Präferenzen insofern statt, als meist nur konstante Präferenzen über selbst konsumierte Gütermengen betrachtet werden. Seltener werden auch in irgendeiner Form altruistische Präferenzen, interdependente und sich ändernde Präferenzen, Statuspräferenzen und relative Präferenzen untersucht. Moralische Präferenzen, „böartige“ Präferenzen und Präferenzen, die sich nicht als Präferenzen über bestimmte Güter rekonstruieren lassen (z.B. Präferenzen über gesellschaftliche Prozesse, die zu bestimmten Güterallokationen führen) werden nicht oder nur in Ausnahmefällen berücksichtigt.

Demgegenüber könnte das Paretokriterium gemäß dem vorgeschlagenen Maßstab nur dann als unproblematisch und allgemein zustimmungsfähig gelten, wenn es sich auf die Präferenzen aller betroffenen Individuen bezieht, *wie auch immer sie sind*.⁷⁸ Denn wenn ein Individuum weiß, daß in bestimmten Situationen seine Präferenzen nicht berücksichtigt werden oder irgendwelche andere Präferenzen für seine Präferenzen substituiert werden, dann könnte es nur noch unter engeren Voraussetzungen (bezüglich seiner Präferenzen oder bezüglich der Menge der möglichen Situationen) einer allgemeinen und konsequenten Verwirklichung des Paretokriteriums zustimmen.

Eine zweite Frage, die sich in der Interpretation des Paretokriteriums stellt, ist die danach, wessen Präferenzen berücksichtigt werden sollen. Auch die Antwort darauf ist mit dem vorgeschlagenen Maßstab für allgemeine Zustimmungsfähigkeit leicht zu geben. Es müssen einfach alle sein, die unproblematisch ihre Zustimmung zu diesem Kriterium geben sollen. Wenn also das Paretokriterium unproblematisch (ohne besondere Voraussetzungen über die Präferenzen oder möglichen Situationen) von der ganzen jetzt lebenden Menschheit akzeptiert werden können soll, dann sind auch die Präferenzen aller jetzt lebenden Menschen im Paretokriterium zu berücksichtigen. Wenn ein Entscheidungskriterium für alle jetzt und zukünftig lebenden Menschen im betrachteten Sinn unproblematisch sein soll, dann müssen auch die Präferenzen all dieser Menschen berücksichtigt werden.⁷⁹

⁷⁸ Mit dieser Aussage möchte ich *nicht* die Empfehlung verbinden, daß tatsächliche Entscheidungskriterien in der Weise paretoinklusive sein sollten, daß sie die hier diskutierte Fassung des Paretokriteriums einschließen. Es hier nur darauf hingewiesen werden, daß andere Fassungen des Paretokriteriums im hier betrachteten Sinn nicht allgemein zustimmungsfähig sind. Das ist insofern kein Problem, als über das Paretokriterium hinausgehende Entscheidungskriterien sowieso nicht allgemein zustimmungsfähig sind.

⁷⁹ Es wird oft gesagt, und ich selber gebrauche immer wieder diese Formulierung, bei bestimmten Entscheidungen seien alle und nur die Betroffenen zu berücksichtigen. *In der hier betrachteten Perspektive* kann das nur heißen, daß alle zu berücksichtigen sind, die nicht gegenüber allen möglichen Alternativen indifferent sind. Da aber diese Indifferenz erst festzustellen ist (in praktischen Entscheidungen setzt man oft eine solche Indifferenz ohne große Nachprüfungen einfach voraus), sind bei der Anwendung des Paretokriteriums alle Individuen zu berücksichtigen, die dem Paretokriterium zustimmen können sollen. Selbstverständlich können andere, voraussetzungsvollere und nicht notwendig allge-

Diese Interpretation des Pareto-kriteriums hat wichtige Konsequenzen: Es ist kaum mehr möglich, dieser Fassung des Pareto-kriteriums nicht zuzustimmen. Wenn das so verstandene Kriterium immer konsequent verwirklicht würde, wenn es anwendbar ist,⁸⁰ so würde aufgrund dieses Kriteriums nie ein Zustand verwirklicht, der nicht für alle Betroffenen – nach eigenen Maßstäben – der beste verwirklichtbare wäre. Jedes Individuum hätte, wenn es ganz allein hätte entscheiden können, in jedem Fall der Umsetzung des Kriteriums entweder genau den gleichen Zustand verwirklicht, den das Kriterium empfohlen hat, oder einen Zustand den es als genauso gut einschätzt (also nicht besser), wie den verwirklichten.⁸¹

Paradoxerweise gilt das selbst für ein Individuum, das das Pareto-kriterium als Entscheidungskriterium „prinzipiell“ ablehnt. Wenn es ein Individuum gäbe, das in jeder Situation einen Zustand, der nicht aufgrund des Pareto-kriteriums verwirklicht wird, einem Zustand, der aufgrund des Paretoprinzips verwirklicht wird, vorziehen würde (auch solche seltsamen Präferenzen müßten wir in der betrachteten Interpretation berücksichtigen), dann würde *aufgrund* des Pareto-kriteriums einfach nie eine gesellschaftliche Entscheidung durchgeführt.⁸²

Ähnlich läßt sich auch im Fall weniger radikaler Kritiker des Pareto-kriterium argumentieren: Wer zum Beispiel das Pareto-kriterium ablehnt, weil darin (in der hier diskutierten Fassung) z.B. auch unmoralische Präferenzen berücksichtigt werden, der würde, wenn er die dem entgegenstehenden moralischen Präferenzen wirklich hat, die Verwirklichung moralisch verwerflicher Zustände, Handlungen oder Maßnahmen aufgrund seiner – zu berücksichtigenden – moralischen Präferenzen verhindern.

Bis jetzt habe ich hauptsächlich die Behauptung diskutiert, Paretoprinzip bzw. Pareto-kriterium seien allgemein zustimmungsfähig. Aber läßt sich das hier vorgeschlagene Kriterium für allgemeine Zustimmungsfähigkeit eines Entscheidungskriteriums auch als Kriterium für den Anspruch der Wertfreiheit des Pareto-kriteriums nutzen?

Es gibt zumindest drei Argumente, die zeigen, daß selbst diese extrem geschwächte (weil anfordernde) Fassung des Pareto-kriteriums nicht wirklich wertfrei ist. Zum ersten ist dieses Kriterium unproblematisch, allgemeiner Zustimmung fähig für die, deren Präferenzen bei seiner Anwendung berücksichtigt werden. Wie aber wird diese Gruppe bestimmt? Wie soll man etwa zwischen den beiden als Beispiele genannten Möglichkeiten wählen? Soll man die jetzt lebenden Menschen oder auch alle zukünftigen Menschen berücksichtigen? Oder ist diese Gruppe zu groß, und es sollten immer nur die Menschen eines Landes berücksichtigt werden? Oder ist die Gruppe zu klein, und es wären auch die Interessen und Präferenzen von Tieren oder gar der ganzen belebten Natur zu berücksichtigen? Daß zumindest höhere Säugetiere eigene individuelle

mein zustimmungsfähige Kriterien anspruchsvoller in der Bestimmung der Gruppe der Betroffenen sein.

⁸⁰ Was unter anderem die Existenz kohärenter Präferenzen für jedes Individuum voraussetzt.

⁸¹ Man sollte beachten, daß das Pareto-kriterium in der hier betrachteten Fassung keine Aussagen über Situationen macht, in denen es nicht anwendbar ist. In solchen Situationen könnten natürlich noch immer Zustände verwirklicht werden (nicht aufgrund des Pareto-kriteriums), die von einigen oder vielen betroffenen Individuen als negativ eingestuft werden.

⁸² Man beachte aber das oben in Anm. 81 Gesagte.

Präferenzen besitzen, läßt sich kaum bezweifeln, und so ist es ethisch zu rechtfertigen – und damit sicher nicht wertfrei – warum sie evtl. nicht berücksichtigt werden sollten.

Zumindest in der Auswahl der Individuen, deren Präferenzen zu berücksichtigen sind, drückt sich also eine Wertauffassung aus, wenn wir das Paretokriterium als Entscheidungskriterium anwenden.

Ein zweites Argument für die Wertgeladenheit des Paretokriteriums ist in seinem ausgesprochen individualistischen Charakter zu sehen. Dieser individualistische Charakter mag ethisch rechtfertigbar sein. Er ist aber sicher nicht wertfrei, da er sich abhebt von ethischen Theorien, die Kriterien nicht individualistischen Charakters vorschlagen. Wer z.B. an die Möglichkeit von irgendwie kollektiven Entitäten (Familie, Gemeinschaft, Volk, Staat, Menschheit) glaubt und davon überzeugt ist, daß es ein Gemeinwohl bezogen auf diese kollektiven Entitäten gibt, das nicht einfach auf das Wohl der Individuen, die diese Entitäten ausmachen, zurückführbar ist, der wird zumindest *prima facie* die Möglichkeit nicht ausschließen können, daß auch eine Entscheidung, die die Zustimmung aller Individuen findet, die Teil einer kollektiven Identität sind, aufgrund des Kriteriums des Gemeinwohls anders ausfallen müßte. Auch alle in irgendeinem Grade deontologischen ethischen Theorien, die nicht das Rechte auf das aus der Sicht der Individuen Gute zurückführen, werden mit der Möglichkeit rechnen müssen, daß das, was von allen gewünscht wird, nicht das ist, was ethisch gefordert und damit zu verwirklichen ist. Das gilt vor allem, wenn sie in der Weise realistisch sind, daß sie nicht annehmen, daß das Rechte immer auch von den Individuen gewollt wird.

Ein drittes Argument für die Wertgeladenheit des Paretokriteriums ist darin zu suchen, daß es ja keineswegs ausgemacht ist, daß sich einem jedem Individuum eindeutig genau eine Präferenzordnung zuordnen läßt. Es ist in jedem Fall nicht von vornherein ausgeschlossen, daß zumindest einige Menschen für verschiedene Kontexte und Fragestellungen verschiedene Präferenzordnungen verwenden. So könnten etwa manche Individuum für Entscheidungen, die als privat eingestufte wirtschaftliche Fragen betreffen, persönliche Präferenzen im Sinne Harsanyis benutzen, und wenn sie wirtschaftspolitische Fragen, oder allgemeiner Fragen öffentlichen Charakters diskutieren und entscheiden, ethische Präferenzen im Sinne Harsanyis. Oder im Sinne der Theorien des „multiple self“ [Elster 1987] könnten manche Individuen zu verschiedenen Zeitpunkten verschiedene Präferenzen haben: Wenn sie in einer stillen Stunde über ihr Leben nachdenken, entscheiden sie sich etwa für eine Lebensführung, die irgendwelchen ethischen Vorstellungen entspricht, denen sie aber, wenn sie vor entsprechenden konkreten Entscheidungen stehen, nicht immer ganz entsprechen.⁸³

Damit muß auch hier wieder gewählt werden. Es muß entschieden werden, welche der verschiedenen Präferenzordnungen der Individuen dem Paretokriterium zugrunde liegen sollen. Je nach dem wird das Paretokriterium zu verschiedenen Ergebnissen führen können, weshalb auch diese Entscheidung ethisch zu rechtfertigen sein wird.

⁸³ Die allgemeine Zustimmungsfähigkeit des Paretokriteriums würde durch multiple Präferenzordnungen nicht gefährdet, solange (a) Individuen mit mehreren Präferenzordnungen trotzdem in jeder Entscheidungssituation eindeutig die jeweils in ihren Augen beste Entscheidung bestimmen könnten, und (b) in jeder Entscheidungssituation für die Anwendung des Paretokriteriums die unter (a) vorausgesetzte Menge der für jedes Individuum besten Entscheidungen zugrundegelegt würden.

2 Grenzen der Wohlfahrtsökonomie

Ökonomen gehen vielleicht allzu zuversichtlich davon aus, daß sich die verschiedenen Präferenzordnungen zu einer kohärenten effektiven und im Sinne der Theorie der offenbarten Präferenzen erhebbarer Präferenzordnung zusammensetzen lassen. Aber obwohl das ohne Zweifel möglich ist, ist es nicht notwendig so. Und auch wenn es so wäre, wäre der wertfreie Charakter des Paretokriteriums damit noch nicht sichergestellt. Denn man könnte ja aus ethischen Überlegungen heraus (wie das etwa Harsanyi tut) verlangen, daß nicht die resultierende Präferenzordnung, sondern eine der zugrundeliegenden ausgewählt wird. Dagegen wäre wieder die Entscheidung für die resultierende effektive Präferenzordnung (ethisch) zu rechtfertigen. Außerdem könnte, selbst wenn es eine solche kohärente resultierende Präferenzordnung gäbe, das betroffene Individuum für bestimmte kollektive Entscheidungen eine der Ausgangspräferenzordnungen anwenden wollen, so daß die Entscheidung die resultierende kohärente Präferenzordnung zu verwenden, sogar den unproblematischen Charakter des Paretokriteriums in Frage stellen könnte. Dieser setzt nämlich voraus, daß immer die Präferenzordnung zugrundegelegt wird, die das Individuum selbst zugrunde legen würde.

Wir sind hier also zwar bei einer wirklich „unproblematischen“ Fassung des Paretokriteriums angelangt, die abzulehnen, auch ein möglicherweise irgendwie betroffenes Individuum wenig Grund hat. Allerdings hat auch diese Fassung ihre Probleme: Neben den drei Fragen, die ich im Zusammenhang der Wertfreiheit des Kriteriums diskutiert haben, betrifft das vor allem seinen höchst anfordernden Charakter. Für diese allgemein zustimmungsfähige Fassung des Paretokriteriums ist besonders deutlich, wie wenig damit zu rechnen ist, daß je bedeutendere (z.B. wirtschaftspolitische) Entscheidungen auf seiner Basis getroffen werden könnten. Situationen, in denen z. B. die ganze jetzt lebende Menschheit in der Einschätzung der zu verwirklichenden Maßnahmen oder Zustände übereinstimmt, wird es nicht allzu häufig geben. Die Anforderung, daß alle, gegenwärtige wie zukünftige Menschen eine Maßnahme als Verbesserung bewerten, ist überhaupt uneinlösbar.

An vielen Stellen der Argumentation spielte es eine Rolle, daß ich den Bereich möglichen Präferenzstrukturen und Präferenzkonstellationen sehr weit gehalten habe. So etwa, wenn es galt zu zeigen, daß über das Paretokriterium hinausgehende Entscheidungskriterien nicht im selben Sinn unproblematisch genannte werden können. Ebenfalls davon betroffen sind die Frage der Möglichkeit verschiedener Präferenzordnungen eines Individuums, die evtl. gar nicht einmal in einer kohärenten effektiven Präferenzordnung resultieren, oder die hohen Anforderungen, die das Paretokriterium in der betrachteten Fassung stellt.

Hiergegen legt sich ein wichtiger Einwand nahe: Wir leben nur in *einer* Welt, in der nur *ein Teil* der möglichen Präferenzkonstellationen verwirklicht sind. Wäre es nicht möglich, daß auf der Basis dieser verwirklichten Präferenzkonstellationen etwa nicht doch noch weitergehende Entscheidungskriterien allgemein zustimmungsfähig wären? Oder daß Menschen eben doch nur eine kohärente Präferenzordnung besitzen, so daß die Möglichkeit mehrerer evtl. nicht kohärent verbindbarer Präferenzordnungen in unserer Welt keine Rolle spielt? Oder daß in unserer Welt in vielen Entscheidungssituationen doch alle Menschen in der Wahl des Besten übereinstimmen?

Dieser Einwand läßt sich noch etwas anders formulieren, wenn wir berücksichti-

gen, daß wir in unserer Reflexion sowieso immer auf Modelle, Repräsentationen von Wirklichkeit angewiesen sind, die immer auch von der wahrgenommenen Wirklichkeit abstrahieren. Es könnte deshalb sinnvoll sein, für die Wahl eines wirtschaftspolitischen Entscheidungskriteriums ein Modell von Wirklichkeit zugrunde zu legen, das weitergehende Annahmen über die Präferenzkonstellationen macht, auch wenn wir davon ausgehen, daß diese Annahmen nicht einfach so in der von unwahrgenommenen Wirklichkeit so erfüllt sind.

Das erste und zweite Argument für die Wertgeladenheit auch eines allgemein zustimmungsfähigen Paretokriteriums bleiben von diesem Einwand unberührt. Anders verhält es sich mit der Frage, ob nicht auch weitergehende Entscheidungskriterien allgemein zustimmungsfähig sein können. Unter weiteren Voraussetzungen über Präferenzkonstellationen und Entscheidungssituationen ist das ohne Zweifel möglich. Allerdings eben nur unter weiteren, einschränkenden Annahmen etwa über die möglichen Präferenzkonstellationen. Deshalb sind alle solche weitergehenden Entscheidungskriterien nicht im gleichen Grad unproblematisch wie das Paretokriterium, sie hängen von der Akzeptanz evtl. durchaus umstrittener Annahmen ab.

Die letzte Bemerkung gilt natürlich immer, wenn man auf der Basis von weiteren Annahmen über mögliche Präferenzkonstellationen argumentiert. Entscheidend ist dann, ob diese Annahmen in der vorliegenden Fragestellung adäquat sind. Das wird etwa im dritten Argument für die Wertgeladenheit des Paretokriteriums deutlich. Wenn wir in einer Welt leben würden, in der alle Menschen in der Tat nur eine einzige kohärente Präferenzordnung hätten, dann würde dieses Argument keine Rolle spielen können, da sich die Frage der Wahl unter verschiedenen Präferenzordnungen nicht stellen würde.

Ökonomische Modelle gehen nun in der Tat im allgemeinen davon aus, daß einem jeden betrachteten Individuum unzweideutig genau eine Präferenzordnung zugeordnet werden kann. Gehen wir einmal davon aus, daß das in typischen und traditionellen Fragestellungen der positiven Wirtschaftswissenschaften auch adäquat sei.⁸⁴ Aber kann man auch bei der Auswahl eines wirtschaftspolitischen Entscheidungskriteriums von dieser Annahme ausgehen, ist sie für wirtschaftspolitische Entscheidungen adäquat? Und was heißt in einem ethischen Diskurs adäquat? Welche Argumente im ethischen Diskurs zählen, welche nicht ethischen Tatsachenurteile und welche evtl. vereinfachende Annahmen bezüglich nicht direkt ethischer Tatsachenurteile vor dem Forum der praktischen Vernunft zugelassen sind, ist selbst eine Frage der praktischen Vernunft und damit Teil des ethischen Diskurses. Welche vereinfachenden Annahmen über Wirklichkeit in einem Diskurs zugelassen sind, entscheidet sich auf der Basis des Ziels diesen Diskurses, im Falle eines ethischen Diskurses also auf der Basis seines ethischen Ziels. Natürlich, wenn es eine unbezweifelte Tatsache wäre, daß Menschen nur eine Präferenzordnung haben, dann könnte und sollte man das einfach als eine nicht ethische Tatsache in den ethischen Diskurs übernehmen. Da die Sache aber umstrit-

⁸⁴ Typische und traditionelle Fragestellungen der Wirtschaftswissenschaften wäre Fragen, die sich auf das gesellschaftliche Subsystem Wirtschaft beziehen. Die Ausweitung ökonomischer Methoden auf z.B. Fragen des Heiratverhaltens oder des Verhaltens von Politikern zähle ich nicht zu den traditionellen Fragestellungen der Ökonomie.

ten ist, und es eine Reihe guter Argumente für die gegenteilige Auffassung gibt, gilt in jedem Fall bezogen auf die zweite Linie der Begründung des Einwands, daß es innerhalb des ethischen Diskurses zu entscheiden ist, ob es eine sinnvolle Vereinfachung ist anzunehmen, daß jeder Mensch nur eine Präferenzordnung hat oder nicht. Die Entscheidung darüber ist also selbst wertgeladen.⁸⁵

Ich habe bis jetzt eine allgemein zustimmungsfähige Fassung des Paretokriteriums diskutiert. Ist nach dem hier angewandten Kriterium auch das Paretoprinzip (wie von mir definiert) allgemein zustimmungsfähig? Das heißt, könnte jedes Individuum unproblematisch einer Entscheidung zustimmen, die unwiderruflich und effektiv festlegt, daß in Zukunft nur noch Entscheidungskriterien angewandt werden, die paretoinklusiv sind? Das ist meinem Eindruck nach nicht der Fall. Dabei ist ein wichtiger Unterschied zwischen dem Paretokriterium und dem Paretoprinzip zu beachten: Das Paretokriterium in der hier diskutierten Fassung würde, wenn es effektiv und unwiderruflich implementiert würde, dazu führen, daß immer dann, wenn alle berücksichtigten Individuen in der Einschätzung der besten Entscheidung übereinstimmen, diese Entscheidung auch durchgeführt wird. Das unterwirft etwaige weitere Entscheidungen und Entscheidungskriterien keinerlei Restriktionen. Das hat zu der paradoxen Situation geführt, daß selbst ein (konsequenter) Gegner des Paretoprinzips der effektiven und unwiderruflichen Approbation des Paretokriteriums zustimmen könnte, da seine (konsequente) Gegnerschaft ja verhindern würde, daß es je zur Anwendung käme.⁸⁶ Das ist im Fall des Paretoprinzips anders. Das effektiv und unwiderruflich beschlossene Paretoprinzip würde allen zukünftigen Entscheidungen und Entscheidungskriterien Restriktionen auferlegen. Wer also irgendein Kriterium kollektiver Entscheidungen vertritt, das nicht paretoinklusiv ist (z.B. ein strikt egalitäres, nicht-individualistisches oder perfektionistisches Entscheidungsprinzip), der würde mit seiner Zustimmung zum Paretoprinzip die Möglichkeit der Implementierung seiner Vorstellungen zu kollektiven Entscheidungen verhindern, und deshalb eine solche Zustimmung nicht geben.

Dieser fundamentale Unterschied ist bemerkenswert, besonders weil zwischen Paretoprinzip und Paretokriterium oft nicht klar unterschieden wird und damit der Eindruck erweckt wird, als wären beide gleichermaßen problemlos und selbstverständlich.⁸⁷

⁸⁵ Ich sehe zumindest drei Argumente, die dafür sprechen, daß Menschen im allgemeinen mehr als eine Präferenzordnung zugrunde legen. (1) Es scheint der internen lebensweltlichen Erfahrung von Menschen eher zu entsprechen, die sich ethischen Zielen und Grundsätzen verpflichtet wissen, oder die sich für einen längerfristigen Lebensplan entschieden haben, ohne im konkreten Handeln dem immer zu entsprechen. (2) In nicht-konsequentialistischen und selbst in nicht egoistisch-konsequentialistischen Ethikentwürfen ist das Rechte oder ethisch Gute dem (subjektiv) Guten vorgeordnet und nicht auf das (subjektiv) Gute reduzierbar. Wenn diese Entwürfe in der Ablehnung des egoistischen Konsequentialismus (als ethischer Theorie) Recht haben, dann wären Menschen ethisch *verpflichtet*, neben einer eventuellen privaten, eigenorientierten Präferenzordnung zumindest einer zweiten Präferenzordnung zu folgen, die nicht einfach auf die erstere reduzierbar wäre. (Die beiden Präferenzordnungen wären evtl. nicht einmal Ordnungen über dieselben Mengen.) (3) Es gibt einen wachsenden Korpus empirischer Studien, die darauf hinweisen, daß neben privaten Präferenzen bei vielen Entscheidungen moralische und Gerechtigkeitsüberlegungen eine Rolle spielen.

⁸⁶ Das ist paradox, weil es ja aufgrund des Paretokriteriums nicht zu Anwendung käme.

⁸⁷ Hier ist allerdings zu beachten, daß im ökonomischen Schrifttum Paretokriterium und Paretoprinzip

Als Fazit läßt sich zunächst festhalten, daß es zwar eine Version des Pareto­kriteriums gibt, die als allgemein zustimmungsfähig und in diesem Sinn als unproblematisch gelten kann, daß diese Version jedoch so anfordernd ist, daß eine praktische Verwendbarkeit für wirtschaftspolitische Fragen ausgeschlossen werden kann. Das Pareto­kriterium in dieser Version erweist sich als nutzlos.

Andere Versionen des Pareto­kriteriums wie auch das Paretoprinzip dagegen können nicht als allgemein zustimmungsfähig gelten. Es gibt einen ganzen Fächer von traditionellen ethischen Positionen, die etwa aus der bloßen Tatsache, daß direkt Betroffene subjektiv und auf der Basis ihrer faktischen Präferenzen eine bestimmte Veränderung als positiv bewerten, nicht folgern, daß diese Veränderung ethisch als positiv zu bewerten ist.

Dies gilt insbesondere für alle – autonome und heteronome – deontologischen Positionen, die das Rechte dem Guten vorordnen und nicht auf das subjektiv Gute oder empirische Präferenzen zurückführen.⁸⁸ Es gilt auch für konsequenzialistische Ethikentwürfe, deren Konzeption des Guten nicht einfach auf individuelle Präferenzen zurückzuführen ist, wenn sie etwa – wie Harsanyi – die empirischen Präferenzen durch sozial verträgliche Präferenzen ersetzen. Und es gilt für perfektionistische Ethikentwürfe, deren Konzeption des Guten nicht aus empirischen, individuellen Präferenzen, sondern aus einem Ideal menschlicher Leistung oder eines guten Lebens abgeleitet sind.⁸⁹

Wenn auch das zentrale Thema dieses Kapitels die These ist, daß Ökonomen die explizite Bezugnahme auf substantielle ethische Positionen nicht vermeiden können, wenn sie substantielle normative Aussagen machen wollen, da Paretoprinzip wie Pareto­kriterium dazu nicht ausreichen, hat der letzte Absatz noch darüberhinaus deutlich machen wollen, daß in seinen üblichen Fassungen Paretoprinzip wie Pareto­kriterium keineswegs ethisches Gemeingut sind. Wer der Paretoprinzip benutzt, bezieht damit eine *umstrittene* ethische Position, und befindet sich schon damit mitten in der ethischen Debatte.

bezogen auf Präferenzen über selbst konsumierte Güterbündel definiert werden. In dieser Fassung ist selbst das Pareto­kriterium im hier betrachteten Sinn dann *nicht* problemlos, wenn Individuen z.B. in ihren Präferenzen Gerechtigkeitskriterien berücksichtigen oder allgemeiner in irgendeiner Weise Präferenzen haben, die gesellschaftliche Zustände nicht einfach nach den selbst konsumierten Güterbündeln ordnen. Daß die von Ökonomen normalerweise betrachtete Fassung des Paretoprinzips nicht allgemein zustimmungsfähig ist, soll nicht heißen, daß diese Fassung falsch ist. Sie mag die für ökonomische Fragestellungen eigentlich angemessene Fassung sein, sie ist aber eben nicht „harmlos“.

⁸⁸ Das betrifft u. a. die kantische Ethik, die Rawlsche Theorie der Gerechtigkeit und verschiedenste Entwürfe einer christlichen Ethik.

⁸⁹ Das betrifft u. a. die aristotelische Tradition, aber etwa auch Nietzsche. Vgl. Rawls [1975, 43, 360-367].

2 Grenzen der Wohlfahrtsökonomie

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

3.1 Paretoeffizienz und Gerechtigkeit: Zur Exposition der Fragestellung.

Es ist ein weitverbreiteter Glaubenssatz unter Ökonomen, daß es einen Trade-off zwischen Effizienz und Gerechtigkeit gebe. Es gilt einfach als selbstverständlich, daß mehr Effizienz in der Regel weniger Gerechtigkeit bedeutet, und mehr Gerechtigkeit nur auf Kosten eines gewissen Maßes an Effizienz zu haben sei.

Natürlich ist ein solcher Glaube nur dann begründbar, wenn man eine halbwegs klare Vorstellung davon hat, was man mit Effizienz und Gerechtigkeit eigentlich meint.

Was Effizienz angeht, so scheint es zunächst klar zu sein, was darunter zu verstehen ist. Man kann vermuten, daß hier Paretoeffizienz gemeint ist. Was Gerechtigkeit angeht, so liegen die Dinge komplizierter. Will es doch ein weiterer ökonomischer Glaubenssatz, daß es praktisch unmöglich ist, über Fragen der Gerechtigkeit Einigkeit zu erzielen. Schon das sollte zur Vorsicht mahnen, denn wenn es viele Vorstellungen von Gerechtigkeit gibt, dann ist es *prima facie* nicht auszuschließen, daß manche davon in keiner Weise Einbußen an Effizienz verlangen.

Es ist nun in der Tat so, daß gängige Vorstellungen von Gerechtigkeit in keinerlei Gegensatz zur Forderung der Verwirklichung von Paretoeffizienz stehen, einfach da sie selbst paretoinklusiv sind.

Das gilt z. B. für den Utilitarismus, für den Nutzenrawlsianismus¹, oder auch für die Gerechtigkeitsvorstellung, die in der auf Nutzen angewandten Leximin-Regel ausgedrückt wird; allgemeiner auch für alle Gerechtigkeitstheorien, die sich in einer paretoinklusiven Wohlfahrtsfunktion formulieren lassen. Paretoinklusive Wohlfahrtsfunktionen vervollständigen auf ergänzende Weise die Pareto-Relation, und helfen deshalb

¹ Damit bezeichne ich Positionen, die das *Maximinkriterium* im Nutzenraum anwenden und es als alleinstehendes Kriterium der Gerechtigkeit benutzen.

Man beachte dabei noch Folgendes: Das Maximinkriterium implementiert das schwache Paretoprinzip, nicht das starke. Trotzdem kann in folgendem Sinn kein Trade-off zwischen dem Paretokriterium und dem Maximinkriterium im Nutzenraum auftreten: Nach dem Maximinkriterium sind gewisse gesellschaftliche Zustände gleich gut, die nach dem Paretokriterium noch einmal strikt geordnet sind. Das Maximinkriterium verpflichtet aber natürlich nicht zur Wahl eines bestimmten indifferenten gesellschaftlichen Zustands (d. h. es ist möglich unter den nach dem Maximinkriterium indifferenten Zuständen einen paretosuperioren Zustand auszuwählen). Deshalb verliert man weder in der Pareto-Ordnung etwas durch die Anwendung des Maximinkriteriums, noch in der Ordnung des Maximinkriteriums durch die Anwendung des Paretokriteriums.

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

unter paretoeffizienten gesellschaftlichen Zuständen auszuwählen, aber eben unter paretoeffizienten.

Solange wir also davon ausgehen, daß Effizienzgesichtspunkte vollständig durch das Paretoprinzip erfaßt sind, kann ein Trade-off zwischen Effizienz und Gerechtigkeit nur entstehen, wenn Gerechtigkeit durch eine Ordnung erfaßt wird, die nicht paretoinklusiv ist.

Beispiele dafür wären Gerechtigkeitsvorstellungen, die Gerechtigkeit als Gleichheit im Güterraum oder im Nutzensraum definieren. Es gibt dann Mengen E von erreichbaren gesellschaftlichen Zuständen, so daß man zwischen Gleichheit im Güterraum bzw. Nutzenraum und Paretoeffizienz wählen muß. Im Falle der Gerechtigkeitsforderung nach Gleichheit im Nutzensraum müßten dazu E und die betrachteten individuellen Ordnungen über E so beschaffen sein, daß es zu einem „Rückwärtsbinden“ der Nutzenmöglichkeitsgrenze im Bereich der Gleichverteilung kommt. Nur als Nebenbemerkung sei erwähnt, daß das in ökonomischen Modellen praktisch nie vorkommt: In den verbreiteten ökonomischen Modellen würde es also nicht einmal dann zu einem Trade-off zwischen Effizienz und Gerechtigkeit kommen, wenn Gleichheit der Nutzen als Gerechtigkeit verstanden würde.

Die Überlegungen bis hierher waren rein formal: Wie muß eine Gerechtigkeitsvorstellung aussehen, daß es überhaupt zu einem Trade-off zwischen Effizienz und Gerechtigkeit kommen kann.

Solche formalen Überlegungen sind aber eigentlich nicht ausreichend. Das Sprechen von einem Trade-off suggeriert die Idee von Kosten, von Verlust, und setzen damit voraus, daß man sich zwei Werten verpflichtet weiß, die miteinander in Konflikt geraten können. Das Paretoprinzip wird aber von denen, die es vertreten, oft als eine Mindestanforderung an akzeptable vollständigere Ordnungen gesellschaftlicher Zustände verstanden. Das macht es eher unwahrscheinlich, daß Anhänger des Paretoprinzips Gerechtigkeitsvorstellungen vertreten, die nicht paretoinklusiv sind. Wer das Paretoprinzip vertritt, wird kaum Gleichheit im Güterraum für eine gute Theorie der Gerechtigkeit halten, in jedem Fall ist es alles andere als selbstverständlich, daß Gleichheit im Güterraum eine gute Theorie der Gerechtigkeit *ist*.

Gerechtigkeitsvorstellungen dagegen, die auf die Implementierung des Paretoprinzips im Nutzenraum verzichten, tun dies in der Regel, weil sie diese Implementierung für irrelevant halten.

Als ein erstes Ergebnis können wir also festhalten, daß ein Trade-off zwischen Paretoeffizienz und Gerechtigkeitsvorstellungen keinesfalls als ein allgemein gegenwärtiges Phänomen einfach vorausgesetzt werden kann.

Der Verdacht liegt also nahe, daß Ökonomen, wenn sie einen Trade-off zwischen Effizienz und Gerechtigkeit für so selbstverständlich halten, in Wirklichkeit einen anderen Begriff von Effizienz benutzen als den der Paretoeffizienz. Bevor ich solche alternativen Begriffe darstelle, möchte ich zunächst noch ganz abstrakt fragen, unter welchen Bedingungen und welchem Sinn man allgemein von einem Trade-off zwischen Effizienz und Gerechtigkeit sprechen könnte.

3.2 Was ist ein Trade-off?

Nehmen wir an, wir besäßen irgendein Kriterium, mit dessen Hilfe wir alle möglichen gesellschaftlichen Zustände nach Effizienz Gesichtspunkten ordnen könnten – genauer wir besäßen eine vollständige Ordnung über alle möglichen gesellschaftlichen Zustände auf der Basis der Effizienz. Für Paretianer würde eine solche Ordnung paretoinklusiv sein, aber man mag auch andere Vorstellungen darüber haben, was Effizienz bedeutet. Daneben besäßen wir auch eine vollständige Ordnung aller gesellschaftlichen Zustände nach dem Kriterium der Gerechtigkeit. Wiederum könnte diese Ordnung paretoinklusiv sein oder nicht. Wenn wir jetzt beiden Gesichtspunkten, dem der Effizienz und dem der Gerechtigkeit verpflichtet sind, und die beiden Ordnungen nicht übereinstimmen, dann brauchen wir, um bei jeder möglichen Menge erreichbarer gesellschaftlicher Zustände zu Entscheidungen zu kommen, ein Kriterium für Situationen, in denen die beiden Ordnungen zu verschiedenen Entscheidungen führen. Hier kann es nun sinnvoll sein, von einem Trade-off zwischen Effizienz und Gerechtigkeit zu sprechen: Wäre ich nur einem der beiden Kriterien verpflichtet, könnte ich einen diesem Kriterium gemäß besseren Zustand erreichen, die Berücksichtigung des jeweils anderen Kriteriums verlangt von mir, etwas an Effizienz bzw. Gerechtigkeit zu opfern. Zwei Punkte sind hierbei wichtig: Ein Trade-off kann nur auftreten, wenn die betrachteten Kriterien zu verschiedenen Ordnungen über bestimmte Mengen gesellschaftlicher Zustände führen. Zum anderen ist es aber wohl auch nur dann sinnvoll von einem Trade-off zu sprechen, wenn man beiden Kriterien verpflichtet ist: Wer Gerechtigkeit für völlig unwichtig hält, für den verliert eine Gesellschaft nichts, wenn um den nach Effizienz Gesichtspunkten besten Zustand zu erreichen, ein nach dem Gerechtigkeitskriterium nicht optimaler Zustand gewählt wird. Entsprechendes gilt auch umgekehrt.

Im vorausgehenden Paragraph war nur von vollständigen Ordnungen die Rede. Selbstverständlich kann aber auch ein Trade-off zwischen Teilordnungen auftreten. Allgemein kann ein Trade-off zwischen zwei (Teil-) Ordnungsrelationen nur dann auftreten, wenn mindestens zwei gesellschaftliche Zustände durch diese (Teil-) Ordnungen verschieden geordnet werden, und es ein übergeordnetes Kriterium gibt, das die beiden (Teil-) Ordnungen so zu einer neuen, dritten (Teil-) Ordnung verknüpft, daß diese mit keiner der beiden ursprünglichen (Teil-) Ordnungen für alle möglichen gesellschaftlichen Zustände übereinstimmt.²

Unter den gegebenen Bedingungen *kann* ein Trade-off zwischen den ursprünglichen Ordnungen auftreten, das muß aber nicht der Fall sein. Insbesondere kann es sein, daß die neue konstruierte Ordnung zwischen Zuständen differenziert, die in einer der beiden ursprünglichen Ordnungen als indifferent, gleich gut geordnet waren. In diesem Fall ist es aber kaum angemessen von einem Trade-off zu sprechen: Wenn z. B. die neue Ordnung einige gesellschaftliche Zustände aus einer Menge von gemäß einer der ursprünglichen Ordnungen indifferenten Zustände als die besten herausfiltert, so verliert

² Wäre die so konstruierte dritte (Teil-) Ordnung mit einer der beiden ursprünglichen einfach identisch, so wäre das Sprechen von einem Trade-off zwischen den beiden ursprünglichen Ordnungen reine Rhetorik. Im folgenden werde ich der Einfachheit halber zwischen Teilordnungen und Ordnungen sprachlich i.d.R. nicht mehr unterscheiden

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

man ja gemäß dieser ursprünglichen Ordnung nichts. Es war gemäß dieser ursprünglichen Ordnung ja gerade egal, welchen der indifferenten Zustände man verwirklicht.³

Wenn wir jetzt noch das Verhältnis von Effizienzordnung und Gerechtigkeitsordnung zum Paretoprinzip betrachten, so sei wiederholt, daß zwischen dem Paretokriterium und einer paretoinklusiven Gerechtigkeitsordnung kein Trade-off bestehen kann. Nur also, wenn man der Meinung ist, daß durch die Pareto-Relation nicht genügend erfaßt wird, was mit Effizienz gemeint ist, und versucht, diese Teilordnung nach irgendeinem überzeugend erscheinenden Gesichtspunkt zu ergänzen, kann es einen Gegensatz zwischen einer so erweiterten Effizienzordnung und einer paretoinklusiven Gerechtigkeitsordnung auftreten. Alternativ könnte man sogar eine Effizienzordnung benutzen, die selbst nicht paretoinklusiv ist. Auch dann kann gegenüber paretoinklusiven Gerechtigkeitsordnungen ein Trade-off zwischen Effizienz und Gerechtigkeit auftreten.

3.3 Realeinkommen als Maßstab für Effizienz.

In seinem einflußreichen Buch „Equality and Efficiency: The Big Tradeoff“, das nicht wenig dazu beigetragen hat, den Glauben an diesen Trade-off zu popularisieren, gibt sich A. Okun wenig Mühe, Effizienz zu definieren: Er gibt gleich am Anfang ein Beispiel für *technische* Effizienz und erklärt dann, daß dieser Begriff der Effizienz impliziere, daß mehr besser ist, „insofar as the ‘more’ consists of items that people want to buy.“

Ich bin davon überzeugt, daß diese Intuition, daß mehr besser ist, (und *nicht* das Paretoprinzip) die Grundlage eines großen Teils der Attraktivität von Effizienzüberlegungen für Ökonomen ist. Und sobald man eine nachvollziehbare Meßvorschrift (z. B. Bruttosozialprodukt) dafür angibt, wie man das Mehr an Dingen, die von Leuten gern gekauft werden, mißt, können auf der Basis *dieser* Effizienzvorstellung in der Tat Politikempfehlungen gegeben werden.

Würden wir in einer Welt von einem Konsumenten und einem Gut leben, wäre es leicht eine solche Meßvorschrift anzugeben: die physische Menge dieses einen Gutes. In unserer Welt vieler Güter und vieler Konsumenten dagegen, ist das schwieriger: Wollte man die Methode, einfach die physischen Mengen zu messen und zu vergleichen, auch hier anwenden, käme man in dem Moment in Schwierigkeiten, wo man Situationen miteinander vergleicht, für die nicht gilt, daß in der einen Situation von allen Gütern mehr vorhanden sind, als in der anderen, sondern von manchen mehr und von anderen weniger.

Wir könnten – konsequenterweise – auf „Effizienzvergleiche“ solcher Situationeneinfach verzichten. Weil in einer Ökonomie, die noch keine Verbrennungsmotoren kennt, mehr Hufeisen hergestellt werden als in einer, die solche Motoren benutzt, werden wir also keine Vergleiche des Realeinkommens solcher Ökonomien versuchen und z. B. nicht behaupten, heute würden in Deutschland mehr Güter hergestellt als vor 200 Jahren.

Diese Vorstellung wirkt absurd: Es scheint selbstverständlich, daß letzterer Vergleich möglich ist. Es stellt sich deshalb die Frage nach einem alternativen Maßstab, der einen

³ Vgl. Anm. 1 oben.

solchen Vergleich möglich macht.

Würden wir uns in einer Vielgüterwelt mit nur einem Konsumenten bewegen, wäre aber auch dieses Problem leicht lösbar, jedenfalls solange wir uns in der neoklassischen Modellwelt bewegen. Dieser eine Konsument hätte eine vollständige Präferenzordnung über alle möglichen Güterbündel, und wir könnten einfach diese Präferenzordnung als Maßstab benutzen: indem wir den Konsumenten einfach befragen, oder ihm trickreich seine Präferenzordnung durch empirische Untersuchungen (z. B. seines Nachfrageverhaltens) entlocken.

Letztlich ist das auch der einzige Maßstab, der dem Anspruch, das Urteil des Konsumenten selbst (allein) entscheidend dafür zu machen, ob die produzierten Güter solche sind, die er kaufen will, gerecht wird.

Aber auch dieser Maßstab versagt, wenn wir ihn auf die Welt, in der wir leben anwenden wollen: Da Menschen unterschiedliche Präferenzordnungen haben, gäbe es nicht einen, sondern viele verschiedene und zu unterschiedlichen Ergebnissen führende Maßstäbe dafür, ob in einer Situation mehr Güter produziert und angeboten werden als in einer anderen.

Das Problem, einen Maßstab für die Menge von Gütern zu finden in einer Welt von vielen Gütern und vielen verschiedenen Individuen, ist das Problem der Messung des Realeinkommens. Das ist also der Maßstab für Effizienz, den wir in diesem Unterabschnitt betrachten.

Bleiben wir noch einen Moment in der Einkonsumentenwelt. Unter einigen weiteren großzügigen neoklassischen Annahmen können wir sogar auf die Kenntnis der Präferenzen dieses Konsumenten verzichten und trotzdem zu einer vollständigen Ordnung z. B. aller möglichen Güterbündel kommen, von der sinnvollerweise gesagt werden kann, daß sie das Realeinkommen beschreibt. Vier Dinge werden dafür über die üblichen Annahmen zu Präferenzordnungen über Güterbündel hinaus vorausgesetzt: (1) Wir kennen für jedes mögliche Güterbündel eine Transformationskurve, die beschreibt, welche Güterbündel für das Individuum vom gegebenen Güterbündel aus erreichbar sind. (2) Wir können voraussetzen, daß das Individuum sicher das für es optimale Bündel aus den gegebenen Transformationskurven wählen kann und wird. (3) Für jedes Güterbündel x und jede Transformationskurve gilt, daß x entweder auf der Transformationskurve liegt, oder ein y auf der Transformationskurve existiert, so daß $x > y$ oder $y > x$. (4) Transformationskurven schneiden sich nicht. Unter diesen Voraussetzungen können wir eine vollständige Ordnung \succ definieren, die wegen Voraussetzung (2) sinnvollerweise als eine Ordnung des Realeinkommens gewertet werden kann:

$$x \succ y \Leftrightarrow \exists x' \text{ auf der Transformationskurve von } x, \\ y' \text{ auf der Transformationskurve von } y : x' > y' \quad (3.1)$$

Wenn im Fall eines Einkonsumentenmodells überhaupt sinnvoll über den Grad an Realismus verschiedener Annahmen gesprochen werden kann, dann ist die Kombination der Annahmen (3) und (4) sicher dann völlig unrealistisch, wenn die Transformationskurven technisch interpretiert werden: Produktionsprozesse sind in der Regel

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

irreversibel und es läßt sich nicht alles aus allem produzieren – selbst in kleinsten Mengen.

Mehr Realismus können diese beiden Annahmen dann in Anspruch nehmen, wenn wir ein Modell eines einzelnen Konsumenten angesichts eines ideal funktionierenden kompetitiven Marktes betrachten. Dann können die Transformationskurven als Budgetgrenzen bei gegebenen Preisen verstanden werden, die bei festen Preisen parallel verlaufen werden.

Das erklärt, warum auch bei Unkenntnis der Präferenzen eines einzelnen Konsumenten oder Haushaltes bei gegebenen oder annähernd festen Preisen Realeinkommen eines Einzelnen ein unproblematischer Begriff ist. Vielleicht kommt aus der lebensweltlichen Erfahrung der unproblematischen Verwendung des Einkommensbegriffs bezogen auf einzelne Individuen und Haushalte, die intuitive Selbstverständlichkeit, mit der wir diesen Begriff auf andere Situationen übertragen wollen. Aber für die ökonomische Analyse fallen Preise nicht vom Himmel. In einem Totalmodell werden sie (neoklassisch) u.a. auf der Basis der Kenntnis der Präferenzen bestimmt.

Auch ohne einzelne Vorschläge zur Messung des Realeinkommens zu betrachten, können wir schon aufgrund prinzipieller Überlegungen einige Aussagen über grundlegende Eigenschaften möglicher Maßstäbe für das Realeinkommen machen.

Das Ideal der Messung des Realeinkommens wäre es, einer jeden Allokation von Gütern (evtl. unter Einschluß der nicht über Märkte gehandelten)⁴ einen Index (das Realeinkommen) zuzuordnen zu können, von dem sinnvoll gesagt werden könnte, er messe die Menge an vorhandenen (und von Endnachfragern gewünschten) Gütern. Die dafür vorgeschlagenen Verfahren sehen sehr verschieden aus und verlangen evtl. keine Informationen über die von einzelnen Individuen nachgefragten Mengen, wie wenn z. B. aggregierte Gütermengen mit ihrem Marktpreis gewichtet und aufaddiert werden: Implizit sollte aber auch bei einem solchen Verfahren einer jeden Allokation ein Index zugeordnet werden.

Damit ist nun aber schon klar, um was es sich bei einer jeden Meßvorschrift für die Messung des Realeinkommens handelt: um eine Ordnung aller möglichen gesellschaftlichen Zustände. Damit gilt sachlich für die Messung des Realeinkommens alles, was wir heute über Wohlfahrtsfunktionen wissen, insbesondere gelten für die Messung des Realeinkommens alle die Möglichkeit- und Unmöglichkeitstheoreme, die in der Folge Arrows hergeleitet worden sind.⁵ „Effizienz“, auch definiert auf der Basis eines Maßstabs für das „Realeinkommen“, ist eben keine Zauberformel dafür, ohne normative Wertmaßstäbe, die über das Paretoprinzip hinausgehen, oder es gar verletzen, normative Urteile (und seien es auch „nur“ Effizienzurteile) herzuleiten.

Jede Meßvorschrift für das Realeinkommen *muß* also (zumindest) eine der Forderungen des Arrowschen Unmöglichkeitstheorems verletzen.

Eine Reihe von Realeinkommensmeßvorschriften verletzt (solange nicht besondere Annahmen über die Präferenzen gemacht werden, womit gegen die Forderung eines

⁴ Die folgenden Überlegungen würden auch dann ihre Gültigkeit behalten, wenn ich die „ökonomistische Engführung“ auf Allokationen vermeiden würde, und allgemeiner von gesellschaftlichen Situationen sprechen würde.

⁵ Eine Zusammenstellung einiger dieser Ergebnisse findet sich etwa in Sen [1986].

3.3 Realeinkommen als Maßstab für Effizienz.

universellen Definitionsbereichs verstoßen wird⁶) das Paretoprinzip. Das gilt insbesondere auch für alle sogenannten „statistischen“ Indexmaße, die aufsummierte Einkommen mit irgendeinem (beliebigen!) Preisindex deflationieren⁷, oder mit den Marktpreisen einer Basissituation (bzw. deren Ersatz) gewichtete Gütermengen aufaddieren.⁸ Solche Maßstäbe für das Realeinkommen und die Effizienz mögen sinnvoll sein, aber sie bedürfen einer eigenständigen normativen Rechtfertigung. Aus der Eingängigkeit des Paretoprinzips können sie sicher nicht abgeleitet werden.

An zwei Beispielen möchte ich diese abstrakten Überlegungen illustrieren.

In der Literatur, die sich mit sogenannten „sozialen Evaluierungsfunktionen“ beschäftigt, wird eine Vorgehensweise gewählt, die es möglich macht, an der Idee eines Trade-offs zwischen Effizienz und Gerechtigkeit sinnvoll festzuhalten.⁹ Es werden voneinander unabhängige Maßstäbe für Gerechtigkeit und Effizienz von bestimmten Einkommenssituationen entwickelt, die dann durch eine Entscheidungsvorschrift bei Konkurrenz der beiden Maßstäbe ergänzt werden, um so zu *einem* Wohlfahrtsmaß zu gelangen.

Weinhardt [1989] zeigt z. B. wie einige wenige Axiome (Additivität¹⁰, Symmetrie, Irrelevanz der Geldeinheit und Unabhängigkeit von der Bevölkerungsgröße) die Zahl der möglichen Effizienzmaßstäbe auf Maßstäbe der Form

⁶ Z. B. ermöglicht nur die Beschränkung auf identische und homothetische Präferenzen für alle Individuen, oder allgemeiner auf Präferenzen mit parallelen und geraden Engelkurven eine konsistente Definition von Realeinkommen über Nutzengrenzen, die sich dazu nicht schneiden dürfen. [Vgl. Gorman 1953, Gorman 1961 u. Blackorby u. Donaldson 1985]

⁷ Dazu gleich im folgenden zu Weinhardt [1989].

⁸ Beide Indexmaße verstoßen schon für ein einzelnes Individuum gegen das Paretoprinzip, da sie die individuellen Präferenzen nicht korrekt wiedergeben. [Vgl. Ahlheim u. Rose 1992, 145-156, 205-207]

Die zweite Art von Meßvorschriften (Mengenindizes) spielt insbesondere in der sogenannten Kosten-Nutzen Literatur eine Rolle. Dort werden allerdings oft *keine* für alle Vergleiche identische Basispreise anzugeben, statt dessen werden z. B. sei es die Preise der *jeweiligen* Ausgangssituation oder der *jeweiligen* Alternativsituationen herangezogen. Das führt zu wechselnden Maßstäben, so daß von *einer* Wohlfahrtsfunktion nicht mehr die Rede sein kann. Das macht die Situation dieser Maßstäbe aber nicht besser: Zu den im Haupttext erwähnten Problemen kommen nur Probleme z. B. der Zirkularität *hinzu*.

⁹ Allerdings sollte man dabei nicht so tun, als sei dieser Trade-off etwas, was man *entdecken* könne, das also dem Erkennen und Werten der Ökonomen irgendwie vorausliege. Es gibt einen Trade-off zwischen Effizienz und Gerechtigkeit, weil es Anhängern dieses Ansatzes sinnvoll erscheint, die Idee der Gerechtigkeit und die der Effizienz zunächst getrennt darzustellen, und dann in einem Wohlfahrtsmaß zu vereinigen: weil man also an der Idee eines solchen Trade-offs aus irgendwelchen Gründen festhalten *will*.

¹⁰ Aus Weinhardt [1989] geht nicht hervor, wie die Additivität als sinnvolle Forderung an ein Effizienzmaßes begründet werden kann. Der (motivierende?) Hinweis S. 5 auf verschiedene Einkommenskomponenten (wie Arbeits- oder Kapitaleinkommen) ist natürlich unsinnig, da die verschiedenen Komponenten ja sowieso ununterscheidbar zu einem Einkommen aufaddiert werden, und die Additivität für ganz *beliebige* Einteilungen dieses *einen* Einkommens gelten soll. Die *Additivität verschiedener Einkommensarten* wird also durch die Vorschrift, alle Einkommensarten zu einem Einkommen aufzuaddieren, sichergestellt. Das hat mit der Additivität des *Effizienzmaßes* nichts zu tun.

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

$$E(p, x) := \beta^m(p) \frac{1}{n} \sum_{i=1}^n x_i \quad (3.2)$$

reduzieren. Dabei sei $\beta^m(p)$ ein vom Grade -1 homogener Preisdeflator, m die Zahl der verschiedenen Güter, p der Preisvektor, n die Zahl der Individuen und x_i das Einkommen des Individuums i . Maßstab für Effizienz ist also das deflationierte durchschnittliche Nominaleinkommen (Realeinkommen).

Dieses Effizienzmaß läßt sich als eine nicht paretianische Wohlfahrtsfunktion verstehen, die alle Preis-Einkommens-Situationen und damit implizit alle Allokationen ordnet. Nur wenn wir uns auf Individuen mit identischen und homothetischen Präferenzen beschränken, können wir dieses Maß als paretianisch verstehen, denn nur dann existieren Preisdeflatoren $\beta^m(p)$, so daß für alle Preis-Einkommens-Situationen (p, x) (mit x als Vektor der n Einkommen) gilt: ¹¹

$$(p, x) \succeq_p (p', x') \rightarrow E(p, x) \geq E(p', x') \quad (3.3)$$

Damit ist auch klar, in welchem Sinne dieses Effizienzmaß auf das Abwägen verschiedener Nutzen, auf Nutzenvergleiche verzichten kann. Etwas salopp gesagt: weil sich dieses Maß um die Präferenzen der Individuen nicht kümmert. Genauer: Es ist im allgemeinen unmöglich, $E(p, x)$ als eine Funktion allein der indirekten Nutzen der Individuen darzustellen. Maßzahlen die ausdrücken sollen, mit welchem Gewicht (bezogen auf eine gegebene Repräsentation der Präferenzordnungen) die Wohlfahrt eines bestimmten Individuums in diesem Maß für Effizienz berücksichtigt werden, sind damit natürlich sinnlos.

Während die bis jetzt betrachteten Indizes also auch bei nur einem Individuum höchstens im Fall homothetischer Präferenzen ein exaktes Maß der Wohlfahrt bilden, in dem Sinne, daß sie die Präferenzordnung des Individuum abbilden, ist der geldmetrische Nutzen in diesem (sehr beschränkten!) Sinne immer ein exaktes Wohlfahrtsmaß.¹² Das konnte zu der Idee führen, Realeinkommen über den aggregierten geldmetrischen (im folgenden Beispiel: indirekten) Nutzen zu definieren:

$$R(p, x) := \sum_i e_i(v_i(p, x_i), p^b) \quad (3.4)$$

¹¹ Wenn 3.3 für alle Einkommen gelten soll, muß diese Implikation insbesondere auch für alle x gelten, bei denen alle Individuen außer einem ein Einkommen von 0 erzielen. 3.3 bedeutet dann nichts anderes, als daß $\beta^m(p)x_i$ eine indirekte Nutzenfunktion des Individuums i sein muß, d. h. die Nutzenfunktionen aller Individuen müssen homothetisch sein. Da $\beta^m(p)$ von i nicht abhängt, müssen die indirekten Nutzenfunktionen aller Individuen durch dieselbe Funktion repräsentierbar sein, d. h. die Präferenzordnungen müssen alle übereinstimmen.

¹² Diese Aussage gilt für geldmetrische Nutzenfunktionen, bei denen ein Basispreisvektor festgehalten wird, nicht aber für die äquivalente (evtl. zirkulär), und noch weniger für die kompensierende (evtl. zirkulär; ordnet im allg. verschiedene Projekte selbst bei gleicher Ausgangssituation anders als die zugrundeliegende Präferenzordnung) Variation.

(Mit v_i als indirekter Nutzenfunktion des Individuums i , e_i seiner Ausgabenfunktion, x_i seinem Einkommen und p^b dem Basispreisvektor.)

Der Glaube, damit in der Suche nach einem wohldefinierten Effizienzmaß einen wesentlichen Schritt weitergekommen zu sein, erweist sich als Illusion, solange keine Angaben darüber gemacht werden, wie man zu dem Basispreisvektor p^b gelangt.

Mit dieser Frage ist ein ernstes Problem benannt: Gäbe es ein Inventar aller möglichen Güter, die in der Welt überhaupt je vorkommen können und für jedes dieser Güter eine eindeutige Meßvorschrift mit definierten Mengeneinheiten, könnte man ein für alle mal einen Basispreisvektor festlegen, den man von da an jedesmal benutzt, wenn es darum geht eine Effizienzordnung auf der Basis des geldmetrischen Nutzens festzulegen. Offen wäre zwar immer noch, wie man den einmal festgelegten Basispreisvektor normativ begründet (da andere Basispreisvektoren zu anderen Effizienzurteilen führen würden), aber es gäbe doch ein relativ einfaches Verfahren, die Definition des Realeinkommens durch Angabe des Basispreisvektors zu vervollständigen.

Allerdings gibt es natürlich kein Inventar aller möglicher Güter. Deshalb müßte die Festlegung des Basispreisvektors – soll sie nicht einfach willkürlich und beliebig sein – auf der Basis eines Verfahrens stattfinden, das aus bestimmten Informationen über die analysierte Situation einen solchen Basispreisvektor auswählt. Nur wenn dieses Verfahren sich auf die Verwendung von Informationen über die individuellen Präferenzordnungen beschränken würde, wäre die resultierende Funktion zur Bestimmung der Effizienzordnung *neutral und ordinal*.¹³ Solange aber ein solches Verfahren nicht angegeben wird, haben wir noch keine Effizienzordnung definiert. Die scheinbare Einfachheit der Vorschrift, den geldmetrischen Nutzen zu addieren, erweist sich als Illusion und es wäre ehrlicher zu schreiben:

$$R(p, x) := \sum_i e_i(v_i(p, x_i), ?) \quad (3.5)$$

Solange wir den geldmetrischen Nutzen nur als einen beliebigen Repräsentanten der Präferenzordnung eines einzelnen Individuums benutzen, genauer, solange wir nur auf die ordinalen Informationen zurückgreifen, die durch den geldmetrischen Nutzen dargestellt werden, ist die Unbestimmtheit des Basispreisvektors kein Problem. Die ordinalen Informationen des geldmetrischen Nutzens werden durch die Wahl des Basispreisvektors nicht berührt.

Die über die Summe des geldmetrischen Nutzens definierte Ordnung sozialer Zustände ist aber sehr wohl vom verwendeten Basispreisvektor abhängig. Dieses Problem wird noch dadurch verschärft, daß die durch die Summe des geldmetrischen Nutzens induzierte Ordnung der sozialen Zustände im allgemeinen lokal nicht quasikonkav ist, wobei die Lage der nicht quasikonkaven Regionen vom Basispreisvektor abhängig ist.

¹³ Neutralität verlangt, daß alle Paare, die von allen Individuen jeweils gleich geordnet werden, auch von der gesellschaftlichen Ordnung gleich geordnet werden. Die Bedingung stellt also sicher, daß nur ordinale Informationen für die gesellschaftliche Ordnung relevant sind. Neutralität impliziert Unabhängigkeit von irrelevanten Alternativen. Vgl. Sen [1970, 72].

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

Das bedeutet aber, daß kleine Veränderungen des Basispreisvektors zu großen Änderungen des sozialen Optimums führen können.¹⁴ [Blackorby u. Donaldson 1988]

Pragmatisch veranlagte Ökonomen mögen vielleicht dazu neigen, das Problem der Wahl des Basispreisvektors als eine theoretische Kuriosität abzutun, um dann mehr oder minder willkürlich irgendeinen Basispreisvektor für konkrete normative Fragestellungen festzulegen. Eine Kandidat dafür ist sicher immer der Preisvektor der Ausgangssituation. Damit hat man aber, und darüber sollte man sich klar sein, das Projekt, ein über verschiedene Fragestellungen hinweg konsistentes – und eventuell gar begründbares – Kriterium sozialer Entscheidungen anzugeben, aufzugeben. Konsequenterweise immer die Ausgangspreise zugrunde zu legen, führt zu Inkonsistenz. Je nach Fragestellung immer wieder mit einem neuen „passenden“ Basispreisvektor zu beginnen, macht die vorgebliche Effizienzordnung zum Ergebnis reiner Willkür.

3.4 Wohlfahrtsverluste: Effizienz und Abstand zur Paretogrenze?

Den folgenden Versuch, eine Effizienzordnung in einer Weise zu bestimmen, die bewußt über das Paretoprinzip hinausgeht, habe ich in der Literatur bisher nirgends dargestellt gefunden. Es ist aber mein Verdacht, daß so etwas als Intuition vielen Überzeugungen von Ökonomen zugrunde liegt. Deshalb scheint es mir berechtigt, diese Intuition einmal etwas genauer darzustellen. Ich beginne mit einem vertrauten Beispiel.

3.4.1 Ineffiziente Transfers

Gemäß einer weit verbreiteten Vorstellung sind aus Gerechtigkeitsgründen unternommene Transfers mit einem Effizienzverlust (im Sinne der Paretoeffizienz) verbunden, wenn nur „verzerrende“ Steuern zur Verfügung stehen, um sie zu finanzieren. Diese Vorstellung ist falsch, wenn Gerechtigkeit durch eine paretoinklusive soziale Ordnung erfaßt wird. Um das zu sehen, betrachten wir ein einfaches Beispiel.

Nehmen wir an, wir hätten eine Gesellschaft vor uns, die aus 2 Individuen bestünde. Unsere Gerechtigkeitsvorstellung sei durch Rawls Maximinkriterium bezogen auf Nutzenniveaus erfaßt. „Außerdem“ sind wir dem Paretoprinzip verpflichtet, ordnen also einen gesellschaftlichen Zustand höher, wenn zumindest eines diesen Zustand höher wertet, während das andere ihn zumindest indifferent sieht.

Nehmen wir an, es stünde das Instrument eines abstrakten Transfers von Zahlungskraft zur Verfügung: Durch Umverteilung könnten wir uns dann entlang der Nutzenmöglichkeitsgrenze bewegen, die bei gegebener Ausgangsausstattung und gegebener Technologie, überhaupt erreichbar ist. Kein Effizienzverlust durch Umverteilung.

Nehmen wir nun aber an, Umverteilung wäre nur durch sogenannte „verzerrende“ Steuern erreichbar. Durch Umverteilung bewegen wir uns jetzt also nicht mehr entlang

¹⁴ Die Tatsache, daß die durch die Summe des geldmetrischen Nutzen induzierte soziale Ordnung in der Regel nicht quasikonkav ist, stellt darüber hinaus diese Summe als *Wohlfahrtsmaß* in Frage, da diese Ordnung dann (zumindest lokal) Ungleichheit bevorzugen würde.

3.4 Wohlfahrtsverluste: Effizienz und Abstand zur Paretogrenze?

der äußersten Nutzenmöglichkeitsgrenze, sondern auf einer weiter innerhalb gelegenen, die die äußerste allerdings an mindestens einem Punkt berührt. Solange nun durch diese Umverteilung die Nutzenmöglichkeitskurve sich nicht nach innen zurückbiegt, wählt unser Kriterium für Gerechtigkeit den Punkt mit Gleichverteilung der Einkommen, der auf der *Paretogrenze dieser Gesellschaft* liegt (in der ja abstrakte Transfers von Zahlungsbereitschaft nicht möglich sind). Auch hier gibt es also keinen „Effizienzverlust“ durch Umverteilung. Nehmen wir nun an, im Bereich der Gleichverteilung würde sich die Nutzensgrenze zurückbiegen, so daß es paretoeffiziente gesellschaftliche Zustände gibt, die den Punkt der Gleichverteilung der Nutzen strikt paretodominieren. Dann würde das auf Nutzen angewandte Maximinkriterium einen dieser paretoeffizienten Zustände wählen.¹⁵ Der entscheidende – und eigentlich selbstverständlich erscheinende – Punkt ist, daß es natürlich keinerlei Effizienzverlust geben kann, wenn eine Gesellschaft um ihre Gerechtigkeitsvorstellungen zu verwirklichen, sich entlang *ihrer* Nutzenmöglichkeitskurve bewegt.

An diesem einfachen Beispiel wird aber vielleicht auch klar, was zu der irrigen Annahme führen kann, Umverteilung sei mit Effizienzverlusten verbunden: Wäre Umverteilung mit abstrakten Transfer von Kaufkraft möglich, dann könnte i.d.R. ein für alle besseres Ergebnis erzielt werden. Steht dieses Instrument der Umverteilung zur Verfügung, dann gehören die auf diese Weise erreichbaren Punkte zur Menge der paretooptimalen, effizienten Allokationen und nicht mehr die Punkte, die durch Umverteilung über verzerrende Steuern erreicht werden können: Diese Punkte sind *dann* ineffizient – allerdings würde uns keine paretoinklusive Wohlfahrtsfunktion zwingen, einen solchen ineffizienten Punkt zu wählen.

Wer Allokationen, die erreichbar sind, wenn nur „verzerrende“ Steuern zur Verfügung stehen, „ineffizient“ nennen, vergißt einfach die Definition der Menge der Paretooptima: die erreichbaren Allokationen, von denen aus niemand verbessert werden kann, ohne jemanden anderen zu verschlechtern. Daß andere, paretoessere Allokationen in einer anderen Welt erreichbar wären, ist wirklich irrelevant dafür (oder sollte es sein), wie wir in der Welt wählen, in der wir tatsächlich leben. „England’s production possibilities would be different if the laws of physics could be disregarded or if we could assume that all workers would do their ‘best’, or ... or.“ [Samuelson 1950] Nur: warum sollen wir nach solchen Möglichkeiten darüber entscheiden, was in unserer Welt effizient und was ineffizient ist?

Mit dem letzten Absatz bin ich schon zur Kritik der Intuition übergegangen, die ich darstellen möchte. Ich vermute aber, daß die verführerischen Kräfte dieser Intuition noch einen Schritt weiter führen, als nur dazu, Paretoeffizienz nicht auf der Basis als erreichbar gedachter gesellschaftlicher Zustände, sondern auf der Basis einer Traumwelt zu definieren. Wenn wir mit Umverteilung beginnen, entfernen wir uns ja mehr und mehr von der fiktiven Paretogrenze. Liegt es nicht nahe, diese wachsende Entfernung als wachsende Wohlfahrtsverluste zu deuten? Diese Idee, Effizienz vom Abstand zur

¹⁵ Das gilt so in einer Zweikonsumentenwelt. Wenn mehr als zwei Konsumenten betrachtet werden, gilt die entsprechende Aussage für das Leximin-Kriterium. Für das Maximinkriterium selbst beachte Anmerkung 1 oben.

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

(fiktiven oder realen) Pareto Grenze her zu definieren, möchte ich im folgenden etwas formaler darstellen.

3.4.2 Effizienz: Abstand formal

A. Sen hat, wie schon erwähnt, den Verdacht geäußert, daß viele Ökonomen insgeheim (nämlich implizit, aber nicht explizit) das Paretoextensionskriterium vertreten. [Sen 1970, 197] Dafür spricht, daß viele Ökonomen normative Urteile nach dem Prinzip „Hauptsache paretooptimal“ zu treffen scheinen, d. h. daß sie paretooptimale Zustände als indifferent betrachten. Außerdem spricht für diesen Verdacht, daß das Paretoextensionskriterium – dem Selbstverständnis vieler Ökonomen entsprechend – zwar nicht ohne über das Paretokriterium hinausgehende Werturteile, wohl aber ohne den expliziten Vergleich von Nutzen auskommt. Allerdings macht die Paretoextensionsregel alle gemäß der Pareto Relation nicht vergleichbaren gesellschaftlichen Zustände indifferent. Nur dann (genau dann) also, wenn ein Zustand pareto besser ist als ein anderer, ist er auch nach dem Paretoextensionskriterium diesem anderen (strikt) überlegen.

In weiten Bereichen der normativ arbeitenden ökonomischen Literatur werden aber auch Zustände, die gemäß der Pareto Relation nicht vergleichbar sind, als verschieden effizient eingestuft. Die im folgenden dargestellten Ordnungen gesellschaftlicher Zustände, würden auch solche Effizienzvergleiche zulassen.

Betrachten wir zunächst die herkömmliche, auf Pareto zurückgehende Definition von Effizienz: Ein gesellschaftlicher Zustand ist effizient, wenn es unmöglich ist, von ihm aus die Situation eines Individuums zu verbessern, ohne die Situation eines anderen zu verschlechtern,¹⁶ d. h. mit anderen Worten, wenn er Element von $P(E)$, der Menge aller paretooptimalen Zustände, ist. Nun wird durch diese Definition zwar bestimmt, was eine effizienter Zustand ist, aber es wird natürlich keinerlei Effizienzordnung festgelegt. Ein Zustand ist effizient oder nicht, aber es gibt keine mehr oder weniger effiziente Zustände.

Die Vorstellung liegt aber doch nahe, daß ein Zustand mehr oder weniger weit von einem effizienten Zustand entfernt sein kann. Warum also nicht versuchen, eine Effizienzordnung über den Abstand von der Pareto Grenze zu definieren.

Die Problematik einer solchen Vorgehensweise – abgesehen davon, daß es sich dabei natürlich um eine Ergänzung des Paretokriteriums handelt – wird schon deutlich, wenn

¹⁶ Pareto sprach von „maximaler Ophelimität für ein Kollektiv“ statt von Effizienz:

“Nous dirons que les membres d’une collectivité jouissent, dans une certaine position, du maximum d’ophelimité, quand il est impossible de trouver un moyen de s’éloigner très peu de cette position, de telle sorte que l’ophelimité dont jouit chacun des individus de cette collectivité augmente ou diminue. C’est-à-dire que tout petit déplacement à partir de cette position a nécessairement pour effet d’augmenter l’ophelimité dont jouissent certains individus, et de diminuer celle dont jouissent d’autres: d’être agréable aux uns, désagréable aux autres. [Pareto 1966, 354]

Ähnlich Pareto [1966, 617 f.] und Pareto [1917, §§ 2128-2130].

Diese Definition unterscheidet sich abgesehen von der Terminologie und Diktion im wesentlichen durch einen Fehler von der im Text verwendeten: Natürlich macht es keinen Sinn, zu verlangen, daß in einem „Maximum“ auch keine *Verschlechterungen* für alle Individuen möglich sind. Vgl. Scapparone [1997, 35] mit einem Hinweis auf Allais [1989, 209 f.].

3.4 Wohlfahrtsverluste: Effizienz und Abstand zur Paretogrenze?

man nur ein paar Minimaleigenschaften einer solchen Effizienzordnung betrachtet.

Betrachten wir also ein beliebiges Abstandsmaß $D(g, E)$, das folgende zwei Mindesteigenschaften erfüllt:¹⁷

1. Das Abstandsmaß soll die Basis für eine paretoinklusive Effizienzordnung bilden können, d. h. es soll gelten:

$$(g \succeq_p g' \rightarrow D(g, E) \leq D(g', E)) \wedge (g \succ_p g' \rightarrow D(g, E) < D(g', E))$$

2. Sei E eine beliebige Entscheidungssituation. Der Abstand aller Elemente von $P(E)$ zu $P(E)$ ist dann gleich:

$$g \in P(E) \wedge g' \in P(E) \rightarrow D(g, E) = D(g', E)$$

Die Effizienzordnung \succeq_e sei dann definiert mit:

$$g \succeq_e g' :\leftrightarrow D(g, E) \leq D(g', E)$$

Die Verwendung dieser Ordnung als Entscheidungskriterium stellt eine radikale Abkehr von der sonst in der „rational choice“ Literatur üblichen Vorgehensweise dar, es ist in jedem Fall unvereinbar mit der Idee einer der Wahlentscheidung zugrundeliegenden und von der *konkreten Entscheidungssituation unabhängigen* Ordnung.

Ein einfaches Beispiel einer Zweipersonenwelt genügt, um das zu demonstrieren. Man betrachte vier gesellschaftliche Zustände a, b, c, d mit $a \succ_1 b \succ_1 c \succ_1 d$ und $c \succ_2 d \succ_2 a \succ_2 b$. Außerdem betrachte man zwei Entscheidungssituationen $E = \{b, c, d\}$ und $F = \{d, b, a\}$. Dann ist $P(E) = \{b, c\}$ und $P(F) = \{d, a\}$. Aufgrund von E müßte also gelten $b \succ_e d$. Aufgrund von F dagegen $d \succ_e b$. Es kann also keine einheitliche und damit von der Entscheidungssituation unabhängige Effizienzordnung geben, die die angegebenen Mindestanforderungen erfüllt.

Das gerade dargestellte Argument läßt sich noch verallgemeinern, indem man die Voraussetzung einer hinter der Entscheidung liegenden Ordnung aufgibt, und nur noch eine von der Entscheidungssituation unabhängige Entscheidungsfunktion verlangt.¹⁸ Mit $C(E)$ als der Menge der aus E gewählten Elemente soll als Mindesteigenschaft der Entscheidungsfunktion also gelten:¹⁹

¹⁷ Meine Verwendung des Begriffs „Abstandsmaß“ sollte nicht über den hohen Grad der Allgemeinheit der folgenden kurzen Ausführungen hinwegtäuschen. Letztlich geht es um jede Effizienzordnung oder weiter unten um jede Entscheidungsfunktion, die die jeweils angegebenen Mindestbedingungen erfüllen.

¹⁸ Als einen Zwischenschritt könnte man noch darauf hinweisen, daß das angegebene Beispiel – allgemeiner – zeigt, daß es keine asymmetrische Relation geben kann, die die geforderten Eigenschaften erfüllt.

¹⁹ Für den Fall einer Effizienzordnung ergibt sich mit $C(E) = \{x \in E : \forall y \in E : x \succeq_e y\}$ diese Mindesteigenschaft aus den oben angegebenen und der Definition von \succeq_e .

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

- Paretooptimale gesellschaftliche Zustände werden in Entscheidungen über Paare bevorzugt:

$$\forall E : x \in P(E) \wedge y \notin P(E) \rightarrow x \in C(\{x, y\}) \wedge y \notin C(\{x, y\})$$

Das oben angeführte Beispiel läßt sich unmittelbar auf Entscheidungsfunktionen übertragen, die die angegebene Mindesteigenschaft erfüllen: Wegen E müßte gelten $b \in C(\{b, d\}) \wedge d \notin C(\{b, d\})$, wegen F dagegen $d \in C(\{b, d\}) \wedge b \notin C(\{b, d\})$, ein Widerspruch.

Aus dieser kurzen Diskussion läßt sich festhalten: Unter Beibehaltung der üblichen Definition von Paretoeffizienz ist es im allgemeinen unmöglich, ein über das Pareto-kriterium hinausgehendes Entscheidungskriterium anzugeben, das die Intuition, daß paretoeffiziente soziale Situationen in irgendeinem Sinn nicht paretoeffizienten Situationen überlegen sind, in ein Kriterium wenigstens von binären Entscheidungen übersetzt. Eine solche Übersetzung wäre nur möglich, wenn die potentiellen Entscheidungssituationen radikal eingeschränkt werden, indem entweder Überkreuzungen der Mengen paretooptimaler Situationen verhindert oder einfacher binäre Entscheidungen generell ausgeschlossen werden.

Die Alternative wäre die Aufgabe der üblichen Definition von Paretooptimalität und der Übergang zu einer Definition von Paretooptimalität auf der Basis einer irgendwie ausgezeichneten Entscheidungssituation. Die so ausgezeichnete Entscheidungssituation müßte dann natürlich angegeben und ihre Wahl begründet werden. Angesichts der Vielzahl von Modellen, in denen normative Urteile formuliert werden und werden sollen, besteht dabei allerdings meines Erachtens in jedem Fall die Gefahr eine ausgeprägten normativen „Adhocery“.

3.5 Das Sezieren einer Wohlfahrtsfunktion: Eine Alternative?

Atkinson u. Stiglitz [1980] können als Kronzeugen für die These dienen, um die es mir im ersten Teil meiner Arbeit eigentlich geht: Das Paretoprinzip ist eine ungeeignete, weil unzureichende Basis für wirtschaftspolitische Entscheidungen.²⁰ Deshalb entscheiden sie sich im normativen Teil ihrer „Lectures on Public Economics“ für einen Ansatz der von einer (allerdings nicht spezifizierten) Wohlfahrtsfunktion ausgeht, um so – abhängig dann von der jeweiligen konkreten Form der Wohlfahrtsfunktion – die Konsequenzen verschiedener, in der Wohlfahrtsfunktion ausgedrückten Wertvorstellungen für die Wirtschaftspolitik aufzeigen zu können. Atkinson und Stiglitz gehen davon aus, daß über das Paretoprinzip hinausgehende Wohlfahrtsfunktionen im allgemeinen irgendeine Form der Vergleichbarkeit von Nutzen verlangen – eine Möglichkeit, die von einem breiten Konsens innerhalb der ökonomischen Profession abgelehnt

²⁰ „On the one hand, the New Welfare Economics does not provide sufficient guidance for public policy decisions – it does not allow a complete ranking of all policies. In practice, governments have to choose between Pareto-efficient outcomes, and only the most die-hard conservative would reject all policies that made any worse off. As Dalton recognized, ‘this is a difficult calculus, but statesman must handle it as best they can, since there is no practical alternative’ (1954, p. 142).“ [Atkinson u. Stiglitz 1980, 352]

3.5 Das Sezieren einer Wohlfahrtsfunktion: Eine Alternative?

wird. Wirtschaftspolitik auf der Basis von Wohlfahrtsfunktionen zu diskutieren, mag deshalb wie eine Rückkehr zur Erbsünde erscheinen.²¹

Obwohl also Atkinson und Stiglitz im normativen Teil ihres Buches jeweils von *einer* als gegeben angenommenen Wohlfahrtsfunktion ausgehen, halten sie es dennoch für möglich, auf dieser Basis eine präzise Unterscheidung zwischen Effizienzzielen und Zielen der Verteilungsgerechtigkeit aufzeigen zu können, auch wenn sie sich bewußt sind, daß diese Unterscheidung im Falle *einer* Wohlfahrtsfunktion, in der beide Gesichtspunkte zusammengefaßt sind, künstlich erscheinen muß.²²

Die in diesem Kapitel bisher angestellten Überlegungen lassen es nun aber von vornherein als sehr zweifelhaft erscheinen, wie so etwas *sinnvoll*, ja bei paretoinklusiven Wohlfahrtsfunktionen sogar überhaupt *möglich* sein soll. Die einzig kohärente und *präzise* Form, ausgehend von einer Wohlfahrtsfunktion einen Trade-off zwischen Effizienz und Gerechtigkeit herzuleiten, wäre es in dieser Sicht, zwei (einander widersprechende) Ordnungen anzugeben, die mit Hilfe eines anzugebenden übergeordneten Kriteriums kombiniert die als Ausgangspunkt gewählte Wohlfahrtsfunktion ergeben. *Wenn* so etwas möglich wäre, wäre es aber zunächst in der Tat künstlich, da ja z. B. in keiner Weise klar ist, warum jemand der in der ursprünglichen Wohlfahrtsfunktion seine Wertvorstellungen ausgedrückt findet, das aufgrund der heraussezierten einander widersprechenden und durch ein weiteres Prinzip zu versöhnenden Ordnungen tun sollte. Außerdem wäre es ja eventuell möglich, die ursprüngliche Wohlfahrtsfunktion auf verschiedene Ordnungen zurückzuführen.

Im konkreten Fall von Atkinson und Stiglitz kommt noch eine weitere Schwierigkeit hinzu: Sie betrachten im normativen Teil ihres Buches hauptsächlich paretoinklusive Wohlfahrtsfunktionen und da, wo sie ihre „präzise“ Interpretation des Trade-offs zwischen Effizienz und Gerechtigkeit ausführen, *nur* paretoinklusive Wohlfahrtsfunktionen.²³ Das führt aber unmittelbar zu dem schon oben erwähnten Problem, daß es zu einem Trade-off nur kommen kann, wenn entweder die heraussezierte Effizienzvorstellung über das Paretoprinzip hinausgeht, oder die heraussezierte Gerechtigkeitsvorstellung nicht paretoinklusiv ist.

Diese vorbereitenden Überlegungen begründen die Erwartung, daß der Versuch von Atkinson und Stiglitz, aus einer gegebenen paretoinklusive Wohlfahrtsfunktion einen Trade-off von Effizienz und Gerechtigkeit herzuleiten, fehlschlagen muß.

Atkinson und Stiglitz versuchen nun allerdings leider gar nicht erst, paretoinklusive Wohlfahrtsfunktionen in eine Effizienzordnung und eine Gerechtigkeitsordnung zu zerlegen. Ihre tatsächliche Vorgehensweise möchte ich am Beispiel ihres Optimalsteu-

²¹ „Some readers will no doubt object strongly to much of the normative analysis of Part Two on the grounds that it makes unwarranted assumptions about the comparability of individual welfares. The great contribution of the New Welfare Economics over the earlier utilitarian analysis of Bentham, Mill, Sidgwick, and Edgeworth was to examine the consequences of not being able to make interpersonal utility comparisons. Thus, in this view, the „New Public Economics“, as represented by Part Two of this book, is a reversion to original sin.“ [Atkinson u. Stiglitz 1980, 351 f.]

²² „The distinction between equity and efficiency is in a sense artificial, since both are subsumed in the objective of maximizing social welfare. [...] We shall therefore [...] provide a more precise interpretation.“ [Atkinson u. Stiglitz 1980, 360]

²³ Vgl. Atkinson u. Stiglitz [1980, 386 ff., 405 ff., 430 ff., 451 ff., 460 f., 494 ff.]

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

ermodells darstellen, das sie S. 386 ff. und leicht verallgemeinert S. 430 ff. vorstellen.

Es gebe H Haushalte, und jeder Haushalt h maximiere $U(\mathbf{X}^h, L^h)$ unter der Nebenbedingung:

$$\sum_i q_i X_i^h = w^h L^h + G \quad (3.6)$$

mit \mathbf{X}^h als Vektor der von h konsumierten n Gütermengen X_i^h und L^h als von h geleisteter Arbeit, q_i als Konsumentenpreise $1 + t_i$, d. h. als Summe der konstant gleich 1 gesetzten Produzentenpreise und der indirekten Steuersätze t_i , Lohnsätzen w^h und einem Lumpsum-Transfer an alle Haushalte G . Die Regierung maximiere eine (indirekte, d. h. auf Basis indirekter Nutzenfunktionen $V^h(\mathbf{q}, w, G)$ formulierte) Wohlfahrtsfunktion:

$$\Psi[V^1(\mathbf{q}, w, G), \dots, V^H(\mathbf{q}, w, G)]$$

unter der Budgetbeschränkung²⁴:

$$R := \sum_i t_i \sum_h X_i^h - HG = R_0$$

Aus der Lagrangefunktion:

$$\mathcal{L} = \Psi + \lambda(R - R_0)$$

erhält man als Bedingungen erster Ordnung:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t_k} &= \sum_h \left[(\lambda - \beta^h) X_k^h + \lambda \sum_i t_i \frac{\partial X_i^h}{\partial t_k} \right] = 0 \quad k = 1, \dots, n \\ -\frac{\partial \mathcal{L}}{\partial G} &= \sum_h \left[(\lambda - \beta^h) - \lambda \sum_i t_i \frac{\partial X_i^h}{\partial G} \right] = 0 \end{aligned}$$

wobei

$$\beta^h := \frac{\partial \Psi}{\partial V^h} \frac{\partial V^h}{\partial M^h}$$

für die marginale soziale Wertschätzung des Einkommens M^h des Haushaltes h steht. Zuletzt führen die Autoren noch den Ausdruck

$$b^h := \beta^h / \lambda + \partial R / \partial M^h$$

als die marginale soziale Nettowertschätzung von Einkommen gemessen in Einheiten von Regierungseinnahmen ein. Dieses b^h soll sowohl den direkten Nutzen von zusätzlichem Einkommen für den Haushalt h , als auch den indirekten Nutzen von dadurch erhöhten Steuereinnahmen messen.

²⁴ R_0 ist das als vorgegeben angenommen Budget.

3.5 Das Sezieren einer Wohlfahrtsfunktion: Eine Alternative?

Da wir hier nicht an Fragen der Optimalsteuertheorie interessiert sind, brauchen wir der Vorgehensweise der Autoren nicht mehr weiter folgen. Das einzige, was uns noch interessiert, ist die Behauptung der Autoren, mit Hilfe von b^h Effizienzkriterien von Verteilungsüberlegungen klar trennen und so den „inhärenten Konflikt“ zwischen diesen beiden Zielen herausarbeiten zu können.²⁵

Dazu identifizieren sie die Situation, daß $b^h = b =$ konstant für alle h , damit, daß die Regierung/Gesellschaft keine Verteilungsziele und im Gegenschluß nur Effizienzziele verfolge. [Atkinson u. Stiglitz 1980, 389, 431]

Zunächst fällt auf, daß es sich bei den b^h um dem Maximierungsproblem der Regierung endogene Größen handelt²⁶, die für bestimmte Ausgangsparameter und individuelle Nutzenfunktionen diese Bedingung für das Nichtvorliegen von Verteilungszielen erfüllen mag, aber in keiner Weise bestimmte Wohlfahrtsfunktionen charakterisiert. Mit anderen Worten, die Regierung mag eine beliebige paretoinklusive Wohlfahrtsfunktion, die ihr Maximierungsproblem über die Bedingungen erster Ordnung lösbar macht, zugrunde legen, für bestimmte Ausgangsparameter und Nutzenfunktionen wird sie nach diesem Kriterium (angeblich) keine Verteilungsziele verfolgen.

Wir können diese Bedingung noch besser verstehen, wenn wir einmal der Regierung neben der Möglichkeit von indirekten Steuern und einer allgemeinen positiven oder negativen Pro-Kopf-Subvention, zusätzlich die Möglichkeit der *Umverteilung*, d. h. positiver oder negativer individueller Transfers T^h unter der Bedingung, daß sich diese Transfers auf Null aufsummieren, einräumen.²⁷

Als Lagrangefunktion erhalten wir dann:

$$\mathcal{L} = \Psi + \lambda [\sum_i t_i (\sum_h X_i^h) - HG - R_0] + \mu \sum_h T^h$$

Daraus ergibt sich als zusätzliche Bedingung erster Ordnung:

$$\frac{\partial \mathcal{L}}{\partial T^h} := \beta^h + \lambda \sum_i t_i \frac{\partial X_i^h}{\partial M^h} + \mu = 0 \quad \text{für } h = 1, \dots, H$$

Das ist natürlich nichts anderes als die von Atkinson und Stiglitz gegebene Bedingung der alleinigen Verfolgung von Effizienzzielen. Zwei Dinge werden jetzt deutlich. Zum ersten haben die Autoren Verteilungsgerechtigkeit über ein *Mittel* definiert: Verteilungsgerechtigkeit ist das, was man im Rahmen der Optimierung einer Wohlfahrtsfunktion über individuelle, sich auf Null aufsummierende Transfers erreicht. Effizienz ist dann offensichtlich einfach alles das, was dann noch zur Optimierung einer Wohlfahrtsfunktion übrigbleibt, d. h. all die Ziele, die man durch Umverteilung nicht er-

²⁵ Was die Autoren offensichtlich für einen Fortschritt ihrer Darstellung gegenüber früheren empfinden, die diesen Gegensatz nicht herausarbeiten konnten. [Atkinson u. Stiglitz 1980, 389 f.]

²⁶ Die b^h sind im nur für die Lösungen der Bedingungen erster Ordnung definiert und sind nur dann relevant, wenn das vorliegende Optimalsteuerproblem sich über diese Bedingungen erste Ordnung lösen läßt.

²⁷ Man beachte, daß die Möglichkeit dieser Transfers zusammen mit der Möglichkeit eines Pro-Kopf-Transfers im Rahmen des Budgets auf die Möglichkeit beliebiger Transfers im Rahmen des Budgets hinausläuft.

3 Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?

reichen kann. Zum andern wird jetzt klar, wann diese Bedingung erfüllt sein wird: dann wenn – völlig unabhängig von der Art der Ziele, die eine Regierung/Gesellschaft verfolgt – sie den Teil der Ziele *erreicht* hat, den sie mit dem Mittel der Umverteilung erreichen kann. Das heißt, Atkinson und Stiglitz erklären uns, daß eine Regierung keine Verteilungsziele verfolgt, wenn sie sie erreicht hat! (Was natürlich irgendwie auch stimmt.) Im oben dargestellten Fall, wo die Regierung nur die Mittel indirekter Steuern und eines Pro-Kopf-Transfers besitzt, heißt das, daß die Regierung dann keine Verteilungsziele verfolgt, wenn sie sie mit Hilfe der ihr zur Verfügung stehenden Mitteln (zufällig!) erreicht.

Diese Überlegungen helfen auch, zum Teil die „eher überraschende Folgerung“ zu verstehen, daß indirekte Steuern mit dem Gerechtigkeitsziel assoziiert seien, während die Kopfsteuer dem Effizienzziel diene.²⁸ Um diese Assoziierung zu begründen, zeigen die Autoren, daß, wenn der Regierung eine Kopfsteuer zur Verfügung steht, und die Bedingung $b^h = b$ (für alle h) erfüllt ist, die indirekten Steuern gleich Null sein werden.²⁹ Aber in Wirklichkeit bedeutet $b^h = b$ (für alle h) im gegebenen Fall nichts anderes, als daß die Regierung mit Erhebung der Kopfsteuer in der Höhe von R_0 zufällig auch schon die bestmögliche Lösung überhaupt getroffen hat. Die Bemerkung, „only if this is felt to be inequitable should indirect taxes be brought into play“, ist im hohen Maße irreführend. In aller Regel wird bei Erhebung einer Kopfsteuer gerade in Höhe $1/H$ des notwendigen Steueraufkommens die Bedingung $b^h = b$ (für alle h) nicht erfüllt sein. Das Problem ist dabei dann weniger, daß diese Situation als ungerecht „empfunden“ werden wird, sondern es ist einfach der Fall, daß durch *Umverteilung* der Zielerreichungsgrad bezogen auf die gegebene Wohlfahrtsfunktion noch verbessert werden kann, was von den Autoren als Vorliegen von Ungerechtigkeit *definiert* wurde.³⁰

Das Argument der Autoren für ihre Form der Zuordnung des Gerechtigkeitsziel zu indirekten Steuern war, daß wenn ihr Kriterium für das Vorliegen eines Gerechtigkeitsziels nicht erfüllt ist (jedenfalls solange der Ansatz der Bedingungen erster Ordnung zur Lösung des Problems ausreicht), auch keine indirekten Steuern erhoben werden sollten. Aber selbst wenn man dieses Kriterium akzeptieren würde, ist diese Folgerung mehr als zweifelhaft. Wenn $R^0 > 0$ und die Möglichkeit von Kopfsteuern nicht zur Verfügung steht, dann müssen indirekte Steuern erhoben werden. In der Regel werden dann die b^h verschieden sein. Wenn jetzt die Möglichkeit von Kopfsteuern hinzukommt, und damit zufällig die bestmögliche Lösung erreicht wird, dann ermöglicht die Kopfsteuer den Übergang von einer Situation der Verteilungsungerechtigkeit in eine

²⁸ Ich sage zum Teil, weil ein Großteil der *Überraschung* das Ergebnis einer semantischen Spielerei ist. Wenn Kaldor, den die Autoren als einen der Kronzeugen der „populären“ Sichtweise, die die direkten Steuern dem Ziel der Gerechtigkeit zuordnet, dies damit begründet, daß man mit indirekten Steuern dieses Ziel nur sehr ungeschickt erreichen könne, und das heißt umgekehrt mit direkten Steuern genauer, so denkt er natürlich bei direkten Steuern nicht an eine Kopfsteuer, die die Autoren hier als Beispiel einer direkten Steuer verwenden! Vgl. Atkinson u. Stiglitz [1980, 428 u. 432].

²⁹ Das Argument setzt voraus, daß der Lösungsansatz über die Bedingungen erster Ordnung ausreicht zur Bestimmung der Optimalsteuern, insbesondere auch, daß die Bedingungen erster Ordnung genau eine Lösung haben.

³⁰ Was nicht notwendig ausschließt, daß auch durch indirekte Steuern der Zielerreichungsgrad noch verbessert werden kann!

3.5 Das Sezieren einer Wohlfahrtsfunktion: Eine Alternative?

Situation der Verteilungsgerechtigkeit. Wäre es dann nicht angebracht, die Kopfsteuer mit dem Gerechtigkeitsziel zu identifizieren?

Noch eine Bemerkung dazu, was ich mit der Behauptung meine, daß eine Situation $b^h = b$ für alle h sich nur „zufällig“ ergeben wird, wenn nur indirekte Steuern und eine einheitliche Kopfsteuer zur Verfügung stehen. Bei einer beliebigen gegebenen pareto-inklusive Wohlfahrtsfunktion bedeutet $b^h = b$ für alle h , daß die optimale Lösung bei Möglichkeit der Umverteilung keine Umverteilung vorsehen würde. Das kann natürlich bei gegebener Wohlfahrtsfunktion nur zufällig, d. h. bei bestimmten Ausgangsparametern (insbesondere bestimmten Lohnsätzen w^h und vorgesehenen Steuereinnahmen R_0) und bestimmten individuellen Nutzenfunktionen der Fall sein.

Wenn wir nur den Fall betrachten, daß der Regierung indirekte Steuern und eine allgemeine Kopfsteuer zur Verfügung stehen, und davon ausgehen, daß die Bedingungen erster Ordnung genau eine Lösung bestimmen, so bedeutet $b^h = b$ immer auch, daß keine indirekte Steuern erhoben werden.³¹ Deshalb ist dann auch $b^h = \beta^h$ und die Bedingung vereinfacht sich zu $\beta^h = \beta$ für alle h oder ausgeschrieben:

$$\frac{\partial \Psi}{\partial V^h} \frac{\partial V^h}{\partial M^h} = \frac{\partial \Psi}{\partial V^j} \frac{\partial V^j}{\partial M^j} \quad \text{für alle } h, j \in I$$

Wenn wir jetzt Variationen der Parameter und eventuell gar der Nutzenfunktionen so zulassen, daß es keine eindeutige Abbildung der Nutzenniveaus auf die Grenznutzen der Einkommen gibt, dann *kann* es keine „neutrale“ Wohlfahrtsfunktion geben, d. h. keine Wohlfahrtsfunktion, die die soziale Wohlfahrt allein auf Basis der individuellen Nutzen ordnet, und die dazu führt, daß bei allen zugelassenen Konstellationen der Parameter (und evtl. sogar der Nutzenfunktionen) die Bedingung des Nichtvorliegens von Gerechtigkeitserwägungen erfüllt ist.

³¹ Siehe dazu gerade oben.

3 *Gibt es einen Trade-off zwischen Effizienz und Gerechtigkeit?*

4 Kritische Anmerkungen zum ethischen Nonkognitivismus

If we disagree about ends it is a case of thy blood or mine.

The Nature and Significance of Economic Science
LIONEL ROBBINS

Hilary Putnam hat die Dichotomie von Werten¹ und Tatsachen als eine gesellschaftliche Institution beschrieben [Putnam 1981, 128 f.]: daß Werte subjektiv sind und sich im Gegensatz zu Tatsachen nicht objektiv begründen lassen, ist eine der Tatsachen, um die man heute einfach weiß.

Wie immer es um die von Putnam beschriebene Diskussionen auf Cocktailparties bestellt sein mag, unter Ökonomen jedenfalls gilt die Unterscheidung als eine Selbstverständlichkeit, der jeder verständige Ökonom zustimmen wird und die deshalb auch keiner sorgfältigen Begründung mehr bedarf.

Diese Selbstverständlichkeit hat große Konsequenzen für den Weg, den die Ökonomie als Wissenschaft gegangen ist: Um sich im hehren Kreis der Wissenschaften zu etablieren, schien es möglich und notwendig, Ökonomie als rein positive Wissenschaft zu betreiben. Die Einbeziehung von Werturteilen in die Ökonomie müßte diese ja als Wissenschaft „zerstören“, da Werturteile von ihrer Natur her subjektiv und nicht objektiv und „wissenschaftlich“ sind.

Ökonomie war aber immer auch eine praktische Wissenschaft: Politik und Öffentlichkeit erwarten von Ökonomen Empfehlungen für wirtschaftspolitisches Handeln. Wie aber soll man zu solchen Empfehlungen kommen, ohne sich auf Werturteile zu beziehen?

Dieses Dilemma hat u.a. zu dem seltsamen Gewächs der Wohlfahrtsökonomie geführt: Aufbauend auf einem als allgemein akzeptiert behaupteten Werturteil (dem Paretoprinzip) hat man versucht, wirtschaftspolitische Empfehlungen herzuleiten. Ein

¹ In verschiedenen Fragestellungen ist es sinnvoll zwischen deontischen Begriffen (z. B. geboten, verboten, erlaubt; sollen, dürfen) und Wertbegriffen (z. B. gut, schlecht, indifferent; besser, schlechter) zu unterscheiden (so Kutschera [1982, 1-20]). Im folgenden werde ich auf eine sorgfältige Abgrenzung und Definition von Wertaussagen, normativen Aussagen, deontischen Aussagen (insbesondere Sollenssätze) verzichten, da sie für die hier verfolgte Fragestellung nicht notwendig sind. Die Zusammenhänge zwischen deontischen Aussagen und Wertaussagen sind so eng (vgl. ebd. 20-24), daß (Non-)Kognitivismus im einen Bereich mit (Non-)Kognitivismus im anderen normalerweise Hand in Hand gehen.

4 Kritische Anmerkungen zum ethischen Nonkognitivismus

weitverbreitetes Werturteil, scheint also den wissenschaftlichen Charakter der Ökonomie nicht in Frage zu stellen.

Die Ökonomen habe die Trennung von Tatsachenurteilen und Werturteilen nicht erfunden, sie haben damit vielmehr eine Überzeugung übernommen, die im Grunde philosophischer (erkenntnistheoretischer) Natur ist, und über einige Zeit vor allem unter analytisch orientierten Philosophen gängige Münze war.² Das hat sich in den letzten Jahrzehnten aber geändert: Selbst unter den analytisch orientierten Philosophen haben sich doch gewichtige Stimmen zu Wort gemeldet, die mit überzeugenden Argumenten (wie mir scheint) der Trennung von Tatsachen und Werten das rationale Fundament entziehen.

Im folgenden möchte ich eine Argumentationslinie darstellen, wie sie in Varianten z. B. von Hilary Putnam und F. v. Kutschera vertreten wird, und die weniger darauf hinausläuft, für ethische Tatsachen eine genauso unbezweifelbare Objektivität wie für Tatsachen der empirischen Wissenschaften einzufordern, als darauf, zu zeigen, daß die rationale Basis, von der aus wir sogenannte empirische Tatsachen akzeptieren, sich nicht wesentlich von der unterscheidet, von der aus wir für ethische Positionen argumentieren.

Zunächst möchte ich die von mir vertretene These präzisieren. Dazu möchte ich von der von mir abgelehnten weitverbreiteten Gegenposition ausgehen. Gemäß dieser Position sind Werten und Erkennen zwei grundsätzlich verschiedene geistige Tätigkeiten. Werturteile sind subjektiv und beruhen auf letztlich nicht weiter hinterfragbaren Einstellungen. Zwar gibt es abgeleitete Werturteile, die aus anderen Werturteilen oder aus der Verknüpfung von Tatsachenurteilen und Werturteilen folgen, aber am Ende des Versuches, Werturteile zu begründen, stehen irgendwelche letzten Werte, die man einfach hat, oder für die man sich entschieden hat, die man in jedem Fall nicht *erkannt* hat.

Tatsachenurteile dagegen sind (wenn sie wahr sind) objektiv. Was das nun heißt, ist nicht ganz so klar und hängt von der jeweils (explizit oder implizit) vertretenen erkenntnistheoretischen Schule ab. Ein Element von Objektivität aber wird es in jedem Fall sein, daß Tatsachenurteile „prinzipiell“ unabhängig von subjektiven Einstellungen erkannt werden können. Es gibt Methoden, Kriterien, die es ermöglichen, unabhängig von Entscheidungen, Werten und Einstellungen der am Prozeß der Erkenntnisgewinnung Beteiligten über die Wahrheit oder Falschheit von Tatsachenurteilen zu entscheiden.

Die von mir hier vertretene Position des Kognitivismus behauptet dagegen, daß es so

² Im ökonomischen Schrifttum stellt sich die Sache oft so dar, als seien der Nonkognitivismus und die Wert-vs.-Tatsachen Dichotomie gesichertes Wissen, auf das man sich mit aller Selbstverständlichkeit berufen könnte. (Die Überzeugung lag etwa als Selbstverständlichkeit dem Bemühen zugrunde, „wert-“ oder „ethikfreie“, „nicht politische“ und „rein wissenschaftliche“ wirtschaftspolitische Empfehlungen herleiten zu können, und wurde von so unterschiedlichen Autoren wie Robbins, Kaldor, Hicks, Scitovsky und Samuelson geteilt. [Vgl. Sen 1970, 56 f.] Auch in die Lehrbuchliteratur und wirtschaftswissenschaftliche Handbücher hat sie Einzug genommen. [Vgl. etwa Rose u. Sauerheimer 1992, 520 und Külpl 1982, 470]) Warum Ökonomen eine *philosophische* Position, die in der Philosophie immer umstritten war, quasi zu einem Dogma erhoben haben, ist eines der Rätsel, die nur wissenschaftshistorische Forschung wird lösen können.

etwas wie *Werterkenntnis* gibt. Ein Werturteil auszusprechen, drückt nicht einfach eine Einstellung des Sprechers, oder der Gruppe und Kultur, der ein Sprecher zugehört, aus, sondern sagt etwas über die bewertete Sache. Werturteile können wahr oder falsch sein³, es gibt Werttatsachen⁴, es gibt ein Wissen bezogen auf Werte.

Diese Position behauptet nicht, daß es keinen Unterschied zwischen Werturteilen und Urteilen der empirischen Wissenschaften gebe. Auch nicht, daß das Erkennen von Werttatsachen einfachhin mit dem Erkennen von z. B. physikalischen oder mathematischen Tatsachen in eins zu setzen sei: Die Annahme, für alle Wissensbereiche seien dieselben Methoden und Vorgehensweisen angemessen, wurde von Aristoteles zu Recht als Merkmal des Dummen apostrophiert.⁵

Die zwischen Kognitivisten und Nonkognitivisten umstrittene Frage ist also, ob es bezogen auf Werte Wissen gibt. Können wir *wissen*, daß ein bestimmtes Bild schön ist, oder, daß einen Mord zu begehen schlecht ist? Um diese Frage zu beantworten, müssen wir zunächst klären, was es eigentlich heißt, etwas zu wissen, und wie wir überhaupt zu Wissen gelangen können.

Etwas zu wissen, heißt etwas, das *wahr ist, aus den richtigen Gründen* (d. h. rational gerechtfertigter Weise), *für wahr zu halten*.⁶ Eine traditionellen Vorstellung davon, wie wir zu Wissen gelangen, geht zunächst einmal davon aus, daß es eine von unserer Erkenntnis unabhängige Realität gibt, bevölkert von Dingen, die wirklich (d. h. unter anderem unabhängig von unserer Erkenntnis) existieren. Ein Satz ist dann wahr, wenn das, was er über ein solches Ding sagt, auch auf dieses Ding in der erkenntnisunabhängigen Wirklichkeit zutrifft. Wissen zu können, würde in dieser naiven Vorstellung bedeuten, daß wir irgendeinen Zugang zu der erkenntnisunabhängigen Realität haben. Um etwas

³ Es hängt in der Tat nicht viel davon ab, ob man daran festhält, daß evaluative Sätze „wahr“ oder „falsch“ sein können, oder ob man statt dessen sagt, sie könnten nur „korrekt“ oder „inkorrekt“ sein, *aber* es seien rationale, kognitive Prozesse, die uns dazu führen können, das eine oder das andere zu behaupten. (So Rescher [1991, 46 ff.].) Trotzdem neige ich eher dazu die Rede von wahr/falsch bezogen auf solche Sätze aufrechtzuerhalten. Normalsprachlich werden solche Sätze gebraucht, um etwas zu behaupten und werden deshalb von normalsprachlichen Sprechern (u.a.) als wahr oder falsch eingestuft. Wenn mit Rescher sowohl an einem „property realism“ (also daran, daß Wertprädikate, die wir bestimmten Objekten zuordnen, diesen wirklich als Eigenschaften anhaften können), als auch an einem „warrant realism“ (also daran, daß Werturteile durch rationale Überlegungen gestützt werden können) festgehalten wird, dann sehe ich keinen Grund, sich einer Sprachregelung zu unterwerfen, die wahr und falsch bezogen auf Werturteile verbietet. Formal läßt sich die Tarski-Semantik von wahr/falsch auf alle Behauptungssätze anwenden. Einwände dagegen wären nur zu erheben, wenn Werturteile pragmatisch nicht dazu verwendet werden könnten (oder faktisch nie dazu verwendet würden), etwas zu behaupten.

Ich folge in diesem Punkt also nicht dem späteren Rawls, der, um den eigenständigen Charakter der *politischen* Theorie der Gerechtigkeit zu betonen, die nicht Teil einer umfassenderen ethischen Theorie oder Weltsicht ist, auf der konsequenten Verwendung von „vernünftig“ (reasonable) statt „wahr“ besteht. Vgl. dazu die Auseinandersetzung Habermas [1995, 122-124, 131] und Rawls [1995, 149 f.].

⁴ Ich benutze hier das Wort „Werttatsachen“ weil es in für Nonkognitivisten paradoxer Weise die Entgegensetzung von Werten und Tatsachen schon sprachlich leugnet. Die verbreitete Alternative von „objektiven Werten“ zu sprechen, ist mißverständlich, da diese Formulierung zu stark an philosophische Positionen erinnert, die die Existenz von einer eigenen Kategorie von „Gegenständen“ – eben Werten behaupten. Trotzdem werde ich der Kürze wegen ab und zu ähnliche Formulierungen verwenden.

⁵ Vgl. Nikomachische Ethik I 3; zitiert bei Putnam [1981, 135].

⁶ Das sind die traditionellen drei Bedingungen für das Vorliegen für Wissen.

zu wissen, müßten wir einfach nachschauen.

Kaum ein Philosoph oder Wissenschaftstheoretiker würde heute eine solche „naive“ Vorstellung vertreten wollen. Viele würden schon die Definition von Wahrheit als Übereinstimmung von Sätzen, Theorien, Überzeugungen mit einer von der Erkenntnis selbst unabhängigen Wirklichkeit als „naiven“ Realismus ablehnen. Selbst wer die eine oder andere Form eines *solchen* (dann natürlich nicht als „naiv“ bezeichneten) Realismus vertritt, wird heute die Vorstellung als „naiv“ abtun, wir hätten irgendwo einen direkten Zugang zu dieser/einer von unsere Erkenntnis unabhängigen Wirklichkeit und müßten einfach nachschauen.⁷

Was eine naiv realistische Erkenntnistheorie unhaltbar gemacht hat, ist akkumuliertes (immer revidierbares) Wissen darüber, wie alle Erkenntnis (theoretisch wie empirisch) vom erkennenden Subjekt abhängt: Von seinem physiologischen Erkenntnisapparat, von seiner Sprache, seiner Kultur, von seinen Theorien (und weiter unten werden wir sehen: von seinen Werten).

Wenn es kein einfaches Nachschauen in einer von unserer Erkenntnis unabhängigen Realität gibt, dann wird deutlich, daß die Beantwortung der Frage „Wann können wir etwas wissen“ zusammenfällt mit der Beantwortung der Frage „Wann sind wir rational gerechtfertigt, etwas für wahr zu halten“.

Eines der zentralen Argumente Putnams [Putnam 1981, 128-135] ist der Versuch zu zeigen, daß das, was wir Wissen nennen, z. B. die Wissenschaften, daß unsere Tatsachen selbst zu einem guten Teil auf Werturteilen aufbauen, und zwar *weil* „*rational rechtfertigbar*“ – *ausgesagt von Theorien oder Sätzen* – *selbst ein Wertprädikat ist*. In diesem Sinn sind Tatsachenurteile und Werturteile hoffnungslos vermischt. Und die Leugnung der Möglichkeit von Wissen über Werte zerstört die Möglichkeit des Wissens über z. B. empirische Tatsachen.

Es gibt eine Reihe von Eigenschaften von Theorien, über die ziemliche Einigkeit besteht, daß Theorien sie besitzen müssen, wenn sie als gute Theorien akzeptiert werden sollen, also rational rechtfertigbar sein sollen. Zu diesen gehören z. B. innere Kohärenz, Einfachheit, Eleganz und technische Effizienz. Diese Eigenschaften sind *Werteigenschaften* von Theorien. Und wenn „rationale Rechtfertigbarkeit“ über (u.a.) solche Eigenschaften von Theorien definiert ist, dann ist sie selbst ein Wertprädikat.

Aber sind Einfachheit und Eleganz wirklich Werte? Während es bei Eleganz kaum möglich ist zu leugnen, daß es sich dabei um einen *ästhetischen Wert* handelt, wäre es doch z. B. bei Einfachheit und vielleicht noch mehr bei innerer Kohärenz möglich, hier von objektiven Eigenschaften zu sprechen, die Theorien haben und die wir Menschen (Wissenschaftler?), wie's der Zufall will, wertschätzen.

Das ist zum einen schon ein großer Schritt in Richtung eines Wertkognitivismus. Denn durch diesen Einwand wird anerkannt, daß es objektive Eigenschaften von „Din-

⁷ Putnam, auf dessen Argumente ich im folgenden häufig zurückgreifen werde, lehnt schon eine Wahrheitstheorie der Entsprechung von Theorie (Satz) und Wirklichkeit als metaphysischen Realismus ab. Dabei stützt er sich auf Probleme der Referenz, die sich bei einer solchen Position ergeben. Ich werde mich im folgenden auf diese Frage nicht festlegen. Im Zusammenhang reicht völlig die Ablehnung eines erkenntnistheoretischen naiven Realismus, eine Position, die (in Schattierungen) heute als eine Art Konsens unter einer großen Zahl von Erkenntnistheoretikern gelten kann.

gen“ gibt, die wir als Menschen (in der Mehrheit? im Allgemeinen?) wertschätzen und die wir deshalb auch oft als Prädikate benutzen, um unsere Anerkennung auszudrücken. So wie eine komplexe (!) Theorie *objektiv einfach* sein kann, so könnte auch ein Gemälde, ein (komplexes) Musikstück objektiv einfach sein. Ästhetische Urteile könnten sich also mindestens zum Teil auf objektive Eigenschaften stützen, von denen wir zusätzlich wissen, daß „wir“ sie schätzen.

Und warum bei ästhetischen Urteilen haltmachen? Kann es dann nicht genauso gut objektive Eigenschaften von Handlungen und Menschen geben, die wir dann zusätzlich auch noch wertschätzen? Warum sollte die Eigenschaft eines Menschen, ehrlich zu sein, weniger objektiv sein als die Eigenschaft einer Theorie, einfach zu sein? Auch ethische Urteile könnten sich damit also zu einem guten Teil auf objektive Eigenschaften von Menschen und Handlungen etc. stützen, von denen wir wissen, daß „wir“ sie schätzen.

Zwar bräuchten wir formal (sowohl im Fall der ästhetischen wie der ethischen Urteile) noch immer „wertende“ Voraussetzungen um von anerkanntermaßen objektiven Eigenschaften wie Einfachheit oder Ehrlichkeit zu Werturteilen wie schön oder gut zu kommen. Aber diese Voraussetzungen hätten dann zum Teil den Charakter von Selbstverständlichkeiten. (Wenn von jemanden objektiv festgestellt wurde, daß er ehrlich, wohlüberlegt in seinen Handlungen, besorgt um sein eigenes Wohl und das derer sei, die von seinen Entscheidungen abhängen, dann wäre es wirklich nur noch ein kleiner Schritt zu sagen, er sei ein guter Mensch.)⁸

Kommen wir noch einmal auf die Einfachheit von Theorien zurück, von der wir für den Moment annehmen, sie sei eine objektive und zusätzlich geschätzte Eigenschaft von Theorien. Wer ist es aber, der diese Einfachheit schätzt? Wenn dieses Wertschätzen nach wie vor subjektiv, idiosynkratisch gedacht (und nicht objektiv, kognitiv) ist, dann hängt noch immer in letzter Analyse die Frage der Akzeptanz von Theorien von subjektiven Werthaltungen ab. Die Trennung von Wertbegriffen wie „Einfachheit“ in eine objektive und eine subjektive Komponente ändert also nichts an der Tatsache, daß die Theoriwahl auch in empirischen Wissenschaften von Werturteilen abhängt.

Putnams Argument gegen den hier diskutierten Einwand, „Einfachheit“, „rationale Rechtfertigbarkeit“ usw. seien eine objektive Eigenschaften, die wir zusätzlich wertschätzen, und keine Wertprädikate, ist es, darauf hinzuweisen, daß diese epistemischen Prädikate zumindest dies mit Wertprädikaten gemein haben, daß sie ähnlich umstritten, ständigem geschichtlichem Wandel unterworfen und kulturabhängig sind. Es gibt also z. B. überhaupt kein allgemein akzeptiertes Verständnis von rationaler Rechtfertigbarkeit. Ohne aber ein Verständnis davon, welche Kriterien rationaler Rechtfertigbarkeit man haben *sollte*, ist es nicht einmal möglich, von bestimmten Theorien als objek-

⁸ Vgl. in etwas anderem Kontext Rescher [1991, 48 ff.], Putnam [1981, 139 ff.]. Weder Rescher noch Putnam vertreten eine „Zweikomponententheorie“ wertender Prädikate, auf die im Text angespielt wird. Putnam weist dagegen zurecht darauf hin, daß wir Prädikate wie z. B. ehrlich in sehr verschiedenen Kontexten gebrauchen, *u.a.* manchmal nur um zu beschreiben, manchmal um zu werten. Der Eindruck eines gewaltigen unüberbrückbaren Grabens zwischen Fakten und Werten geht sicher auch darauf zurück, daß man bei Fakten oft an physikalistisch beschriebene Fakten denkt, nicht aber an Fakten, wie sie z. B. in der Alltagspsychologie aber durchaus auch in der klinischen Psychologie eine Rolle spielen.

tive Tatsache zu behaupten, sie seien im Sinne von ... rational rechtfertigbar. Ohne daß wir wissen, welche Eigenschaften von Theorien *wir* wertschätzen, was *für uns* rationale Rechtfertigbarkeit bedeutet, gibt es keine objektive Rede davon, was andere als rational gerechtfertigt akzeptieren.

Hat man einmal die Idee aufgegeben, man könne die Wahrheit einer Theorie, oder allgemein von Sätzen, dadurch nachprüfen, daß man auf irgendeine Weise in der von unseren Erkenntnis unabhängigen Wirklichkeit „nachschaue“, dann ist objektive Erkenntnis überhaupt nur noch möglich, wenn auch objektive Werterkenntnis möglich ist. Die Position also, es gäbe eine objektive Erkenntnis in einem Bereich wertfreier Tatsachen, aber keine objektive Werterkenntnis ist unhaltbar. Wollen wir überhaupt an der Idee der Möglichkeit objektiver Erkenntnis festhalten, müssen wir auch an der Idee objektiver Werterkenntnis festhalten.

Bis jetzt wurde gezeigt, daß zumindest einige Werte den Charakter objektiver Erkenntnis besitzen müssen. Diese Werte besitzen nun aber keinen direkt ethischen Charakter. Wie steht es also um die Möglichkeit objektiver ethischer Werterkenntnis?

Für die Möglichkeit objektiver *ethischer* Werterkenntnis ist mir kein ähnlich stringentes Argument wie für die Möglichkeit objektiver Werterkenntnis überhaupt bekannt. Aber vielleicht ist ein solches Argument auch nicht mehr nötig, weil viele Einwände gegen die Möglichkeit ethischer Werterkenntnis Einwände gegen die Möglichkeit von Werterkenntnis überhaupt sind und *diese* Einwände können nicht stichhaltig sein.

Auch speziell im Bereich der ethischen Werte ist die Trennung von Werten und Tatsachen nicht so eindeutig und einfach, wie das oft zunächst scheinen mag.⁹ Bei wirklichen Konflikten über Werte und Ziele wird es in der Regel nicht stimmen, daß „wir über die Fakten übereinstimmen und nur in den Werten unterschiedlicher Meinung sind“. Unterschiedliche Wertauffassungen werden mit unterschiedlichen Auffassungen über Tatsachen einhergehen und umgekehrt. Dieser Zusammenhang von Werten und Tatsachen läuft nicht nur in der Richtung von den Tatsachen zu den Werten. Unterschiedliche Wertauffassungen werden auch die Auffassungen davon, was der Fall ist, ändern. Das gilt sicher über das Scharnier der Relevanz: Unterschiedliche Wertauffassungen machen unterschiedliche Tatsachen relevant. Der Korpus des Wissens darüber, was der Fall ist, wird deshalb im Falle deutlich unterschiedlicher Wertauffassungen ebenfalls sehr verschieden sein. Das gilt zwar auch schon in gewissem Grade für die Wahrnehmung der uns umgebenden Natur, aber in noch viel stärkerem Maße für die Wahrnehmung kultureller, sozialer, psychischer Wirklichkeit. Korpus des Wissens meint dabei aber nicht nur eine unterschiedliche Ansammlung von als wahr betrachteten Sätzen: Das ganze Begriffssystem, die Strukturierung der Wirklichkeit wird sich unterscheiden.¹⁰

Solange man eine Auffassung vom Erkenntnisprozeß hat, in der es irgendwelche Möglichkeiten gibt, in der theorieunabhängigen Wirklichkeit nachzuschauen, was der Fall ist, kann man diese unterschiedlichen Sichtweisen der Welt so verstehen, als hätte man aufgrund verschiedener Wertauffassungen auf verschiedenen Seiten des Buches

⁹ Siehe dazu schon oben Anm. 8 und dazugehöriger Textabschnitt.

¹⁰ Putnam illustriert das u.a. am Beispiel von „Super-Benthamiten“ [Putnam 1981, 139-141].

des möglichen Wissens nachgeschaut, und wenn wir, selbst ohne die Werte der anderen zu teilen, uns überreden lassen, auf denselben Seiten nachzuschauen, werden wir dasselbe entdecken, und das kann dann additiv zu *einem* Wissensschatz (der in den Augen der Anhänger der verschiedenen Wertsysteme ziemlich irrelevante Aussagen enthalten wird) zusammengefaßt werden.¹¹ Sobald man aber anerkennt, daß es ein solches Nachschauen nicht gibt, daß unsere Wirklichkeit von uns immer *auch* konstruiert ist (und zwar entsprechend unserer *Ziele*), ist es nicht mehr klar, ob solche aufgrund unterschiedlicher Werte stark unterschiedliche Wissenssysteme überhaupt additiv zusammenpassen. So wie aufgrund unterschiedlicher Auffassungen über das, was rationale Rechtfertigung heißt, Entscheidungen für unterschiedliche, evtl. inkompatible Theorien gefällt werden können, so können auch unterschiedliche Wertesysteme über das Kriterium der Relevanz zu Entscheidungen für inkompatible Theorien führen.

Die letzteren Überlegungen lassen sich sehr schön mit Beispielen aus der Ökonomie illustrieren. Lionel Robbins [1984, XXII f., XXXVI f., 54-63] hat zurecht darauf hingewiesen, daß nicht nur Grundbegriffe der Wohlfahrtstheorie (wie z. B. die Menge der paretooptimalen Allokationen), sondern auch Begriffe der Makrotheorie wie das „Volkseinkommen“ nicht einfach Worte sind für Dinge, die unabhängig von unseren Theorien und Werten „da draußen in der Wirklichkeit“ existieren, sondern soziale Konstrukte, denen wir *auch* deshalb theoretische Bedeutung zumessen, weil sie etwas erfassen, was wir wertschätzen. (Lionel Robbins wollte deshalb allerdings solche Begriffe aus der positiven, „exakten“ Wissenschaft verdammen.) Was Robbins nicht gesehen hat, ist, daß das natürlich auch für die Begriffswelt der Mikrotheorie gilt. Robbins hat sich noch der Illusion hingeeben, daß der Homo oeconomicus evidenterweise das wirkliche Verhalten von Menschen in Knappheitssituationen beschreibt, ging also davon aus, daß „da draußen in der Wirklichkeit“ Menschen existieren, die sich so verhalten, wie’s die Theorie behauptet.

Heute wird eher ein Konsens in dem Sinne bestehen, daß auch der Homo oeconomicus nur eine Modellkonstruktion ist, ohne daß es allerdings eine einheitliche metatheoretische Antwort auf die Frage gibt, *weshalb* diese Konstruktion eigentlich verwendet werden sollte.¹² Aber auch wenn wir hier auf eine Antwort auf diese Frage verzichten und auch darauf, auf die Frage einer kausalen Richtung einzugehen, scheint es doch nicht weit hergeholt, zu sagen, daß methodischer Individualismus in positiven Fragen mit methodischem Individualismus in normativen Fragen Hand in Hand geht.

Ein großer Teil ökonomischer positiver wie normativer Arbeiten (im Bereich der Wachstumstheorie, der Außenhandelstheorie z. B.) verwendet auch heute noch die Konstruktion von Ein-Konsumenten Nationen. Läßt sich das aber überhaupt erklären, ohne die Annahme, daß den entsprechenden Ökonomen Verteilungsfragen und Strukturfragen nicht besonders relevant erscheinen? (Wer meint die Aufgabe der Ökonomie sei es, eine Theorie der Verteilung des Volkseinkommens zu erarbeiten, wird mit

¹¹ Eine solche (hier zugegebenermaßen leicht karikierte) Vorstellung liegt Webers Version von Wertfreiheit der Sozialwissenschaften zugrunde (vgl. Weber [1973b, 155]; Weber [1973c, 499, 511 f.]).

¹² Eine weit verbreitete Antwort ist der Verweis darauf, daß es keine bessere Alternative gäbe. So etwa Binmore [1994, 21]. Weitere Beispiele und eine kritische Auseinandersetzung mit diesem Argument finden sich bei Hausman [1992, 234-244].

4 Kritische Anmerkungen zum ethischen Nonkognitivismus

Ein-Konsumenten Modellen wenig anfangen können.) Ein-Konsumenten Modelle und Modelle mit mehreren verschiedenen Konsumenten führen aber nicht nur „normativ“ sondern auch „positiv“ zu sehr verschiedenen Ergebnissen. Gerade die Verwendung von Ein-Konsumentenmodellen zeigt also sehr klar, wie im Bereich der Ökonomie unterschiedliche Auffassungen darüber, was relevant ist, zu sehr verschiedenen, inkompatiblen Theorien führen können.

Alle Überlegungen bisher hatten zum Ziel, die weitverbreitete Überzeugung, wir könnten unsere Weltsicht sauber in zwei Teile aufteilen, eine Welt der Tatsachen und eine Welt der Werte, wobei vor allem die ersteren völlig unabhängig von der zweiten wären, zu erschüttern. Unsere Welt der Tatsachen ist grundlegend auch durch unsere Welt der Werte bestimmt. Für eine Verteidigung eines Wertkognitivismus ist es aber auch notwendig, aufzuzeigen, wie rationale Werterkenntnis aussehen könnte, und zu zeigen, warum diese Art von Erkenntnis wirklich rationalen Ansprüchen genügt. Solche Vorschläge haben z. B. Putnam, Kutschera, Rescher geliefert. Im folgenden möchte ich – allerdings inspiriert durch diese Autoren – meine eigene Sicht von Werterkenntnis skizzieren.

Grundlegend ist dabei die Einsicht, daß es in keinem Wissensbereich die Möglichkeit gibt, einfach in einer theorieunabhängigen Wirklichkeit nachzuschauen, um zu sehen, was der Fall ist.¹³ Deshalb kann es auch keine absolut sichere Erkenntnis geben: Alle Erkenntnis ruht auf einer gewaltigen Masse von Überzeugungen, die wir einfach für wahr halten, ohne sie je überprüft zu haben. Diese Überzeugungen, die wir faktisch ohne weitere Überprüfung für wahr halten, bestehen aus sehr verschiedenem: Überzeugungen oder Denkstrukturen, die letztlich durch unsere menschliche Art bestimmt sind; Überzeugungen, die wir aus der Tradition übernommen haben; Überzeugungen, die uns sehr sicher erscheinen, die wir deshalb kaum je aufgeben werden; Überzeugungen, die schon durch wenig Evidenz oder einfache Argumente zu erschüttern wären; Überzeugungen, die wir als Teil unserer unmittelbaren Erfahrung verstehen (auch wenn diese Erfahrung selbst durch unsere physiologische Konstitution, durch unsere Sprache und Kultur, durch unsere Traditionen geprägt ist); Überzeugungen, die für uns fundamentale Prinzipien der Wirklichkeit oder des Erkenntnisprozesses darstellen. Der Erkenntnisprozeß kann nun in gar nichts anderem bestehen, als daß wir einige (den Großteil dieser Überzeugungen) festhalten, um an anderen zu arbeiten: Widersprüche, Inkohärenzen, Uneinheitlichkeiten und unnötige Kompliziertheiten auszuräumen; Lücken zufüllen, neue Erfahrungen einzuarbeiten.¹⁴

¹³ Vgl. z. B. Kutschera [1982, 251-255].

¹⁴ Eine Bemerkung zum erkenntnistheoretischen und ethischen Relativismus ist hier am Platz, da die hier vertretene Position mit einer relativistischen Position verwechselt werden könnte. Erkenntnistheoretischer Relativismus ist letztlich selbstwidersprüchlich und deshalb keine besonders attraktive Position [Putnam 1981, 119-124]. Wenn ein ethischer Relativismus mit dem Wunsch, damit Toleranz zu fördern, vertreten wird, dann ist dazu zu bemerken, daß aus einer *rein* relativistischen Ethik sich keine ethischen Forderungen ableiten lassen, deshalb auch keine Forderung nach Toleranz [Putnam 1981, 161 f.].

Die von mir vertretene Position ist zunächst einmal einfach eine realistische Schilderung der Situation, in der wir uns erkenntnistheoretisch befinden: Wer wird z. B. bezweifeln wollen, daß seine theoretischen Überzeugungen auch von den Traditionen mitbestimmt wären, in denen er steht. Was meine

Es wird deutlich, daß so wie ich hier den Erkenntnisprozeß beschrieben haben, positive Erkenntnis und normative Erkenntnis im selben Boot sitzen. So wie wir uns mit einem ganzen System von Überzeugungen u. a. über grundlegende Prinzipien und einfache Erfahrungstatsachen im Bereich z. B. naturwissenschaftlicher Erkenntnis vorfinden, so tun wir das auch im Bereich der Ethik: Wir finden uns vor mit einem System ethischer Überzeugungen (u. a. haben wir einige sehr grundsätzliche, uns unmittelbar einleuchtende Prinzipien) und machen in unserem Alltag Werterfahrungen. Dieses System können wir systematisieren, Widersprüche ausräumen, einfacher gestalten, neue (Wert-)Erfahrungen einbauen etc. – ganz ähnlich wie Erkenntnisfortschritt im Bereich positive Erkenntnis vonstatten geht.

Wenn das eine halbwegs zutreffende Beschreibung der Situation ist, in der wir uns vorfinden, dann wäre zu fragen, ob es gute Gründe dafür gibt, im Sinne des Nonkognitivismus nun einfach auf einen Streich Werteinsichten, Wertüberzeugungen aus unserem System des Wissens zu streichen. Ob also im Prozeß der Erkenntnisgewinnung unsere Weltsicht an Kohärenz, Widerspruchsfreiheit, Einfachheit, Übereinstimmung mit Erfahrungen gewinnt, wenn wir Werteinsichten daraus streichen.

Lebensweltlich würde unser System von Überzeugungen dadurch sicher nicht an Kohärenz sondern im Gegenteil an Inkohärenz gewinnen. Wenn es eine positive Tatsache ist, daß wir Menschen Wertüberzeugungen haben, diese ausdrücken, über sie diskutieren, Ansprüche daraus ableiten, manchmal danach handeln etc., dann kämen wir als Nonkognitivisten in ernste Schwierigkeiten: Wenn wir sagen, dieses oder jenes Verhalten, sei unethisch, dieses oder jenes Gesetz sei ungerecht etc., dann wollen wir ja nicht sagen, wir *empfinden* dieses Gesetz als ungerecht (oder was auch immer wir als nonkognitivistische Rekonstruktion unserer ethischen Äußerungen benutzen wollen). Der ethische Nonkognitivismus würde kohärenterweise von uns verlangen, ohne ethischen Überzeugungen zu *leben*, und das ist etwas, was wir eben nicht tun.¹⁵

Das hier gegebene Argument für die lebensweltliche Inkohärenz, in der uns der Nonkognitivismus zu leben verdammt, hängt davon ab, daß in der Tat der Nonko-

Position von einem Relativismus unterscheidet, ist die prinzipielle Offenheit des jeweils „erreichten“ Wissenstandes für rationale Kritik und Weiterentwicklung nach rationalen Kriterien. Solche Kriterien sind natürlich selbst wieder nicht unabhängig von unserer Kultur etc., aber auch selbst möglicher Revision unterworfen. *Ein* Input in diesem Prozeß sind Begegnungen mit Theorien und Positionen, die von anderen kulturellen Hintergründen und Traditionen geprägt sind.

Letztlich geht es um die Erkenntnis, daß es aufgrund der Sprach-, Kultur-, Traditionsabhängigkeit unseres Wissens kein *sicheres, absolutes, unhinterfragbares* Wissen geben kann. *Wissen* aber kann es geben (wir können rational gerechtfertigt sein, bestimmte Aussagen und Theorien für wahr zu halten, die wahr sind) und wahre Aussagen und Theorien kann es geben: Z. B. ist die Aussage „Schnee ist weiß“ wahr, denn Schnee *ist* weiß.

¹⁵ Hier liegt die eigentliche Basis für das, was als sogenannte Versuche der Letztbegründungen von Ethik, z. B. aus den Voraussetzungen der wissenschaftlichen Kommunikation vorgelegt wurde. Die wissenschaftliche Kommunikation setzt (als eine *Praxis*) faktisch gewisse ethische Ansprüche und Zugeständnisse voraus, ohne die sie keinen Sinn mehr machen würde. In Wirklichkeit ist das aber nur ein Bereich, wo wir in unserem faktischen Verhalten bestimmte ethische Normen zugrundelegen (müssen). Nur springt die Inkohärenz eines Teilnehmers am wissenschaftlichen Diskurs, der gleichzeitig (implizit) die Anerkennung gewisser ethischer Ansprüche verlangt und ihre Basis leugnet, besonders deutlich ins Auge.

gnitivismus das Ende einer jeden Ethik bedeutet. Diese Konsequenz wird von einer ganzen Reihe von Nonkognitivisten genau so geteilt, so z. B. von Ayer und Nowell-Smith [Kutschera 1982, 96-99]. Von anderen aber wird diese Konsequenz geleugnet, wie z. B. im deutschen Sprachraum von Trapp [1988].

Wenn wir Ethik als die Antwort auf die Frage „Was sollen wir tun?“ verstehen, dann ist der Nonkognitivismus nichts anderes als die Behauptung, daß es auf diese Frage keine rational begründbare und wahrheitsfähige Antwort – und damit keine Ethik geben kann. Nehmen wir nur zum Beispiel die weit verbreitete Position des Dezisionismus, die behauptet für irgendwelche letzten Werte müßten¹⁶ wir uns einfach *entscheiden*, ohne daß diese Entscheidung selbst noch einmal als gesollte verstanden werden kann. Aber wir entscheiden uns für das, was wir *wollen*, und nicht für das, was wir *sollen*. Wenn ich dagegen frage, was ich tun soll, dann frage ich eben gerade danach, was mir jenseits meines Wollens (vielleicht unter seiner Berücksichtigung) als ein Anspruch begegnet. Als Antwort auf das Wissen darum, was ich soll, kann ich mich *dann* entscheiden, das dann auch zu wollen (und zu tun).

Wen dieses einfache Argument nicht überzeugt, für den können wir noch genauer untersuchen: Läßt sich aus einer Entscheidung für bestimmte Basiswerte oder Basisnormen irgendein Sollen ableiten?

Ich habe mich (unter anderem aufgrund meiner verquerten Tatsachenauffassungen) für die ethischen Forderungen des Nationalsozialismus entschieden¹⁷. Gibt es als Folge davon nun irgendeine ethische Forderung an mich? Wenn ich mich für das (perverse) Ziel wirklich entschieden habe, Europa judenfrei zu machen, dann werde ich faktisch, soweit es in meiner Macht steht, dazuhelfen. Aber *soll* ich das? Der erste Aspekt dieser Frage ist, ob jemand, der sich für irgendein ethisches System entschieden hat, dann wirklich ethisch nicht nur gerechtfertigt sondern sogar dazu verpflichtet ist, diesem System entsprechend zu handeln, und zwar so, daß dieses Gerechtfertigt- und Verpflichtetsein auch von denen anzuerkennen ist, die andere Entscheidungen getroffen haben. Beispiele wie die des Nationalsozialismus zeigen klar, daß das natürlich nicht der Fall ist, nicht der Fall sein kann. Daraus folgt aber ganz allgemein, daß die Entscheidung für ein gewisses ethisches System kein irgendwie geartetes Sollen außerhalb dieses Systems begründet. Das ist auch für weniger abstruse ethische Systeme völlig klar: Ein solches Sollen müßte ja ein Sollen sein, für das ich mich nicht entschieden habe, und das ist im Dezisionismus explizit ausgeschlossen. Alle Sätze der Art „A soll x, denn x ist eine Konsequenz der ethischen Prämissen, für die er sich entschieden hat“ sind im Rahmen des Dezisionismus unhaltbar, wenn sie mehr sein wollen als eine *Beschreibung* des ethischen Systems (inklusive seiner Konsequenzen), für das sich A entschieden hat.

Der zweite Aspekt der Frage ist es aber, ob ich selbst, weil ich mich für ein ethisches System entschieden habe, nun aus meiner Binnenperspektive verpflichtet bin, diesem System und seinen Konsequenzen zu folgen.

¹⁶ Dieses „müssen“ darf im Rahmen des Dezisionismus nicht ethisch, als ein „Sollen“ verstanden werden, da dieses ja dann einer Entscheidung vorausliegen würde. Es kann nur entweder existentialistisch als eine Unausweichlichkeit („Wir sind dazu verdammt, uns ethisch zu entscheiden.“) oder rein faktisch deskriptiv („Menschen entscheiden sich für ethische Überzeugungen.“) verstanden werden.

¹⁷ Oder für die Durchsetzung Großserbiens, oder die Vernichtung von Tutus ...

Einen Fall sollten wir gleich ausschließen: Wenn ich Kognitivist bin (oder aus der Sicht des Dezipionisten der kognitivistischen Illusion unterliege), mich also für das ethische System E entschieden haben, weil ich glaube, daß ich das aufgrund bestimmte Gründe soll, weil das die ethisch richtige Entscheidung ist, dann werde ich selbstverständlich der (möglicherweise illusionären) Überzeugung sein, daß ich diesem System und seinen Konsequenzen in meinem Handeln nun auch folgen soll.

Wie aber, wenn ich Nonkognitivist bin, also „weiß“, daß ich mich für das System E eben entschieden habe? Nun, wenn mir aus irgendeinem Grund (z. B. irgendeiner Konsequenz, die mir nicht paßt) das System E nicht mehr gefällt, entscheide ich mich eben um. Warum *sollte* ich das nicht tun? Dieses Sollen könnte ja selbst nur ein Teil von E sein, für neue Entscheidungen offen wie jeder Teil von E. Man könnte einwenden, daß wir Menschen *faktisch* unsere ethischen Systeme nicht ständig wechseln. Das mag der Fall sein, aber wie jeder Nonkognitivist sich beeilen sollte dazuzufügen: Aus einem „ist“ folgt kein „sollen“. Daß wir normalerweise bei unseren ethischen Systemen über einige Zeit bleiben, begründet nicht, warum ich mich nicht umentscheiden soll. Zumal diese „Umentscheidung“ nichts radikales sein muß: Unsere ethischen Überzeugungen haben entschieden nicht den Charakter von axiomatischen Systemen, aus denen dann für jede mögliche Situation eindeutig logisch folgt, was wir tun sollen. Ethische Überzeugungen müssen auf verschiedene Situationen hin immer wieder neu ausgelegt und interpretiert werden. Ich entscheide mich also zum Zeitpunkt t für ein ethisches System E. Wäre ich zu diesem Zeitpunkt gefragt worden, welche Konsequenzen aus diesem System für die Situation S folgen, hätte ich geantwortet: die Handlung h. Zum Zeitpunkt t' befinde ich mich in der Situation S, die Konsequenz h aber gefällt mir nicht. Vielleicht reicht es jetzt aber, E einfach ein kleines bißchen anders zu interpretieren und schon erspare ich mir h. Warum soll ich das nicht tun? Weil es gegen den „Geist“ von E verstößt? Woher weiß ich das? Weil ich so etwas zum Zeitpunkt t gesagt hätte? Und warum soll ich dem „Geist“ von E folgen? Weil ich mich dafür entschieden habe? Wenn ich mich aber nicht dafür entschieden habe? Und ganz unabhängig von der spezifischen Position des Dezipionismus, ist es überhaupt vernünftig, mich in der Weise an ein ethisches System zu binden, daß ich mich an alle Konsequenzen binde, die ich zum Zeitpunkt der „Entscheidung“ für dieses System daraus gezogen hätte?

Es ist wohl klar genug geworden, daß aus einer Entscheidung für ein ethisches System weder extern noch intern ein wirkliches Sollen folgt: Auf der Basis einer Entscheidung für ein ethisches System gibt es über eine Beschreibung dieses Systems hinaus keine Antwort auf die Frage: Was sollen wir tun? Der Grund dafür sei noch einmal wiederholt: Entscheidungen treffen wir aufgrund dessen, was wir wollen. Wenn wir anders wollen, werden wir deshalb auch unsere ethischen Entscheidungen anders treffen.

Als Illustration der Unmöglichkeit auf nonkognitivistischer Basis eine Ethik aufzubauen, sei es mir erlaubt, noch ein paar Bemerkungen zu Trapps entsprechenden Versuch zu machen. Im ersten Kapitel seines Werkes [Trapp 1988] versucht Trapp zunächst einmal, die Position des Nonkognitivismus zu begründen (nicht überzeugend, wie mir scheint). Trapp ist Dezipionist und in seiner Version des Nonkognitivismus kann man zu ethischen Positionen nur gelangen, indem man sich für „nicht weiter begründbare

normative Axiome“ entscheidet. „Normative Urteile“ seien nämlich „allein deduktiv überhaupt begründbar“ [Trapp 1988, 59].

Und was macht Trapp dann? Dann stellt er uns sein ethisches System auf der Basis seiner ethischen Axiome vor, nicht aber etwa mit einer Bemerkung wie: „Dafür habe ich mich entschieden. Sind sie nicht schön, attraktiv?“, die seine Entscheidung und einen Appell ausdrücken würde, ihm zu folgen, sondern er *begründet*¹⁸ sie mit mehr oder minder weit ausholenden Argumenten, der Art wie man sie von einem Kognitivist erwarten würde. Trapp kündigt das als einen „systematisch begründeten Appell an jeden einzelnen Adressaten, sich in freier Entscheidung zu ihren [der Ethik] obersten Normen und Prinzipien zu bekennen“ an. Aber das versichere ich dem Leser, daß das vor allem ein Appell an seinen Intellekt ist, an seine Fähigkeit, rationalen Argumenten zu folgen und ihren Konsequenzen zuzustimmen, falls sie *überzeugen*. Es dürfte Trapp sehr schwer fallen, klar zu machen, weshalb sich seine Argumente grundsätzlich von Argumenten unterscheiden sollten, wie sie etwa in der Diskussion von Grundlagenfragen auch der „härtesten“ Wissenschaften auftreten, in Wissenschaftstheorie und Erkenntnistheorie (man vergleiche nur die Argumente in Kapitel I seines Buches: Tatsachenfrage: Gibt es ethische Erkenntnis? und in Kapitel II seines Buches: normative Frage: Welches sind die normativen Axiome, auf die wir uns stützen sollen?), und allgemein in fast allen Wissenschaften vom Menschen auftreten.¹⁹

¹⁸ So wörtlich in der Überschrift des entsprechenden Gesamtabschnittes und auch der auf einzelne Axiome bezogenen Unterabschnitte: „Erläuterung und Begründung der notwendigen Bedingungen von GU“, „Begründung und Erläuterung von P₁“ etc. [Trapp 1988, 310 ff.]. Wie diese natürlich *nicht* deduktiven „Begründungen“ von Axiomen mit der Versicherung zu verbinden sind, es gäbe nur deduktive Begründungen normativer Sätze, bleibt mir ein Rätsel. Sicher: Es handelt sich hier nur um einen kleinen semantischen Fauxpas, aber doch einen verräterischen, wie mir scheint.

¹⁹ Ein Beispiel: Trapp [Trapp 1988, 192-199] möchte zeigen, daß das Habermassche „Verallgemeinerungsprinzip“ inhaltlich abzulehnen sei. (Was nach Trapp auch die Begründungsansprüche dieses Prinzips, d. h. die metaethische, nicht selbst normative Frage, ob Normen sich begründen lassen, in Frage stellen soll!) Zur Einleitung dieses Arguments gebraucht er eine seiner für einen Nonkognitivist doch eher überraschenden Formulierungen: „Nicht nur die Berechtigung bestimmter Begründungsansprüche hinsichtlich des Prinzips U [des Verallgemeinerungsprinzips], sondern dessen normative Geeignetheit betrifft schließlich eine viel wichtigere weitere *Tatsache*: Es gibt *Gründe*, [...] auch U selbst als inhaltlich nicht generell annehmbar anzusehen.“ (Hervorhebungen vom Verfasser, A.G.) Als einen dieser Gründe versucht er dann zu zeigen, daß dieses Prinzip möglicherweise selbst eine so grundlegende Norm wie „das allgemeine prima-facie-Tötungsverbot anderer Menschen“ in Frage stellen könnte. Das ist nun sachlich und in der Formulierung eine Vorgehensweise, mit der ein Kognitivist nur zufrieden sein könnte: Um ein Prinzip (ein Axiom?) in Frage zu stellen, zeige ich, daß es Konsequenzen hätte, die eine grundlegende Wertintuition verletzen würde. Trapps Argument baut darauf, daß seine Leser (wie er selbst) diese Intuition *haben*, und es deshalb für widersinnig halten würden, wenn diese Einsicht verletzt wird. Wenn Trapp nun behaupten möchte, dies sei keine Einsicht, sondern eben eines der Grundprinzipien, für die man sich entschieden habe (Warum zählt er dieses Prinzip dann nicht unter seinen Prinzipien auf? Wann habe ich mich für dieses Prinzip entschieden, das ich für eine Einsicht halte, die in keiner Weise in meiner Disposition steht?), dann muß man nur auf die strikte Parallelität mit entsprechenden nicht-normativen Kontexten verweisen. Warum steht hinter der Intuition, daß es nicht erlaubt sein kann, ohne weitere gute Gründe Menschen nach Belieben umzubringen, eine Entscheidung, ein nicht-kognitiver Prozeß, hinter der Intuition, daß die Interaktion zwischen räumlich voneinander entfernten Teilchen höchstens mit Lichtgeschwindigkeit erfolgen kann (die viele Physiker dazu gebracht hat, die Bohmsche Interpretation der Quantenphysik abzulehnen) eine Einsicht, ein ko-

Welche positiven Gründe für die Aufgabe des Wertkognitivismus könnten nun aber angeführt werden?

Als erstes käme da wohl Humes Gesetz in Frage, also die These, daß sich Werte aus Tatsachen nicht ableiten lassen. Aber warum soll das ein Einwand gegen den Wertkognitivismus sein? Wie wir gesehen haben: Einige Werte brauchen wir, selbst um unsere Tatsachenwelt zu konstruieren. Wenn es aber diese, aus Tatsachen nicht ableitbaren Werte gibt, warum soll es dann nicht andere geben? Es gibt viele Bereiche des menschlichen Wissens, die aus anderen Bereichen nicht ableitbar sind: Aus unseren phänomenalen Erfahrungen ist die Existenz einer Welt außerhalb dieser Erfahrungen nicht ableitbar [Putnam 1981, 142]. Aus unserem gesamten empirischen Wissen ist die Mathematik nicht ableitbar. Aus unserem naturwissenschaftlichen Wissen sind keine Theorien der Referenz oder der Wahrheit ableitbar [Putnam 1981, 22-48]. Werte nur anerkennen zu wollen, wenn sie aus „wertfreien“ Fakten ableitbar sind, ist eine eher willkürliche Vorentscheidung gegen den Wertkognitivismus.

In meiner Darstellung von Werterkenntnis spielten Werterfahrungen und Wertintuitionen eine Rolle. Ein Einwand gegen den Wertkognitivismus könnte es sein, dahinter einen seltsamen, physiologisch nicht lokalisierbaren „achten“ ethischen Sinn zu vermuten. Was Wertintuitionen angeht, braucht man dagegen nur auf die Rolle mathematischer Intuitionen zu verweisen: Gewisse mathematische und logische Axiome erscheinen dem ausgebildeten Mathematiker und zum Teil auch dem mathematischen Laien intuitiv einleuchtend, und u.a. deshalb haben Mathematiker darauf mathematische oder logische Systeme gebaut. Diese mathematischen Intuitionen setzten aber keinen besonderen mathematischen Sinn voraus, sondern hängen mit fundamentalen Strukturen unseres Geistes und vielleicht auch unserer Sprache zusammen. Ähnlich

gnitiver Prozeß? Oder ist das letztere auch eine Entscheidung? Dann spielen aber nicht nur im Bereich normativer Fragen, sondern auch in manchen Bereichen der modernen Physik – der Entscheidung für und wider bestimmter physikalischer Theorien – nicht-kognitive Prozesse eine *entscheidende* Rolle? Ist Trapp ein Vertreter einer dezisionistischen Physik?

Oder nehmen wir eine Intuition, die in Trapps Argument gegen den Wertkognitivismus eine Rolle spielt: Ist es eine grundlegende epistemologische Einsicht (kognitiv) oder eine Entscheidung, wenn Trapp relativistische Vergleichskriterien ablehnt, weil sie zu der unannehmbaren Folge führen würden, daß nach diesem Kriterium „Theorien“ (wie die eines beliebigen Sektenführers), wenn sie nur in sich konsistent sind, in ihrem Rationalitätsstatus von der modernen Mikrophysik nicht mehr unterscheidbar wären? Trapp macht in diesem Zusammenhang den „kaum als extravagant zu bezeichnende[n] Vorschlag: Man sollte in der Philosophie Vergleichskriterien immer mit einem Blick auf die Folgen wählen“. Ein *Vorschlag* ist ein etwas abgeschwächter Appell? Ein Appell an die Freiheit des Erkenntnistheoretikers, der Vergleichskriterien *wählt*? Ist die Frage Kognitivismus vs. Nonkognitivismus (keine normative Frage!) also eine Frage der Entscheidung?

Es zeigt sich hier wie an anderen Stellen in Trapps Argumentationsgang eine weit verbreitete Tendenz, Wissen und Wissenschaft paradigmatisch auf Naturwissenschaften und insbesondere der Physik zu beschränken. Mein bescheidener Vorschlag wäre es, die Kriterien für Rationalität nicht so zu wählen, daß dann plötzlich nur noch *Teile* der Naturwissenschaften als Bereich von möglicher kognitiver Erkenntnis übrigbleibt. Vor allem sollte ein Philosoph, der einen Nonkognitivismus in der Ethik und einen Kognitivismus in der Mikrophysik vertreten möchte, nicht – bei konsequenter Anwendung *seiner* Kriterien – bei einem Nonkognitivismus der Erkenntnistheorie (speziell der Begründung des nicht-kognitiven Status der Ethik *und des kognitiven der Mikrophysik*) landen. Dann hätte vielleicht doch *er* zu viel bewiesen.

4 Kritische Anmerkungen zum ethischen Nonkognitivismus

kann man auch bezogen auf die unmittelbare, auf den Einzelfall bezogene Werterfahrung argumentieren: Wenn das Kind gegen die ungleiche Aufteilung des Kuchens als ungerecht protestiert, dann braucht es dazu keinen „achten“ Sinn, sondern es zeigt, daß es den Gebrauch des Wortes „ungerecht“ korrekt gelernt hat, und wie tief in unserer *kognitiven* Struktur die „Forderung“ nach Gleichbehandlung des Gleichen verankert ist.

Ein letztes Argument könnte noch der pluralistische Charakter der Werterkenntnis sein. Ist es nicht so, daß auf der Ebene ethischer Theorien, auf der Ebene der Prinzipien, und auch auf der Ebene elementarer Werterfahrungen eine so große Diversität besteht, daß im Unterschied von Erfahrungswissenschaften eine intersubjektive Verständigung auf gemeinsame Werttheorien auch als eine Art Grenzbegriff vollständig illusorisch sein muß? Und ist nicht die Möglichkeit, daß verständige Menschen zum selben Ergebnis kommen müßten, einer der Aspekte, die unseren Begriff von Wissen zentral prägen?

Dieser Einwurf ist aus mehreren Gründen heraus nicht sehr überzeugend. Auch wenn ich zugestehen möchte, daß unsere empirische Erfahrung physikalischer Objekte *wahrscheinlich* einheitlicher – und deshalb eine Verständigung über solche Erfahrung leichter möglich – als bei Werterfahrungen ist, handelt es sich bei allen anderen Vergleichen in diesem Zusammenhang um sehr komplexe empirische Fragen, die zugleich schwere konzeptionelle Fragen aufwerfen.

So ist z. B. zu klären, wer denn unter die verständigen Menschen gehört, unter denen da Einigkeit erzielt werden soll: Sind das im Grunde alle Menschen? Dann ist es aber eine empirisch offene Frage, ob es nicht doch mehr Menschen gibt, die von bestimmten ethischen Werten (fundamentale Menschenrechte, Demokratie, Wohlstand) überzeugt sind, als etwa von den Grundprinzipien der modernen Physik. Oder ist es nur die Gruppe der im entsprechenden Gebiet arbeitenden Wissenschaftler? Hier ergibt sich aber das Problem, daß die Zugehörigkeit zu dieser Gruppe im wesentlichen über die Übernahme gewisser Methoden, Grundüberzeugungen, „Wissen“ definiert wird. Aber selbst wenn wir die jeweiligen wissenschaftlichen Experten als Referenzgruppe akzeptieren, so würde man vielleicht zugeben müssen, daß unter Physikern etwa größere Einigkeit besteht als unter Ethikern. Aber wie ist es mit Psychologen, Soziologen, Politologen? Allgemein mit allen Wissenschaften, die sich mit komplexen spezifisch menschlichen Phänomenen beschäftigen. Ist da die Einheit wirklich größer als unter Ethikern? Und wenn wir etwa von den empirischen Grunddaten eines großen Teils der Psychologie sprechen, sind das Erfahrungen, auf die man sich leichter einigen kann, als die Erfahrungen, die ethischer Reflexion zugrundeliegen? Zumal, da es sich dabei ja zu einem guten Teil um die gleichen Erfahrungen handeln wird?

Ich glaube es wird klar, wenn man anfängt zu fragen, wieviel Einigkeit über bestimmte Fragen besteht, und wohl auch in nächster Zeit bestehen wird, dann wird die Trennlinie *nicht* einfach zwischen Werten und „wertfreien Tatsachen“ verlaufen.

Eine noch grundsätzlichere Frage ist es, ob das Erreichen einer einheitlichen Theorie, der alle zustimmen, überhaupt *fundamental* für unser Wissen in dem Sinne ist, wie es dieser Einwand behauptet, oder ob es nicht eher eine der Grenz-, oder Leitideen ist, die unserem wissenschaftlichen Suchen eine Richtung geben. Faktisch sind wir in vie-

len Wissensgebieten weit entfernt davon. Selbst in so paradigmatischen Wissenschaften wie der Physik und der Mathematik gibt es diese Einigkeit nicht. Und ist das noch verwunderlich, wenn die Grundthese, auf die ich hier vertrete, korrekt ist, daß unsere Wirklichkeit, alle unsere Theorien, all unser Wissen immer eine Konstruktion ist, die auch auf Werten aufbaut? Wenn z. B. Theorien über soziale Phänomene auch davon bestimmt sind (inhaltlich!) was wir für relevant halten, muß es dann nicht unterschiedliche sozialwissenschaftliche Theorien geben, solange es unterschiedliche Wertauffassungen gibt?

Zusammenfassend scheint es also keine überzeugenden Gründe dafür zu geben, den Wertkognitivismus aufzugeben, um dafür eine fundamentale Inkohärenz zwischen unserer philosophischen Position zu Werten (Nonkognitivismus) und unserer lebensweltlichen Überzeugungen (im Alltag, in der Politik, im Wirtschaftsverkehr und im Wissenschaftsbetrieb) in Kauf zu nehmen.

4 Kritische Anmerkungen zum ethischen Nonkognitivismus

5 Eine Theorie der Gerechtigkeit

Ich beginne mit dem Versuch einer konzentrierten Darstellung des Rawlsschen Entwurfs einer Theorie der Gerechtigkeit. Es wird unvermeidlich sein, daß ich mich dabei auf einige wichtigen Aspekte des reichen Werkes von Rawls beschränke. Vier Kriterien werden mich bei der Auswahl leiten: Es soll ein möglichst treues Bild der Position Rawls' entstehen, es sollen einige der Mißverständnisse, die sich in einen Teil der ökonomischen Literatur eingeschlichen haben aufgeklärt werden; eng damit verknüpft soll schon die Darstellung einige populäre Einwände beantworten; und schließlich soll deutlich werden, was innerhalb der Konzeption Rawls' die Aufgabe der Ökonomie sein könnte.

Eines will die hier vorgelegte Darstellung explizit *nicht* leisten: Die Position Rawls' in ihrer geschichtlichen Entwicklung nachzeichnen. Ich beschränke mich im wesentlichen auf das 1971 erschienene Hauptwerk, „A Theory of Justice“ [Rawls 1973]¹, und ziehe frühere und spätere Werke nur heran, wenn sie Ausführungen in diesem Hauptwerk ergänzen, bestimmte Punkte verdeutlichen, Lücken in der Argumentation füllen, oder auf – meines Erachtens – plausible Weise Einzelpunkte auch korrigieren, ohne die Grundposition zu gefährden.² Für solche Ergänzungen benutze ich hauptsächlich „Justice as Fairness. A Restatement“ [Rawls 2001], ein Werk, das man als Rawls' Darstellung „von letzter Hand“ seiner Theorie betrachten kann.

5.1 Der Gegenstand der „Theorie der Gerechtigkeit“

Schon in „A Theory of Justice“ hat Rawls großen Wert darauf gelegt, seinen Lesern deutlich zu machen, daß der Gegenstand seiner Theorie einen „Spezialfall des Gerechtigkeitsproblems“ bildet. Er möchte „einen vernünftigen Gerechtigkeitsbegriff für die

¹ Sowohl in der Urfassung wie in der Fassung des von Rawls revidierten Textes [Rawls 2001], der der deutschen Ausgabe von 1975 [Rawls 1975] zugrunde liegt. Um den Text lesbarer zu gestalten, werde ich im folgenden zum Teil aus dem deutschen Text zitieren. Die Unterschiede zwischen den beiden Fassungen herauszuarbeiten, überlasse ich den Rawls-Exegeten. Im Interesse der Lesbarkeit werde ich mich in diesem Abschnitt auch darauf beschränken, jeweils nur am Ende eines Absatzes *zusammenfassende* Hinweise auf Belegstellen zu geben.

² Mir scheint, daß diese Weise der Lektüre den Schriften, die Rawls nach der Theorie der Gerechtigkeit veröffentlicht hat, im Ganzen durchaus gerecht wird. Zu den wenigen Ausnahmen gehört m.E. die Neupräsentation der Theorie der Gerechtigkeit als einer dezidiert *politischen* Theorie, die nicht Teil einer umfassend-moralischen Theorie ist, und die Konsequenzen, die Rawls daraus zieht: neben einer Klärung im Verständnis der Grundgüter besonders die Einbettung der Theorie in das Selbstverständnis moderner, pluralistischer Gesellschaften und die damit einhergehende Einschränkung des Anspruchs, der im Hauptwerk durchaus noch als universal verstanden werden kann. [Vgl. Rawls 2001, XVI-XVIII, 57-61] Die Neuformulierung der zwei Prinzipien und die Neustrukturierung des Arguments für sie bedeuten dagegen wohl keine grundlegende Revision. [Vgl. Rawls 2001, XVI f., 43-45, 94-130]

5 Eine Theorie der Gerechtigkeit

Grundstruktur der Gesellschaft“ formulieren, verstanden als „die Art, wie die wichtigsten gesellschaftlichen Institutionen Grundrechte und -pflichten und die Früchte der gesellschaftlichen Zusammenarbeit verteilen.“ Ausdrücklich weißt Rawls schon in seinem Hauptwerk auf die Möglichkeit hin, daß „die für die Grundstruktur richtigen Grundsätze“ der Gerechtigkeit für andere Anwendungsfälle eines allgemeineren Gerechtigkeitsbegriffs *keine* Geltung besitzen. [Rawls 1975, 23 f.]

Rawls beschränkt seine Fragestellung noch weiter: Zum einen arbeitet er – was den Kern seiner Theorie angeht – mit der unrealistischen Vorstellung einer „Gesellschaft als geschlossenes System [...], das keine Verbindung mit anderen Gesellschaften hat.“³ Zum anderen entwickelt er eine „ideale Theorie“ „vollständiger Konformität“, die Theorie einer „wohlgeordneten Gesellschaft unter günstigen Umständen“⁴, in der „jeder gerecht handelt und seinen Teil zur Erhaltung der gerechten Institutionen beiträgt.“ Auch wenn Rawls „die Probleme der Theorie der unvollständigen Konformität“ für dringlicher hält, beginnt er mit der idealen Theorie, die seiner Auffassung nach „die einzige Grundlage für eine systematische Behandlung dieser dringlicheren Probleme abgibt.“ Rawls gibt in seinem Werk auch Beispiele dafür, wie man von einer idealen Theorie aus Fragen der nichtidealen Theorie angehen kann. Dabei sind zwei Teile einer solchen nichtidealen Theorie zu unterscheiden: Ein Theorie der unvollständigen Konformität, in der es darum geht, wie faktisch vorliegenden Ungerechtigkeiten begegnet werden sollte, und eine Theorie, die nach der Gerechtigkeit unter weniger günstigen Umständen fragt, die sich aus „natürlichen Beschränkungen“ und „geschichtlichen Zufälligkeiten“ ergeben. [Rawls 1975, 24 f., 277-80]

5.2 Die Methode

Schon in einem sehr frühen Aufsatz aus dem Jahr 1951 [Rawls 1999e] hat Rawls einen Vorschlag zur methodischen Vorgehensweise in der Ethik entwickelt, den er auch seinen Überlegungen in „Eine Theorie der Gerechtigkeit“ zugrunde legt.

Rawls geht von der Vorstellung aus, daß Menschen moralische Fähigkeiten (darunter einen Gerechtigkeitssinn) besitzen, die sie befähigen, moralische Urteile (u.a. über die Gerechtigkeit bzw. Ungerechtigkeit bestimmter Handlungen, Institutionen etc.) zu fällen und solche Urteile zu begründen. Diese moralischen Urteile unterscheiden sich in ihrem Grad der Allgemeinheit, aber auch im Grad der Zuversicht, mit der sie gefällt werden. Eine besondere Bedeutung kommt der Klasse von Urteilen zu, die Rawls „wohlüberlegt“ nennt: Urteile, die mit großer Zuversicht und unter Bedingungen gefällt werden, die „für das Denken und Urteilen überhaupt günstig sind“ und insbesondere auch Faktoren wie Angst oder zu große Rücksicht auf eigene Interessen ausschließen, die gerade für moralische Urteile die Gefahr des Irrtums erhöhen. Diese wohlüberlegten Urteile sind die vorläufigen Fixpunkte der ethischen Reflexion. [Rawls 1975, 37]

³ Wo Rawls sich mit Fragen des Kriegs und des Völkerrechts auseinandersetzt, muß er diese Vorstellung natürlich aufgeben.

⁴ Das sind die Umstände *moderater* Knappheit, s. u. S. 136. Eine Gesellschaft ist wohlgeordnet, wenn sie „durch eine öffentliche Gerechtigkeitskonzeption wirksam geregelt ist“. [Rawls 2003, 25]

f., 65f, 68; Rawls 2001, 30]

Eine moralische Theorie kann nun in einer ersten Annäherung als ein Versuch der Beschreibung und Analyse unserer moralischen Fähigkeiten verstanden werden. Zwei Dinge sind dabei aber zu beachten. Zum einen wird man erwarten können, daß eine adäquate Beschreibung und Analyse unserer moralischen Fähigkeiten mehr verlangen wird, als eine bloß systematischere Auflistung unserer moralischen Alltagsurteile, die wir mit ihrer Hilfe fällen, und der Begründungen, die wir dafür vorlegen. Sie verlangt eine eigene wirkliche (philosophische) Theorie mit eigenen Grundsätzen, Methoden und Verfahren der Argumentation und theoretischen Konstruktion, die in gewissem Grade unabhängig von den Urteilen ist, die sie analysieren soll, wenn auch frei, einsichtige Annahmen und bekannte Tatsachen aus anderen Wissensgebieten zu übernehmen. (Gerade deshalb kann eine solche Theorie helfen, Fragen zu entscheiden, über die wir uns zunächst – vorgängig zu dieser Theorie – unsicher waren.) Zum anderen sind sowohl die moralischen Urteile wie die Theorie der Moral revidierbar: Unter der Bedingung, daß die Grundsätze, Annahmen und philosophischen Argumente der Theorie uns überzeugend erscheinen, sind wir bereit, unsere ethischen Alltagsurteile zu ändern. Das betrifft zunächst die Urteile, die nicht als wohlüberlegt gelten können, gilt dann aber eventuell auch für einzelne wohlüberlegte Urteile, wenn diese mit der Theorie nicht vereinbar sind und die betroffenen Grundsätze dieser Theorie überzeugender erscheinen, als die – wenn auch wohlüberlegten – Alltagsurteile. Umgekehrt kann aber auch das Gewicht einzelner unserer wohlüberlegten Urteile so groß sein, daß sie uns zur Revision der vorläufigen Theorie der Moral zwingen. Dieser Prozeß des Messens der Theorie an wohlüberlegten Urteilen und der wohlüberlegten Urteile an der Theorie kann zu einem vorläufigen Gleichgewicht kommen, das Rawls dann ein Überlegungsgleichgewicht nennt.⁵ [Rawls 1975, 37 f., 65-68, 70; Rawls 2001, 29-31]

Rawls unterscheidet drei Niveaus des Überlegungsgleichgewichts. Ein „enges Überlegungsgleichgewicht“ liegt vor, wenn eine Person den beschriebenen Prozeß in Bezug auf eine Theorie abgeschlossen hat, die seinen ursprünglichen Urteilen nahe lag und deshalb relative wenige Revisionen (der ursprünglichen Urteile oder der Theorie) erforderte. Ein „weites Überlegungsgleichgewicht“ dagegen liegt vor, wenn eine Person in diesem Prozeß, „die wichtigsten politischen Gerechtigkeitskonzeptionen, die in unserer politischen Tradition zu finden sind, in Betracht gezogen und die Kraft der verschiedenen Argumente und Gründe dafür“ abgewogen hat, „wobei auch Anschauungen berücksichtigt wurden, die den Gerechtigkeitsbegriff selbst kritisch beurteilen“. Die Moralphilosophie zielt auf ein solches weites Überlegungsgleichgewicht. Ein „volles Überlegungsgleichgewicht“ liegt schließlich vor, wenn in einer wohlgeordneten Gesellschaft jeder Bürger sich in einem weiten Überlegungsgleichgewicht befindet und alle Bürger dabei dieselbe politische Gerechtigkeitskonzeption bejahen.⁶ [Rawls 1975, 68 f.;

⁵ Sehr zu recht weist Rawls darauf hin, daß dieser Prozeß der wechselseitigen „Anpassung von Grundsätzen und überlegten Urteilen“ auch den Erkenntnisprozeß außerhalb normativer Wissenschaften charakterisieren kann und verweist dazu auf Nelson Goodman [Rawls 1975, 38, Anm. 7]. Große Nähe besteht auch zur Beschreibung des Prozesses des (natur-)wissenschaftlichen Fortschritts in den Werken Quines [Rawls 1975, 628].

⁶ Dieser letztere Begriff beschreibt offensichtlich einen nur *in* der Theorie erreichbaren Zustand.

5 Eine Theorie der Gerechtigkeit

Rawls 2003, 61 f.]

Wie jede Theorie bleibt auch die Theorie der Moral im Überlegungsgleichgewicht offen für Revisionen: Neue Einzelfälle stellen moralische Fragen, die wir „wohlüberlegt“ anders beantworten, als es die Theorie nahelegt. Neue Erkenntnisse und Einsichten aus Wissensbereichen, auf die diese Theorie frei zurückgreift, lassen die Theorie plötzlich fragwürdig erscheinen. Oder es wird eine neue Theorie der Moral vorgeschlagen, die besser den bestehenden wohlüberlegten Urteilen im ursprünglichen Überlegungsgleichgewicht entspricht. [Rawls 1975, 38, 69-72, 148, 629-631; Rawls 2001, 83]

Rawls – ein Meister auch der Beschränkung – verzichtet in „Eine Theorie der Gerechtigkeit“ auf eine Untersuchung der Frage, ob es im Erkenntnisprozeß so etwas wie eine Annäherung an ein eindeutig bestimmtes ideales Überlegungsgleichgewicht gibt, da er eine solche Frage für verfrüht hält. Zunächst lädt er nur seine Leser dazu ein, die von ihm vorgelegte Theorie an den eigenen wohlüberlegten Urteilen zu überprüfen. [Rawls 1975, 69 f.]

5.3 Gerechtigkeit als Fairneß

Als einem Teilgebiet einer Theorie der Moral geht es einer Theorie der Gerechtigkeit um die Analyse des Gerechtigkeitssinns als *einer* unserer moralischen Fähigkeiten. Die These von Rawls ist es nun, daß seine „Theorie der Gerechtigkeit als Fairneß“ von den bekannten Gerechtigkeitseinstellungen am besten dieser Aufgabe gerecht wird. Vor dem Hintergrund seiner allgemeinen Auffassung über den Erkenntnisprozeß in der Ethik bedeutet das dreierlei: Rawls erhebt damit den Anspruch, daß die Theorie der Gerechtigkeit als Fairneß am besten zu den wohlüberlegten moralischen Urteilen im Überlegungsgleichgewicht paßt, d. h. zu den Urteilen, an denen wir festhalten, nachdem wir sie mit verschiedenen Vorschlägen zu einer Theorie der Gerechtigkeit konfrontiert haben. Sie ist im Einklang mit den übrigen Tatsachen und Überzeugungen, insbesondere über den Menschen und die Gesellschaft, die wir etwa anderen Wissenschaften verdanken. Zuletzt ist sie vernünftig in dem Sinn, daß auch gewichtige philosophische Argumente für sie sprechen. [Rawls 1975, 69, 471, 629]

Im Zentrum der Theorie der Gerechtigkeit als Fairneß steht die Idee, daß gerecht das ist, auf was sich vernünftige Menschen frei in einer auf geeignete Weise spezifizierten *fairen* Verhandlungssituation – dem Urzustand – einigen würden. Bezogen auf die Aufgabenstellung, Gerechtigkeitsgrundsätze für die gesellschaftliche Grundstruktur zu finden, läßt sich dann also genauer sagen: Diese „sind diejenigen Grundsätze, die freie und vernünftige Menschen in ihrem eigenen Interesse in einer anfänglichen Situation der Gleichheit zur Bestimmung der Grundverhältnisse ihrer Verbindung annehmen würden.“ [Rawls 1975, 28]

Nicht nur die Verteidigung der Theorie der Gerechtigkeit als Fairneß als solcher, sondern auch die Spezifikation des Urzustands rekurriert auf die oben genannten drei Kriterien. Dabei ist zu beachten, daß im Sinne der Idee des Überlegungsgleichgewichts und des „nonfoundationalist“ Charakters der Vorgehensweise die Übereinstimmung mit wohlüberlegten Urteilen im Überlegungsgleichgewichts nur am Ende der Entwicklung

der Theorie abschließend überprüft werden kann. Das erste Kriterium ist also am Ausgangspunkt der Überlegungen nur provisorisch, vorläufig anzulegen, was den beiden anderen Kriterien für die Begründung der genaueren Ausgestaltung des Urzustands zunächst ein größeres Gewicht gibt. [Vgl. Rawls 1975, 37, 140, 629; Rawls 2001, 31]

Unproblematisch sind zunächst einige Elemente des Urzustands, die diesen als fair charakterisieren. Dazu gehören die Freiheit und Gleichheit (und Gleichberechtigung) der Verhandlungspartner im Urzustand: Es wäre wohl kaum zu begründen, weshalb ein Grundsatz gerecht sein soll, dem einige Verhandlungspartner nur unter Zwang zugestimmt haben, oder der Ergebnis eines Verhandlungsprozesses ist, auf dessen Ausgang einige der Verhandlungspartner deshalb wenig oder keinen Einfluß hatten, weil sie in der Verhandlung weniger Rechte als andere hatten.⁷ [Rawls 1975, 34-37, 176; Rawls 2001, 14 f.; Rawls 1993b, 24 f.]

Größte Berühmtheit unter den Bedingungen des Urzustands hat der „Schleier des Nichtwissens“ gefunden, hinter dem sich die Vertragspartner während ihrer Verhandlungen befinden sollen. Diesen sollen im Urzustand die meisten Einzeltatsachen über sich und ihre Gesellschaft unbekannt sein:

„Vor allem kennt niemand seinen Platz in der Gesellschaft, seine Klasse oder seinen Status; ebensowenig 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.“ [Rawls 1975, 160]

Diese Bedingung dient vor allem dazu, sicherzustellen, daß die beschlossenen Grundsätze nicht von den zufälligen und moralisch willkürlichen Situationen beeinflusst werden, in der sich Menschen tatsächlich befinden. Kein Verhandlungspartner (auch keine Koalition von Verhandlungspartner) kann im Urzustand auf Grundsätze hinwirken, die einfach auf die eigenen Interessen zugeschnitten sind. In einer kantischen Deutung wird auf diese Weise erreicht, daß im Urzustand Grundsätze einer *autonomen* Moral beschlossen werden, die sich freie und gleiche Vernunftwesen geben würden, die nur den allgemeinen menschlichen Lebensbedingungen unterworfen sind. Die Bedingung erleichtert auch wesentlich das Entscheidungsproblem, das der Urzustand vorlegt, da sie sicherstellt, daß alle Beteiligten dieselben Grundsätze wählen werden, und unter der Voraussetzung, daß die Annahmen über allgemeine Tatsachen (s. u.) konstant gehalten werden, auch, daß diese Grundsätze unabhängig von Geschichte und Geographie sein werden. [Rawls 1975, 28, 36, 159, 162-164, 283-285]

Um das Entscheidungsproblem des Urzustands definit zu machen, reicht es aber nicht, festzulegen, was die Parteien des Urzustands *nicht* wissen, es muß auch festgelegt werden, was sie wissen. Rawls nimmt an, daß sie zum einen wissen, daß in ihrer Gesellschaft die „Anwendungsverhältnisse der Gerechtigkeit“ gegeben sind: Die

⁷ Vgl. hierzu auch die Überlegungen zu einer autonomen Moral weiter unten.

5 Eine Theorie der Gerechtigkeit

Menschen befinden sich in einer Situation des kooperativen Konflikts, da sich im Hinblick auf ihre unterschiedlichen Interessen und Ziele zwar aus der gesellschaftlichen Zusammenarbeit für alle Vorteile ergeben können, u.a. aber die *moderate* Knappheit der zur Verfügung stehenden Mittel einen Konflikt darüber entstehen läßt, wessen Ziele in welchem Maße verwirklicht werden. [Rawls 1975, 148-151; Rawls 1999g, 109 f.]

Darüber hinaus nimmt Rawls an, daß den Parteien im Urzustand alle relevanten allgemeinen und allgemein anerkannten Tatsachen über menschliche Gesellschaften bekannt sind, wie sie etwa von Psychologie, Soziologie, Politologie und den Wirtschaftswissenschaften erforscht werden. Wären sie nämlich unbekannt, könnten in Urzustand Grundsätze beschlossen werden, die sich in einer menschlichen Gesellschaft als nicht verwirklichtbar erweisen. Zu den bekannten allgemeinen Tatsachen der Moralphysikologie gehört auch – wenigstens für die Zwecke der idealen Theorie – die Annahme Rawls, daß die Menschen einen formalen Gerechtigkeitssinn besitzen, dem sie auch folgen wollen. (Diesen formalen Gerechtigkeitssinn mit einer inhaltlichen Gerechtigkeitseinstellung zu füllen, ist Aufgabe der Parteien im Urzustand.) [Rawls 1975, 160 f., 168 f.]

Zur Spezifikation des Entscheidungsproblems im Urzustand gehören noch Annahmen darüber, wie und nach welchen Kriterien die Parteien im Urzustand entscheiden. Hier nimmt Rawls zunächst an, daß sie zweckrational und rein selbstinteressiert entscheiden. Nach Maßgabe dessen, was sie von ihren Zielen wissen, werden sie die Grundsätze wählen, die diese Ziele am besten fördern werden. Weder haben die Parteien im Urzustand altruistische Motive, noch kennen sie ein Gefühl des Neides. Sie sind auch sonst frei von irgendwelchen speziellen psychologischen Charaktereigenschaften (wie ein besonderes Herrschafts- bzw. Unterwerfungsbedürfnis oder besondere Risikoaversion bzw. Abneigung). Sie haben keinerlei Interesse am Wohlergehen der anderen Parteien des Urzustands und sind bei ihrer Wahl keinen vorgegebenen moralischen Einschränkungen unterworfen.⁸ [Rawls 1975, 150-152., 166-169, 575, 633 f.; Rawls 2001, 87]

Diese Annahmen ergeben sich im wesentlichen aus dem Zweck der Konstruktion des Urzustands, der „Herleitung aller Pflichten und Verpflichtungen der Gerechtigkeit aus anderen vernünftigen Bedingungen“. Die Imputation moralischer Motive für die Parteien im Urzustand könnte bedeuten, daß das ganze Verfahren zirkulär wird, in jedem Fall würde sie die Basis moralischer Voraussetzungen verbreitern. In dem Maß, in dem die unterstellten moralischen Motive selbst klärungsbedürftig bzw. umstritten sind, würde dies die Überzeugungskraft der Theorie der Gerechtigkeit als Fairneß schwächen. (Moralische Beschränkungen, für die allgemeine Anerkennung vorausge-

⁸Wobei die Parteien aber davon ausgehen, daß sie aufgrund eines „formalen Gerechtigkeitssinns“ ihre Vereinbarung einhalten werden. Außerdem suchen sie eine Konzeption der *Gerechtigkeit* und schließen deshalb Prinzipien aus, die den „formalen Bedingungen für den Begriff des Rechten“ nicht genügen. [Rawls 1975, 152-159]

Es sei hier betont, daß dies zunächst Annahmen über die Parteien im Urzustand sind und so nicht durchgängig für die Menschen in der „Realität“ gelten, für die die Parteien Grundsätze wählen. Es kann sich etwa herausstellen, daß die eigenen Interessen, die eine Partei rein im Eigeninteresse verteidigt hat, die Interessen einer altruistischen Person sind. Vgl. auch unten S. 141.

setzt werden kann, wären in dieser Hinsicht kein Problem.) Spezifische Charaktereigenschaften der Parteien anzunehmen, würde in ähnlicher Weise das Gewicht der Ergebnisse einschränken, insoweit sie nun von möglicherweise umstrittenen Annahmen (etwa der Risikoaversion) abhängen würden. Die beschlossenen Grundsätze würden nur noch für Menschen gelten, deren Freiheit so eingeschränkt wäre, daß sie mit diesen Eigenschaften vereinbar wäre. Der Verzicht auf die Annahme etwa altruistischer Motive ist auch in dem Sinne problemlos, als bei den gegebenen Voraussetzungen über das Wissen der Vertragsparteien ihre Einführung sowieso keinen Unterschied bedeuten würde. Ihre Einführung hätte also nur Sinn zusammen mit der Lockerung der Annahmen über die Beschränktheit des Wissens. Von einer solchen Kombination steht zu befürchten, daß sie das Entscheidungsproblem unlösbar machen würde. [Rawls 1975, 151 f., 164, 168 f., 171-173, 286 f., 575 f., 632-635; Rawls 2001, 87]

Während die Annahme der Zweckrationalität der Individuen im Urzustand eine große Nähe zum Modell des Homo oeconomicus herstellt, markiert eine weitere Voraussetzung einen deutlichen Unterschied zu ökonomischen Modellen rationalen Entscheidens. Rawls nimmt nämlich an, daß die Parteien im Urzustand angesichts ihres sehr begrenzten Wissens (s)einer schwachen Theorie des Guten anhängen. Sie nehmen an, daß es bei aller Verschiedenheit der Lebenspläne und Ziele der Menschen – vielleicht aufgrund der Natur des Menschen – bestimmte Güter gibt („Grundgüter“), die sich bei der Verfolgung aller dieser Ziele als nützlich erweisen oder sogar notwendig dafür sind, und von denen deshalb alle Menschen lieber mehr als weniger haben wollen. Zu diesen Grundgütern rechnet Rawls in „Eine Theorie der Gerechtigkeit“ „Rechte, Freiheiten und Chancen [...] Einkommen und Vermögen“ und das Selbstwertgefühl.⁹ [Rawls 1975, 112 f., 166, 433-435, 472 f.]

Zuletzt bleibt noch zu klären, unter welchen „Gegenständen“ die Akteure des Urzustand wählen können. Aus der Aufgabenstellung einer Theorie der Gerechtigkeit ergibt sich zunächst, daß die Parteien unter möglichen Gerechtigkeitsauffassungen wählen sollen. Sie sollen also die Grundsätze festlegen, nach denen die Gerechtigkeit der Grundstruktur einer Gesellschaft beurteilt werden kann. Es ist dagegen nicht ihre Aufgabe, eine bestimmte (ideale) Gesellschaftsform oder gar einen (idealen) Zustand der Welt zu bestimmen. (Vielmehr soll etwa die Wahl der Gesellschaftsform unter der Berücksichtigung der Gerechtigkeitsprinzipien erfolgen.) [Rawls 1975, 23 f., 27 f., 223-229]

Es erscheint allerdings unmöglich, die Menge aller möglichen Gerechtigkeitsauffassungen überhaupt zu bestimmen und dann für den Zweck der Entscheidung im Urzustand darzustellen. Rawls schlägt deshalb vor, den Parteien im Urzustand eine Liste möglicher (bekannter) Gerechtigkeitsvorstellungen vorzulegen und sie unter diesen – mit der Hilfe paarweiser Vergleiche – wählen zu lassen.¹⁰ Das ist zwar insofern un-

⁹ Ähnlich wie für die genauere Ausgestaltung dessen, was mit Freiheit gemeint ist (s. u. S. 139 Anm. 12), betont der spätere Rawls die Bedeutung der politischen Konzeption der Person als frei und gleich für die Ausgestaltung der Liste der Grundgüter. [Rawls 2001, 57 f., 168-176] Das Grundgut „Selbstwertgefühl“ wird später präziser als die „sozialen Grundlagen der Selbstachtung gefaßt. [Rawls 2003, 101 f., 180]

¹⁰ Die Liste die Rawls für seine Untersuchung zugrunde legt, enthält neben seinen eigenen zwei Grundsätzen zwei Varianten des Utilitarismus, den Perfektionismus, je drei „intuitionistische“ (pluralistische)

befriedigend, da auf diese Weise kein endgültiges Ergebnis über die Gerechtigkeitsvorstellung zu erreichen ist, die aus der Theorie der Gerechtigkeit als Fairneß als die notwendigerweise beste ableitbar ist. Wenn Rawls allerdings in der Überzeugung recht hat, daß es von den Grundideen her sowieso nur eine relative kleine Zahl relevanter („philosophisch interessanter“) Gerechtigkeitsvorstellungen gibt, dann wiegt diese zugegebene Schwäche der Verwendung einer Liste vielleicht nicht allzu schwer. Neue Vorschläge, von denen sich vernünftigerweise behaupten läßt, daß sie wesentliche neue Gesichtspunkte einbringen und deshalb den traditionellen Theorien möglicherweise überlegen sind, sind einfach in die Liste aufzunehmen. Wenn dazu die von Rawls vorgeschlagene Liste in der Tat „die führenden herkömmlichen Theorien enthält“ und diese „den historischen Konsensus darüber darstellen, welche Moralvorstellungen bisher als die vernünftigeren und praktikableren erscheinen“, dann ist mit ihrer Ordnung ohne Zweifel schon viel gewonnen. [Rawls 1975, 144-148, 629-631; Rawls 2001, 83; Rawls 1993b, 254]

5.4 Die zwei Grundsätze der Gerechtigkeit

Rawls argumentiert, daß in einem solchermaßen charakterisierten Urzustand die Parteien sich für folgenden zwei Grundsätze der Gerechtigkeit entscheiden würden¹¹:

a. Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.

b. Social and economic inequalities are to satisfy two conditions. First, they must be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they must be to the greatest benefit of the least advantaged members of society. [Rawls 2001, 42]

Einige Hinweise zur Interpretation:

- Die Ordnung, in der die zwei Grundsätze und die zwei Bedingungen für Ungleichheiten im zweiten Grundsatz angeführt werden, entspricht ihrer Rangordnung für die Beurteilung der Grundstruktur einer Gesellschaft: Der erste Grundsatz ist dem zweiten, und die erste Bedingung der zweiten lexikalisch vorgeordnet. [Rawls 2001, 43; Rawls 1975, 82 f.]
- Was den ersten Grundsatz angeht, vertritt Rawls nicht den Standpunkt, daß es möglich wäre, allein aus der Analyse eines emphatischen Begriffs von Freiheit oder auch allein aus der Vorstellung des Urzustands abzuleiten, wie ein „fully adequate scheme of equal basic liberties“ aussehen könnte. Statt dessen bildet eine Liste solcher Freiheiten Teil des den Parteien vorgelegten Vorschlags. Diese Liste ist – so Rawls – nicht einfach aus der politischen wie philosophischen

und egoistische Auffassungen und zuletzt drei (wohldefinierte) Mischformen.

¹¹ Ich zitiere hier die revidierte Fassung, wie sie sich in Rawls [2001] findet, und folge, soweit es um den ersten Grundsatz geht, auch der Darstellung dort. Rawls selbst hält die Revision des ersten Grundsatzes und der Deutung seiner Priorität für die wichtigste Änderung in seiner Theorie der Gerechtigkeit [Rawls 2001, 42, Anm. 2]. Die Neufassung und ihre Gründe werden dargestellt in Rawls [2001, 42-47] und Rawls [1993a].

5.4 Die zwei Grundsätze der Gerechtigkeit

Tradition entnommen, sondern so gestaltet, daß das sich aus ihr ergebende System der Freiheiten, wirkungsvoll zentrale Fähigkeiten freier und gleicher Personen schützt.¹² Die von Rawls konkret verwendete Liste ist s. E. hinreichend, um die Überlegenheit der zwei Grundsätze gegenüber den anderen diskutierten Gerechtigkeitseinstellungen aufzuweisen. Alternative Listen könnten theoretisch mit Hilfe des Verfahrens der Urzustandes mit der Liste Rawls verglichen werden. [Rawls 2001, 44 f., 104, 112 f.; Rawls 1993a, 292 f., 331-334]

- Die erste Bedingung des zweiten Grundsatzes verlangt, daß über die formale Chancengleichheit hinaus, die erfüllt ist, wenn die Laufbahnen, die mit einem höheren Anteil an sozialen und ökonomischen Gütern verknüpft sind, allen Befähigten und Willigen offenstehen, auch die faire Chancengleichheit gegeben ist. D.h. alle Mitglieder einer Gesellschaft, soweit sie ähnliche natürliche Begabungen besitzen und ähnliche Bereitschaft zeigen, sie einzusetzen, haben auch – unabhängig von ihrer sozialen Schicht – ähnliche Chancen, die Positionen zu erreichen, die mit einem höheren Anteil an sozialen und ökonomischen Gütern verbunden sind. [Rawls 1975, 86-96, 105]
- Der Maßstab für die Erfüllung der zweiten Bedingung des zweiten Grundsatzes (des Unterschiedsprinzips) ist ein Index von sozialen Grundgütern. Nachdem die Freiheitsrechte und Zugangschancen in einer wohlgeordneten Gesellschaft *vorrangig* allen Mitgliedern gleich gewährt werden, gehen von der Liste der Grundgüter nur die mit der beruflichen Stellung verbundenen Befugnisse und Vorrechte, Einkommen, Vermögen und die soziale Basis der Selbstachtung in diesen Index ein, wobei zur Vereinfachung die soziale Basis der Selbstachtung zunächst außer acht gelassen wird. Die Bedingung ist nicht auf einzelne Individuen anzuwenden, sondern auf „repräsentative Personen“, die für die verschiedenen sozialen Positionen und beruflichen Stellungen stehen, die im Rahmen der Grundstruktur einer Gesellschaft errichtet worden sind. Rawls geht davon aus, daß es möglich ist, solchen sozialen Positionen – und damit den sie repräsentierenden Personen – zu erwartende Aussichten auf die Erfüllung von Lebensplänen zuzuschreiben (ausgedrückt in ihrem Zugang zu Grundgütern, wobei über die Annahme einer positiven Korrelation zwischen Befugnissen und Vorrechten, im wesentlichen nur noch Einkommen und Vermögen berücksichtigt werden).¹³ Das Unterschiedsprinzip ist also *keine* Regel darüber, wie irgendwelche Güter auf Einzelpersonen zu verteilen sind, sondern darüber wie Institutionen auszugestaltet sind, und hier genauer, welche sozialen Vorteile (insgesamt, d. h. im Blick auf das

¹² Genauer sichern sie „die adäquate Entwicklung und den vollen Einsatz der beiden moralischen Vermögen freier und gleicher Personen“, d. h. der Fähigkeiten zur Entwicklung und Anwendung eines Gerechtigkeitssinns und einer Konzeption des Guten. Dazu und zu den für den späten Rawls wichtigen *politischen* Konzeptionen freier und gleicher Personen und ihrer moralischen Vermögen s. Rawls [1993b, 18-22, 29-35, 293, 299-304, 310 -324] und Rawls [2001, 18-24, 44 f., 112-114].

¹³ Das sind also die „für die Beurteilung sozialer und wirtschaftlicher Ungleichheiten“ relevanten gesellschaftlichen Positionen. Für Fragen, die die Grundfreiheiten betreffen, ist dagegen die Position gleicher Bürgerrechte relevant.

ganze System von Institutionen) mit den im Rahmen dieser Institutionen errichteten Positionen verbunden sein sollen. [Rawls 1975, 85, 114-121]

- Ein oft übersehener Aspekt der zwei Grundsätze der Gerechtigkeit Rawls ist, daß es sich dabei – im Gegensatz zum Utilitarismus – um eine Gerechtigkeitsauffassung handelt, die „erschöpfbar“ ist, d. h. wenn einmal ein gerechter Zustand erreicht ist, ist dieser zwar zu bewahren, aber es besteht keine Verpflichtung zu einer ständigen weiteren Verbesserung des Zustands. Das ergibt sich bezogen auf den ersten Grundsatz und die erste Bedingung des zweiten schon aus der Natur der entsprechenden Forderungen, da diese keine Skala angeben, auf der eine fortlaufende Verbesserung möglich wäre: Ist ein völlig adäquates System gleicher Freiheiten – vorgestellt in der Form einer Liste – einmal eingerichtet, bleibt in Bezug auf die gleichen Grundfreiheiten nichts mehr zu verwirklichen. Die Chancengleichheit verlangt Gleichheit der Chancen und nicht Maximierung (was immer das bedeuten sollte) von Chancen. In Bezug auf die zweite Bedingung des zweiten Grundsatzes ist es sogar möglich, daß eine Gesellschaft gerecht genannt werden kann, selbst wenn institutionelle Umgestaltungen möglich wären, die allen Mitgliedern dieser Gesellschaft einen Zuwachs in Bezug auf den verwendeten Index der Grundgüter geben würden: Der Grundsatz verlangt (bei Rawls) nicht die Maximierung des Zugangs zu sozialen Grundgütern für die am schlechtesten gestellte Gruppe, sondern gibt nur an, bis zu welchem Grad Ungleichheiten im Zugang zu sozialen Grundgütern zulässig sind, nämlich insoweit als die Ungleichheit zum Wohl der am schlechtesten Gestellten beiträgt.¹⁴ Eine Gesellschaft könnte sich dementsprechend für ein höheres Maß an Gleichheit entscheiden, auch wenn das mit schlechteren Aussichten für alle (und damit auch für die am wenigsten bevorzugte Gruppe) verbunden wäre. [Rawls 1975, 99 f.; Rawls 2001, 63 f., 159 f.]
- Ein letzter, oft übersehener Aspekt der Gerechtigkeitsauffassung Rawls' ist, daß diese – worauf Rawls großen Wert legt – wichtige Elemente einer reinen Verfahrensgerechtigkeit enthält.¹⁵ Wenn der grundlegende institutionelle Aufbau einer Gesellschaft gleiche Grundfreiheiten und faire Chancengleichheit für alle sichert, und der Zugang zu sozialen Grundgütern für die am schlechtesten gestellte gesellschaftliche Gruppe nicht durch die Verschlechterung des Zugangs zu diesen Gütern für eine besser gestellte Gruppe verbessert werden kann, dann ist jede

¹⁴ In Rawls [1975, 99] nennt Rawls einen Zustand, in dem ein entsprechendes Maximum erreicht wird, „vollkommen“ gerecht und betrachtet ihn „als die beste gerechte Möglichkeit“. Eine Gesellschaft, die das Unterschiedsprinzip erfüllt, ohne eine solche Maximum zu verwirklichen ist aber immer noch „durchweg gerecht“ und mehr wird von seiner Theorie der Gerechtigkeit nicht gefordert.

¹⁵ Reine Verfahrensgerechtigkeit liegt vor, wenn der alleinige Maßstab für die Gerechtigkeit des Ergebnisses eines Verfahrens die Gerechtigkeit des Verfahrens ist. Vollkommene Verfahrensgerechtigkeit liegt vor, wenn es einen vom Verfahren unabhängigen Maßstab für die Gerechtigkeit des Ergebnisses gibt, und das Verfahren ein nach diesem Maßstab gerechtes Verfahren garantiert. Unvollkommene Verfahrensgerechtigkeit schließlich liegt vor, wenn es einen vom Verfahren unabhängigen Maßstab gibt, und das Verfahren zwar auf ein gerechtes Ergebnis zielt, aber es nicht garantieren kann. [Rawls 1975, 106-108]

5.4 Die zwei Grundsätze der Gerechtigkeit

persönliche und strukturelle Verteilung (an sozialen Grundgütern), die sich aus dem normalen Funktionieren der gesellschaftlichen Institutionen ergibt, auch gerecht.¹⁶ Natürlich enthält sie auch Elemente der unvollkommenen Verfahrensgerechtigkeit: Der Aufbau ihrer Institutionen soll ja etwa gleiche Freiheiten für alle und gleiche Chancen für gleichbegabte und -motivierte Mitglieder sichern und es ist wohl kein institutioneller Entwurf möglich, der diese Gleichheit im Sinne einer vollkommenen Verfahrensgerechtigkeit sichern kann. Entsprechendes gilt für die Sicherung des Zugangs der am schlechtesten gestellten gesellschaftlichen Gruppe zu den sozialen und ökonomischen Grundgütern. Verfahren und Institutionen, die zum diesen Zweck in einer Gesellschaft eingerichtet werden, können – jedenfalls außerhalb der idealen Theorie – ihr Ziel auch verfehlen. [Rawls 1975, 105-110, 395-397; Rawls 1999b, 262 f.]

Die Argumentation für die hier kurz dargestellten Grundsätze erfolgt (schematisiert¹⁷) in drei Schritten¹⁸: Rawls versucht erst zu zeigen, daß die Parteien im Urzustand die zwei Grundsätze wählen würden, wenn nicht nur sie selbst als frei von Neid und anderen speziellen psychologischen Eigenheiten gedacht werden, sondern wenn sie dies auch von den Personen annehmen, die sie vertreten. Im zweiten Schritt geht es darum, zu überprüfen, ob eine Gesellschaft, die diesen Grundsätzen genügt, auch stabil ist, d. h. die Haltungen und Überzeugungen erhält, erzeugt oder verstärkt, die sie zu ihrer Erhaltung braucht. Dafür wird – in einem Teilschritt – (als Beispiel) die Voraussetzung der Neidfreiheit aufgehoben, und untersucht, ob in einer Gesellschaft, die entsprechend der zwei Gerechtigkeitsgrundsätze geordnet ist, Neid in einem Maß produziert würde, der die Gerechtigkeit der Gesellschaft gefährden könnte. Außerdem wird überprüft, ob die vorgeschlagene Gerechtigkeitskonzeption in einer auf ihrer Grundlage gestalteten Gesellschaft ausreichend Unterstützung finden könnte. Dies ist nach Rawls der Fall, da (so in Rawls [1973]) in ihr Gerechtigkeitsinn und unsere Vorstellung vom Guten einander ergänzen oder da (so in Rawls [2001]) Gerechtigkeit als Fairness in einer von einem vernünftigen Pluralismus geprägten Gesellschaft von einem „überlappendem Konsens“ getragen sein kann. In einem dritten Schritt schließlich müssen wir die Gerechtigkeitsgrundsätze und ihre Konsequenzen mit unseren bestehenden (hoffentlich wohlüberlegten) moralischen Überzeugungen konfrontieren, um sicherzustellen, daß sie zusammen Teil eines Überlegungsgleichgewicht bilden können. [Rawls 1999g, 124 f., 441, 464-474, 496-505, 507 f.; Rawls 2001, 88 f., 132, 136, 180-202]

Der erste Schritt erfolgt – wie schon ausgeführt – in der Form eines paarweisen Vergleichs einer Liste von vorgeschlagenen Gerechtigkeitskonzeptionen.¹⁹ Jedenfalls für

¹⁶ Die Mitglieder einer Gesellschaft haben dann auch Anspruch auf die Erfüllung der innerhalb dieser Gesellschaft erworbenen „gerechtfertigten Erwartungen“ (z. B. auf die Auszahlung der Löhne oder auch von Dividenden in einer gerechten Gesellschaft mit privatem Eigentum an Produktionsmitteln). [Rawls 1975, 109, 344-350]

¹⁷ Die tatsächliche Reihenfolge im Text entspricht nicht strikt diesem Schema.

¹⁸ So in Rawls [1999g, 507 f.]. An anderen Stellen spricht Rawls nur von zwei Teilen der Argumentation, da er dort den dritten Schritt nicht als solchen zählt. Demzufolge ist auch die – sachlich unerhebliche – Numerierung der Schritte uneinheitlich.

¹⁹ In der Darstellung dieses Schritts folge ich eher der Präsentation in Rawls [2001], die meines Erachtens

den in der Sicht Rawls' entscheidenden Vergleich seiner zwei Prinzipien mit dem Utilitarismus²⁰ bedient sich Rawls dazu der Maximinregel der Entscheidung bei Unsicherheit.²¹ Er zeigt, daß wenigstens im Vergleich seiner zwei Prinzipien mit dem Utilitarismus die Bedingungen für die Anwendung der Maximinregel gegeben sind, und daß die Anwendung der Regel zur Bevorzugung der zwei Prinzipien führt. [Rawls 1999g, 105-109, 131-135, 153-160; Rawls 2001, 83, 94-104, 115-119]

Drei Bedingungen führt Rawls an: 1. Die Parteien haben keine verlässliche Grundlage für Wahrscheinlichkeitsschätzungen. 2. Ein (ziemlich oder vollständig) befriedigendes Ergebnis ist sicherbar. 3. Das schlechtestes Ergebnis der anderen Alternativen ist deutlich schlechter als als das sicher erreichbare oder ist überhaupt unerträglich. [Rawls 1999g, 134; Rawls 2001, 98]

Die Bedingungen müssen nicht vollständig erfüllt sein, insbesondere würde es reichen, daß die dritte Bedingung vollständig erfüllt ist (Ergebnisse der betrachteten Alternativen sind unerträglich) und die beiden anderen in einem von Rawls definierten Sinn nur teilweise: Die Parteien können sich bei ihren Wahrscheinlichkeitseinschätzungen jedenfalls auf keine gut etablierten Tatsachen über die Welt stützen, und das sicherbare Ergebnis ist zumindest ziemlich zufriedenstellend. Für den Fall des Urzustands sind aber – nach Rawls – die zweite und dritte Bedingung beide im hohen Maße erfüllt.²² [Rawls 2001, 99]

Das Argument besteht also im wesentlichen darin, zu zeigen, daß die betrachteten Alternativen (der Utilitarismus, der Perfektionismus) zu unerträglichen Ergebnissen führen können (das ist nach Rawls der Fall, da sie eine Einschränkung der Grundfreiheiten verlangen können) und daß das sicherbare Ergebnis der Konzeption der zwei Prinzipien zufriedenstellend ist (das ist nach Rawls der Fall, da die zwei Prinzipien eine in hohem Masse zufriedenstellende soziale Welt begründen). Das Argument wird bestätigt durch Überlegungen zu den „strains of commitment“: Die Parteien können guten Glaubens keine Vereinbarung treffen, von der sie wissen, daß sie sie nicht ein-

klarer strukturiert ist als die in Rawls [1973]. Ich beschränke mich auf die Grundlinien der Argumentation.

²⁰ So in Rawls [2001]. In seinem Hauptwerk scheint Rawls der Maximinregel der Entscheidung eine grundsätzlichere Rolle zu spielen: Sie zeichnet die zwei Prinzipien gegenüber allen anderen Konzeptionen der Liste aus. Später schließt das Rawls für eine der Mischkonzeptionen explizit aus und geht auf die übrigen Elemente der Liste nicht mehr ein. Vgl. Rawls [1999g, 132-136, 153] und Rawls [2001, 43, 94-104].

²¹ Sowohl in Rawls [1999g, 153] als auch in Rawls [2001, 99] spricht Rawls von der Maximinregel als einer Heuristik, die mehr der Darstellung dient und nicht notwendigerweise gebraucht wird. Gerade aber die Darstellung in Rawls [2001] scheint mir deutlich zu machen, daß die inhaltliche Verknüpfung der Anwendungsbedingungen der Regel [Rawls 1999g, 134; Rawls 2001, 97 f.] mit der Entscheidungssituation im Urzustand [Rawls 1999g, 134 f.; Rawls 2001, 99-104, 115-119] im wesentlichen das Argument jedenfalls für die zwei Prinzipien versus Utilitarismus zusammenfaßt. Dieser Interpretation folgt die Darstellung im Text.

²² Für die erste Bedingung nimmt Rawls in Rawls [2001, 99, 101 f.] in der Tat nur noch an, daß sie teilweise erfüllt ist. In Rawls [1999g] hatte die Bedingung noch eine größere Rolle gespielt. Die Änderung verdankt sich vor allem der Tatsache, daß Rawls berücksichtigt, daß seine Position zur Wahrscheinlichkeitstheorie angesichts subjektivistischer und bayesianischer Interpretationen umstritten ist, weshalb er sich so wenig als möglich darauf stützen möchte. Meine eigene Konstruktion im folgenden Kapitel hängt in keiner Weise von dieser Bedingung ab.

halten können, da sie sie überfordert. [Rawls 1999g, 135, 153 f., 287-290; Rawls 2001, 99-104, 115-119]

Das Argument für die zwei Prinzipien gegenüber Mischkonzeptionen, die das erste Prinzip beibehalten, aber das Differenzprinzip für eine Alternative – etwa eine utilitaristische Verteilungsregel – eintauschen, ist in Rawls' Augen weniger eindeutig. Eine Reihe von Argumenten zusammengenommen lassen aber doch die Waagschale zugunsten des Differenzprinzips ausschlagen: Insbesondere ist dieses besser mit der Bedingung der Öffentlichkeit, der Idee der Reziprozität und der Anforderung der Stabilität vereinbar. Das utilitaristische Verteilungsprinzip dagegen ist unterbestimmt, da uns ein öffentliches Maß für interpersonelle Nutzenvergleiche fehlt; es verlangt von den Schlechtergestellten zu viel, insbesondere überfordert es ihre Sympathie für Bessergestellte; das in der Mischkonzeption notwendige soziale Minimum²³ ist nicht auf befriedigende Weise zu bestimmen. [Rawls 2001, 119-130]

5.5 Von den Grundsätzen zu den Institutionen der Grundstruktur

Die explizite Aufgabe einer Theorie der Gerechtigkeit, wie Rawls sie definiert, ist *theorieintern* mit der Angabe und der Begründung der Gerechtigkeitsgrundsätze und des Aufweises der Stabilität einer entsprechend gestalteten Gesellschaft erfüllt. Doch verlangt die Methode des Überlegungsgleichgewichts noch einen zusätzlichen wichtigen Schritt. Die Gerechtigkeitsgrundsätze sollen ja daraufhin überprüft werden, ob sie mit unseren wohlüberlegten Einzelurteilen und der Begründung in einem Überlegungsgleichgewicht übereinstimmen. Diese beziehen sich auf die Beurteilung der Gerechtigkeit von Gesellschaften und ihrer Grundstrukturen. Deshalb ist es notwendig, Vorstellungen darüber zu entwickeln, wie denn nun eine Gesellschaft (in ihren Grundinstitutionen) aussehen würde, die den Grundsätzen der Gerechtigkeit entsprechen würde, um dann zu fragen, ob wir eine solche Gesellschaft in der Tat als gerecht beurteilen würden (und Gesellschaften, die in wesentlichen Punkten den Prinzipien der Gerechtigkeit nicht genügen, als ungerecht). [Rawls 1999g, 132; Rawls [2001, 136]]

Ein weiterer Grund für die Beschäftigung mit der institutionellen Ausgestaltung einer gerechten Gesellschaft ist der „Wunsch nach Verdeutlichung der Unterscheidung zwischen einer Demokratie mit Eigentumsbesitz, in der alle von beiden Gerechtigkeitsprinzipien artikulierten politischen Hauptwerte umgesetzt werden, und einem kapitalistischen Wohlfahrtsstaat“. [Rawls 2003, 211]

Diesen Anliegen widmet Rawls den zweiten Teil seines Hauptwerkes, den ich im folgenden allerdings nur insoweit kurz und zusammenfassend referieren werde, als Fragen behandelt werden, die im Blick auf eine normative Ökonomie von Bedeutung sind, und die ich nicht in einem der folgenden Kapitel noch ausführlicher behandeln werde.²⁴

²³ Ohne ein solches würden die Parteien nach Rawls der Mischkonzeption nie zustimmen.

²⁴ Das gilt insbesondere für die Frage der gerechten Sparrate und Fragen der internationalen Verteilungsgerechtigkeit.

5 Eine Theorie der Gerechtigkeit

Zunächst beschreibt er einen idealtypischen Weg, auf dem man von den Grundsätzen der Gerechtigkeit zu einem vollausgestalteten institutionellen System einer Gesellschaft gelangt. Dieser Weg erfolgt in vier Schritten: Beim ersten Schritt handelt es sich um den Urzustand. Beim zweiten und dritten Schritt handelt es sich um die verfassungsgebende Versammlung und das Stadium der Gesetzgebung. Rawls stellt sich dabei vor, daß zwischen der Bestimmung einer gerechten Verfassung und der Bestimmung gerechter Gesetze hin und her gewechselt wird, da das wesentliche Kriterium für eine gerechte Verfassung die Frage ist, ob sie zu (im Sinne der Gerechtigkeitsgrundsätze, die hier als „unabhängiger Maßstab“ vorausgesetzt werden) gerechten Gesetzen führt. Im Unterschied zum Urzustand kennen die Mitglieder der verfassungsgebenden Versammlung und der Gesetzgeber „die wesentlichen allgemeinen Tatsachen über ihre Gesellschaft, nämlich die natürlichen Bedingungen und Hilfsquellen, den wirtschaftlichen und politischen Entwicklungsstand“ wie auch die „Ansichten und Interessen, die die Menschen in dem System haben werden“, und die „politischen Taktiken, die sie unter den gegebenen Umständen vernünftig finden werden“. Rawls stellt sich dabei vor, daß dabei die Forderungen und Konsequenzen des ersten Prinzips der Gerechtigkeit im wesentlichen durch die Verfassung abgedeckt werden, während der zweite Grundsatz „bei der Gesetzgebung zum Zuge“ kommt. Die letzte Stufe beschreibt Rawls dann in der folgenden Weise: „Die letzte Stufe ist die Anwendung von Regeln auf Einzelfälle durch die Verwaltung und die Justiz und die Befolgung von Regeln durch die Bürger im allgemeinen. Auf dieser Stufe kennt jeder alle Tatsachen.“ [Rawls 1975, 223-229; Rawls 2001, 48 f.]

Zwei Bemerkungen sind zum Verständnis des Vier-Stufen-Gang noch wichtig: Dieser gesamte Ablauf ist *Teil der Theorie* und soll nicht tatsächliche Vorgänge in verfassungsgebenden Versammlungen, Parlamenten, Gerichten usw. beschreiben. Außerdem geht Rawls nicht davon aus, daß in dieser Vorgehensweise auf alle Fragen eindeutige Antworten gefunden werden können. Alle Regelungen und Institutionen sind als gerecht anzusehen, die in diesem Ablauf hätten beschlossen werden können. Allerdings erhofft sich Rawls von seiner Theorie der Gerechtigkeit doch, daß „sie den Bereich des Gerechten in besserer Übereinstimmung mit unseren wohlüberlegten Urteilen abgrenzt als andere Theorien“ und „die gewichtigeren Übelstände, die eine Gesellschaft vermeiden sollte, schärfer herausstellt“. [Rawls 1975, 229]

Die gerade angesprochene „Unbestimmtheit der Gerechtigkeitstheorie“ [Rawls 1975, 229] zeigt sich z. B. in der Rawlsschen Behandlung der Frage des grundlegenden Wirtschaftssystems. Rawls geht davon aus, daß sowohl eine „Demokratie mit Privateigentum“ (anders als ein Laissez-faire-Kapitalismus oder ein kapitalistischer Wohlfahrtsstaat) als auch ein liberaler, demokratischer Sozialismus (anders als ein Staatssozialismus) gerechte Systeme sein können. Den entscheidenden Unterschied zwischen diesen beiden Systemen sieht Rawls darin, daß in einer Demokratie mit Privateigentum Produktionsmittel eben Privateigentum sind, während sie in einem liberalen Sozialismus öffentliches Eigentum sind. Als Konsequenz ist die Verteilungsfunktion von Preisen im Sozialismus weit weniger wichtig als in einer Demokratie mit Privateigentum, da abgesehen von Arbeit den Preisen von Produktionsmitteln keine Einkommen von Privatpersonen entsprechen. Außerdem werden in sozialistischen Systemen „Vorstellungen

der Planer oder kollektive Entscheidungen oft größeren Einfluß auf die Ausrichtung der Produktion haben.“ [Rawls 1975, 299, 304, 307 f.; Rawls 2001, 136-140]

Viele andere Aspekte aber hätten die beiden Systeme gemeinsam. Dies gilt auch insbesondere für den großen Raum, dem der Markt in seiner allokativen Funktion in beiden Systemen eingeräumt würde. Beide Systeme erstellen außerdem einen „konstitutionellen Rahmen für demokratische Politik, beide garantieren die Grundfreiheiten mit dem fairen Wert der politischen Freiheiten und mit fairer Chancengleichheit, und sie regulieren die ökonomischen und sozialen Ungleichheiten mit Hilfe eines Prinzips der Gegenseitigkeit, wenn nicht mit dem Differenzprinzip.“ [Rawls 2003, 215]

Des weiteren gibt es noch Fragestellungen, für die mit der Wahl des Wirtschaftssystems noch keine Vorentscheidung gefallen ist. Dazu gehört das Ausmaß, in dem öffentliche Güter zur Verfügung gestellt werden, wobei es für Rawls aber außer Frage steht, daß zur „Aufrechterhaltung gerechter Rahmeninstitutionen“ ein gerechter Staat in einem gewissen Ausmaß öffentliche Güter zur Verfügung stellen muß, deren Bereitstellung er auch notwendigerweise mit Zwangsmitteln sichern wird. [Rawls 1975, 299-304, 318]

Für den Fall einer Demokratie mit Privateigentum skizziert Rawls ein solches System von Rahmeninstitutionen, die die Verteilungsgerechtigkeit sicherstellen sollen. Zentral ist dabei, daß Rawls dabei den Begriff der „reinen Verfahrensgerechtigkeit“ angewendet sehen möchte. Zwar wird das vorgeschlagene System zunächst danach beurteilt, ob es den beiden Gerechtigkeitsgrundsätzen entspricht, aber wenn das der Fall ist, und das System tatsächlich auf unparteiische Weise zur Anwendung kommt, dann ist die sich ergebende Verteilung gerecht, unabhängig davon, wie sie sich zu den bekannten Wünschen und Bedürfnissen bestimmter Menschen verhält. Es gibt – in der Rawlschen Theorie der Gerechtigkeit – kein unabhängiges Kriterium, nach der man die sich ergebende bestimmte Verteilung beurteilen könnte. [Rawls 1975, 105-110, 308 f.]

Die Rahmeninstitutionen der *Verteilungsgerechtigkeit* für sich genommen, können die Gerechtigkeit des gesamten Systems – das ja beiden Prinzipien der Gerechtigkeit genügen muß – nicht sicherstellen. Deshalb faßt Rawls zunächst die weiteren Bedingungen einer gerechten Gesellschaft kurz zusammen:

First of all I assume that the basic structure is regulated by a just constitution that secures the liberties of equal citizenship [...]. Liberty of conscience and freedom of thought are taken for granted, and the fair value of political liberty is maintained. The political process is conducted, as far as circumstances permit, as a just procedure for choosing between governments and for enacting just legislation. I assume that there is fair (as opposed to formal) equality of opportunity. This means that in addition to maintaining the usual kinds of overhead capital, the government tries to insure equal chances of education and culture for persons similarly endowed and motivated either by subsidizing private schools or by establishing a public school system. It also enforces and underwrites equality of opportunity in economic activities and in the free choice of occupation. This is achieved by policing the conduct of firms and private associations and by preventing the

establishment of monopolistic restrictions and barriers to the more desirable positions. Finally, the government guarantees a social minimum either by family allowances and special payments for sickness and employment, or more systematically by such devices as a graded income supplement (a so-called negative income tax). [Rawls 1999g, 243]

In der einen oder anderen Weise haben alle diese vorausgesetzten Elemente einer gerechten Gesellschaft auch wirtschaftliche Konsequenzen. So kann der *gleiche Wert* der politischen Freiheiten nur gesichert werden, wenn Eigentum und Reichtum nicht zu ungleich verteilt sind, weshalb staatliche Eingriffe in die Vermögensverteilung notwendig werden. [Rawls 1999g, 198 f.] Aus verschiedenen dieser Elemente ergibt sich die Notwendigkeit der staatlichen (Teil-)Finanzierung bestimmter Leistungen, wie etwa eines Bildungssystems oder von Foren der politischen Diskussion, und einer staatlichen Parteienfinanzierung²⁵, und damit auch die Notwendigkeit von Steuern. Die Sicherung der wirtschaftlichen Chancengleichheit verlangt die Regelung und Kontrolle des Verhaltens privater wirtschaftlicher Akteure. Zuletzt verlangt die Sicherung eines Existenzminimums eine staatliche Politik der Einkommensverteilung.

Für die Erfüllung der verschiedenen wirtschaftlichen und sozialen Aufgaben des Staates sieht Rawls – idealtypisch – vier Regierungsabteilungen vor, die sich – gegliedert in verschiedene Behörden – den aus den Forderungen der Gerechtigkeit ergebenden Aufgaben widmen, ergänzt durch eine weitere Abteilung, die Maßnahmen ermöglichen soll, die sich nach dem Prinzip des „allseitigen Vorteils“ als sinnvoll erweisen.

1. Die Allokationsabteilung soll Marktmacht verhindern und die Bedingungen der Konkurrenz auf Märkten sicherstellen. Außerdem soll sie durch Steuern, Subventionen und Änderungen im Eigentumsrecht dafür sorgen, daß Preise korrekt den gesellschaftlichen Nutzen und Aufwand erfassen.
2. Die Stabilitätsabteilung soll – u.a. durch die Förderung einer wirksamen Nachfrage – Vollbeschäftigung sichern und damit die freie Berufswahl ermöglichen.
3. Die Umverteilungsabteilung dient der Sicherung des Existenzminimums. Sie bedient sich dabei – wie schon zitiert – einer negativen Einkommenssteuer oder einer Kombination von Familienbeihilfen und Zuschüssen bei Krankheit und Arbeitslosigkeit. Maßstab des Existenzminimums ist das Differenzprinzip. Solange dieses erfüllt ist, können ansonsten die Einkommen auf (durch die Allokationsabteilung korrigierten) Märkten bestimmt werden.
4. Die Verteilungsabteilung hat eine doppelte Aufgabe. Zum einen soll sie einen Grad der Ungleichheit der Vermögensverteilung verhindern, der faktisch mit dem gleichen Wert der politischen Freiheiten und der fairen Chancengleichheit unvereinbar wäre. Instrumente dafür könnten u.a. auf Seiten der Empfänger progressive Schenkungs- und Erbschaftssteuern sein. Orientierungspunkt ist wieder

²⁵ Die letzteren dienen wieder der Sicherung des gleichen Wertes der politischen Freiheiten. [Rawls 1999g, 198 f.]

das Differenzprinzip, jetzt bezogen auf das Vermögen. Die zweite Aufgabe ist die „Aufbringung der Mittel, die die Gerechtigkeit fordert“. Rawls schlägt dazu im Rahmen einer idealen Theorie eine proportionale Verbrauchssteuer vor, hält aber im Rahmen der nichtidealen Theorie auch eine progressive Verbrauchssteuer für möglich, um Ungerechtigkeiten zu korrigieren, die durch tatsächlich vorhandene Institutionen verursacht werden.

5. Auch wenn das eben beschriebene System, wie Rawls annimmt, in der Tat zu einer gerechten Einkommens und Vermögensverteilung führen würde, folgt aus der Logik öffentlicher Güter, daß durch die öffentliche Zurverfügungstellung solcher Güter Paretoverbesserungen möglich sein können. Die Austauschabteilung hat die Aufgabe, solche Güter zur Verfügung zu stellen, wenn für ihre Erstellung eine Form der Finanzierung gefunden werden kann, sodaß das Gesamtpaket dem Wicksellschen Einstimmigkeitskriterium genügt.²⁶

²⁶ Das Einstimmigkeitskriterium ist nur gerecht, soweit eine gerechte Einkommens- und Vermögensverteilung vorausgesetzt werden kann. Die Regelungen dieser Abteilung unterliegen keinen Informationsbeschränkungen, die bauen vielmehr auf den konkreten Bedürfnissen der Mitglieder der Gesellschaft auf. [Rawls 1975, 317 f.]

6 Der Urzustand: Ein Rekonstruktion

6.1 Das Problem

Die Behauptung, daß rationale Parteien in einem passend ausgestalteten Urzustand sich aus Eigeninteresse für seine zwei Prinzipien der Gerechtigkeit entscheiden würden, nimmt in Rawls Argumentationsgang eine zentrale Stelle ein. In seinen vielen Kurzdarstellungen seiner Position wird diese Behauptung immer auch an zentraler Stelle anführt, und hat er sie mehrmals, zuerst in seinem Hauptwerk, ausführlicher begründet.¹

Einer weit verbreiteten – von Rawls aber nur bedingt akzeptierten² – Vorstellung entsprechend läuft die Begründung Rawls auf die mehr oder minder direkte Anwendung der Maximin-Entscheidungsregel auf die Situation des Urzustands hinaus. Rawls hätte in dieser Sicht zwei Dinge zu zeigen: 1. In der Situation des Urzustands ist es rational, die Maximinregel zu verwenden. 2. Die Anwendung der Maximinregel führt zur Wahl der zwei Prinzipien.

Dieser Vorstellung entspricht dann auch eine der wichtigsten Kritiklinien gegen Rawls: Vor allem Ökonomen stehen der Verwendung der Maximinregel sehr skeptisch gegenüber. Rationales Handeln in Situationen der Unsicherheit ist in dieser Sicht gleichbedeutend mit Handeln entsprechend der (subjektiven) Erwartungsnutzentheorie. Rawls „muß“, wenn diese Identifizierung korrekt ist, an seiner ersten Aufgabe scheitern, da es dann keine Situationen der Unsicherheit gibt, in der die Verwendung der Maximinregel rational sein kann. Für eine Reihe Ökonomen, die sich mit Rawls beschäftigt haben, gilt es deshalb als für ausgemacht, daß Rawls hier irrt, und die Parteien des Urzustands selbstverständlich der Erwartungsnutzentheorie folgen werden.³ Verknüpft mit der Verteidigung des Prinzips des fehlenden Grundes würde das formal für die einzelnen Parteien in den Utilitarismus à la Harsanyi führen, ohne ein solches Prinzip zu einer Art allgemeineren Utilitarismus mit eventuell unterschiedlicher Gewichtung der Nutzen verschiedener Individuen.

Wenn diese Kritik an Rawls korrekt wäre, wäre das gleichbedeutend mit dem Scheitern seines Projekts, aus der Einigung von als selbstinteressiert unterstellter Parteien in einer fairen Verhandlungssituation brauchbare Prinzipien für die Gestaltung der Grundstruktur einer gerechten Gesellschaft herzuleiten. Die Probleme beginnen schon damit, daß es zunächst keinen Grund gibt, anzunehmen, daß die Parteien sich auf eine gemeinsame Einschätzung der Wahrscheinlichkeiten und auf gemeinsame Nutzenfunktionen einigen können. Um dieses Problem zu lösen, müßte man zu einer Version

¹ Rawls [1975, 174-220, 587-594], Rawls [1974], Rawls [1999f] und Rawls [2001, 94-134]

² Z. B. in [Rawls 2001, 94-96, 99].

³ Prominente Beispiele sind Harsanyi [1976a] und – wenn auch vorsichtiger – Arrow [1977a].

der Harsanyi Doktrin greifen, die versichert, daß rationale Individuen bei gleicher Informationslage auch zu denselben Urteilen gelangen. Aber während es für die Wahrscheinlichkeiten mit der Gleichverteilung wenigstens einen Fokuspunkt gibt, von dem man mit gutem Willen annehmen kann, daß dies der Punkt sei, auf den sich gleich nicht-informierte rationale Individuen einigen werden, gibt es keinen solchen Punkt für die Nutzenfunktionen. Die Behauptung, die Parteien im Urzustand würden dieselbe Nutzenfunktion zugrunde legen, ist also rein formal, und hilft in keiner Weise *uns*, uns auf ein System vergleichbarer und in eine utilitaristische Wohlfahrtsfunktion einzusetzende Nutzenfunktionen zu einigen. Nur in einem rein formalen und inhaltlich leeren Sinn wären wir damit der Lösung des von Rawls gestellten Problems, Prinzipien zu finden, die eine öffentlich anerkannte Basis bilden könnten, von der aus gesellschaftliche Konflikte gelöst werden können, näher gekommen.

Gerade die Selbstverständlichkeit, mit der Ökonomen z.T. davon ausgehen, daß im Rawlsschen Urzustand die Erwartungsnutzentheorie anzuwenden sei und damit das Rawlssche Argument für seine zwei Prinzipien widerlegt sei, lassen es als eine dringliche Aufgabe erscheinen, die Argumentation Rawls genauer zu rekonstruieren und auf ihre Stichhaltigkeit hin zu untersuchen.

Für eine Verteidigung der Position Rawls, die sich auch an Ökonomen richten möchte, ist es darum vordringlich, sich gerade mit dieser Fragestellung auseinanderzusetzen. Im Folgenden möchte ich deshalb eine Rekonstruktion des Rawlsschen Arguments für seine Prinzipien versuchen, die nachvollziehbar macht, weshalb die Parteien im Urzustand für diese Prinzipien entscheiden sollten.

Es gibt ein zweites Problem in der von Rawls vorgeschlagenen Lösung der von ihm gestellten Aufgabe. Sein zweites Prinzip der Gerechtigkeit macht Gebrauch von einem Index von Grundgütern mit dessen Hilfe die schlechtest Gestellten einer Gesellschaft bestimmt werden sollen, und auf den das Differenzprinzip anzuwenden ist. Rawls hat zwar an der einen oder anderen Stelle seines Werkes Hinweise darauf gegeben, wie er diesen Index verstehen möchte und auch sehr vage darauf, wie er erstellt werden sollte.⁴ Aber man wird zugeben müssen, daß sie nicht ausreichen, um irgendeine Idee davon zu bekommen, wie man denn nun konkret vorgehen sollte, um einen solchen Index zu erstellen.

Für Ökonomen liegt es nahe, die Aufgabe der Erstellung eines Index so zu verstehen, daß es darum geht, auf der Basis von individuellen Nutzenfunktionen über die Grundgüter eine Wohlfahrtsfunktion zu erstellen, die gewisse wünschenswerte Eigenschaften besitzt. Und weiter liegt es nahe, diese Eigenschaften mit denen zu identifizieren, die Arrow in seinem berühmten „Unmöglichkeitstheorem“ verlangt hatte. Das Ergebnis ist klar: Das „Indexproblem“ ist unlösbar, außer man greift zu den Mitteln, die allgemein zu einer Überwindung des Unmöglichkeitstheorem zur Verfügung stehen, wie etwa die Zulassung von Nutzenvergleichen. Aber wenn Rawls selbst Nutzenvergleiche für möglich halten müßte, um die Konstruktion seines Index von Grundgütern zu ermöglichen, schwächt das dann nicht sein Argument gegen den Utilitarismus, das zumindest teilweise auf seiner Skepsis Nutzenvergleichen gegenüber zu beruhen scheint?

⁴ Etwa Rawls [1973, 90-95], Rawls [1993b, 39 f., 187-190] und Rawls [2001, 57-61, 170-176, 188 f.].

Rawls zeigte sich in einem gewissem Maß von diesem Problem beeindruckt, hielt es aber über die Annahme gelöst „that all citizens have a rational plan of life that requires for its fulfillment roughly the same kind of primary goods“. Annahmen dieser Art seien notwendig, denn: „Without restrictive assumptions of this kind, the index problem is known to be insoluble.“ [Rawls 1993c, 180 f., Anm. 8]

Aber es ist die Frage, ob Rawls in der Tat *die* restriktiven Annahmen über die Ähnlichkeit von Nutzenfunktionen machen möchte, die notwendig wären, um das Indexproblem in der Weise zu lösen, wie er es im Text der zitierten Anmerkung vorschlägt. In der Tat scheint Rawls das Problem auf eine weit radikalere Weise gelöst zu haben, indem er nämlich an seinen Index einfach nicht die Anforderungen stellt, die das Problem erst entstehen lassen. Deutlich versteht er ja seinen Index *nicht* als ein Maß für einen irgendwie aggregierten Nutzen der Individuen,⁵ noch verlangt er von der gesuchten Entscheidungsregel, daß es sich dabei um die Wahl auf Grundlage einer Ordnung über alle möglichen gesellschaftlichen Zustände handelt. Ohne solche Ansprüche verschwindet aber das Problem.

In meiner Rekonstruktion werde ich dieser radikaleren Lösung des Indexproblems folgen, auch wenn es dabei eher um die Zerschlagung eines gordischen Knotens handelt: Das heißt, auch wir verlangen keine Verknüpfung zwischen den „tatsächlich“ vorliegenden Präferenzen und der hier rekonstruierten Lösung des Gerechtigkeitsproblems.

Rekonstruktion soll zwei Dinge bedeuten: Zum einen den Anspruch, den Intuitionen Rawls sehr nahe zu sein und besser als z.T. in der Literatur die tatsächliche Struktur der Arguments in TG wiederzugeben.⁶ Zum anderen aber – und das gehört zum Wesen einer Rekonstruktion – gibt es Punkte, in denen das rekonstruierte Argument sich von seinem Original unterscheiden wird. Der Anspruch hier ist, daß es sich bei den Unterschieden einerseits um freundliche „Verbesserungen“ handelt, die die Lücken in der ursprünglichen Argument füllen, ohne seine Intuitionen zu verraten. Andererseits – und auch das ist für eine etwas formaleren Rekonstruktionen eines normalsprachlich vorgetragenen Arguments nicht außergewöhnlich – muß zugegeben werden, daß gegenüber der Vielschichtigkeit des Rawlsschen Originals meine Rekonstruktion ärmer sein wird. Von den vielen Fäden, die Rawls in seinem kunstvollen Teppich seines Argumentes zusammenbindet, werden sich im Gewebe der Rekonstruktion nur einige

⁵ Technisch gesprochen legt Rawls seiner kollektiven Entscheidungsregel weder das Paretoprinzip, noch das Prinzip der Unabhängigkeit von irrelevanten Alternativen, noch das Prinzip der Nichtexistenz eines „Diktators“ auf. Seine Entscheidungsregel ist von den tatsächlich vorliegenden Präferenzen einfach unabhängig.

⁶ Der Anspruch stützt sich darauf, daß in mir bekannten Rekonstruktionen a) das unten in Abschnitt 6.2 entwickelte und meines Erachtens zentrale Element der Rawlsschen Argumentation keine Rolle spielt, und b) das resultierende Entscheidungskriterium nicht auf einem Index von Grundgütern beruht, sondern etwa auf Nutzenwerten oder allgemeiner auf einer Ordnung über Paare von sozialen Positionen, oder c) keine Extension des Maximinkriteriums und eventuell gar ein utilitaristisches Kriterium darstellt oder d) statt *einer* Lösung nur ein Kontinuum von Lösungen anbietet, unter denen eine Extension des Maximinkriteriums nur einen Sonderfall bildet, oder e) Voraussetzungen gemacht werden, die im Kontext des Urzustandes weder gewissermaßen selbstverständlich sind, noch von Rawls selbst gemacht werden. Beispiele finden sich in Sen [1970, 156-158], Sen [1986, 1114-1121], Arrow [1977a, 214 f.], Arrow [1977b], Yamashige [1995], Deschamps u. Gevers [1978], Binmore [1998, vor allem S. 422-453].

wenige wiederfinden.

Von den vielen Unterschieden zum Rawlsschen Original möchte wenigstens einige wichtige auflisten und kurz diskutieren:

- Zwar hat Rawls in seinem Hauptwerk [Rawls 1973, 62 f., 151, 303] kurz diskutiert, was er die allgemeine Konzeption der Gerechtigkeit nennt (kurz gesagt, die Anwendung des Differenzprinzips auf alle Grundgüter⁷), doch gibt es keinen Zweifel, daß sein eigentliches Interesse dem gehört, was er die (eine) spezielle Konzeption nennt, genauer seine zwei Prinzipien der Gerechtigkeit, in denen einem der Grundgüter (den Grundfreiheiten) ein Vorrang eingeräumt wird. Dieser speziellen Konzeption sind weite Teile seines Hauptwerks gewidmet, in den späteren Werken wird die allgemeine Konzeption meistens nicht einmal mehr erwähnt. In Rawls' eigenen Worten:

In developing justice as fairness I shall, for the most part, leave aside the general conception of justice and examine instead the special case of the two principles in serial order. [Rawls 1973, 63]

Demgegenüber lege ich hier hauptsächlich und zuerst eine Rekonstruktion dessen vor, was ich hier in Anlehnung an Rawls als allgemeine Konzeption der Gerechtigkeit bezeichnet habe. Die spezielle Konzeption diskutiere ich nur sekundär, und weise dabei auf einige Voraussetzungen hin, die sich wenigstens in meiner Rekonstruktion für eine Herleitung des Vorrangs *gleicher* Freiheit als notwendig erweisen, und die man als problematisch erachten könnte.

- Meine Rekonstruktion führt nicht wirklich zum Differenzprinzip oder zu einer Maximin-Verteilungsregel, sondern zu dem was ich „beschränkte Leximin-Ordnung“ nenne, die doch immerhin eine gewisse Familienähnlichkeit mit der Maximin-Verteilungsregel hat. Die Schwierigkeiten der Maximin-Verteilungsregel werde ich kurz diskutieren.
- Rawls möchte das Differenzprinzip nicht auf Individuen angewendet wissen, sondern auf Repräsentanten, typische Vertreter von Gruppen, die durch irgendeine Form der Aggregation gewonnen werden müssen. Dies betrachtet er als eine – unvermeidliche und legitime – praktische Vereinfachung, die aber schon von den Parteien im Urzustand so übernommen wird. [Rawls 1973, 98] Demgegenüber führt meine Rekonstruktion zu einer Regel, die im Prinzip auf Individuen anzuwenden ist. Das ist angesichts des individualistischen Charakters der von Rawls

⁷ Das ist zumindest eine mögliche Interpretation von Rawls [1973, 62, 303]. Allerdings diskutiert Rawls in den entsprechenden Passagen nicht, wie genau in der allgemeinen Konzeption verschiedene Güter kommensurabel gemacht werden, insbesondere führt hier die Idee eines Index nicht ein. Da er aber in der „letzten Formulierung“ seiner zwei Prinzipien ebenfalls auf einen Hinweis auf einen Index verzichtet und die „letzten Formulierungen“ der allgemeinen und besonderen Konzeption in dieser Hinsicht parallel sind [Rawls 1973, 302 f.], scheint mir die Interpretation im Text gerechtfertigt. Rawls selbst nennt seine Formulierung der allgemeinen Gerechtigkeitskonzeption „extremly vague“ und verweist auf die Notwendigkeit der Interpretation [Rawls 1973, 62].

vorgeschlagenen Entscheidungssituation bei einer formalen Rekonstruktion mit den Mitteln der formalen Entscheidungstheorie kaum anders zu erwarten. (Auch bei ihrer Anwendung mag ein gewisses Maß der Aggregation notwendig sein, aber diese Aggregation wäre nicht Teil der im Urzustand gewählten Regel.)

- Während ich eine Form von zwei Prinzipien mit einem Vorrang gleicher Grundfreiheiten für alle immerhin diskutiere, und Voraussetzungen aufzeige, unter denen sie sich herleiten ließe, behandle ich keiner Weise der Vorrang der Chancengleichheit innerhalb des zweiten Prinzips. Allerdings wird man sagen müssen, daß sich auch bei Rawls nirgends eine ausführliche Begründung für den Vorrang der Chancengleichheit aus dem Urzustand findet, die vergleichbar wäre mit seiner Begründung für den Vorrang der Grundfreiheiten oder für das Differenzprinzip. In seinen späteren Arbeiten wird der Vorrang fairer Chancengleichheit gewissermassen eingeschmuggelt: Es gibt eine Argumentation für das Ganze der zwei Prinzipien, die für Rawls wesentlich vom Vorrang der Grundfreiheit getragen wird, und dann ein Argument für das Differenzprinzip unter Voraussetzung des Vorrangs der Grundfreiheiten, die Chancengleichheit dagegen wird einfach im Ganzen mitgenommen.⁸ Nur in seinem Hauptwerk findet sich ein kurzes Argument in zwei Schritten: Zunächst wird angesichts der Komplexität einer Gerechtigkeitskonzeption, die jeweils alle Änderungen individueller Positionen in Betracht ziehen wollte, die praktische Überlegenheit einer reinen Verfahrensgerechtigkeit begründet, dann die Notwendigkeit der Chancengleichheit dafür als "evident" behauptet, daß distributive Gerechtigkeit rein prozedural implementiert werden kann. [Rawls 1973, 86] Dieses Argument scheint durch eine formale Rekonstruktion wenig gewinnen zu können.
- Rawls setzt voraus, daß die Parteien die Größe ihrer Gesellschaft unbekannt ist. [Rawls 1973, 200] Dagegen ist in meiner Rekonstruktion eine gewisse Größe N der Gesellschaft als gegeben angenommen. Diese Annahme ist nur als Vereinfachung gedacht, und könnte in einer Weiterentwicklung der Rekonstruktion aufgegeben werden.
- Idealweiser sollte nach Rawls das Argument im Urzustand rein deduktiv sein, man sollte sich um eine "moralische Geometrie" bemühen. [Rawls 1973, 121] Dies läßt vermuten, daß formale Rekonstruktionen seiner Argumente – wie die vorliegende – seinen Intentionen nahe kommen sollten. Allerdings ist hier auf ein wichtiger Unterschied zwischen formalen Argumenten, wie sie etwa aus der Social-Choice-Literatur vertraut sind, und dem, was Rawls vorschwebt, hinzuweisen: Deduktive Strenge soll für ihn nur heißen, daß alle (u.a. psychologischen) Voraussetzungen in der Darstellung des Urzustands explizit gemacht sind. Einer im engeren Sinn formalen Darstellung entzieht sich die Argumentation aber aus zwei Gründen: Die Zahl der möglichen Überlegungen (considerations), die sich im Urzustand als Argumente anführen lassen, ist unbegrenzt. Ein Argument kann

⁸ Das ist die Struktur des Argumentes etwa in Rawls [2001, 94-134].

deshalb immer nur vorläufig sein. Verschiedene Überlegungen sprechen für verschiedene Gerechtigkeitskonzeptionen. Eine Entscheidung verlangt deshalb immer ein Abwägen. Dieses Abwägen läßt sich zwar vielleicht formal darstellen, aber nicht letztlich formal begründen. [Rawls 2001, 133 f.] Meine Rekonstruktion ist im engeren Sinn formal und axiomatisch und verzichtet ganz auf eine Rekonstruktion eines Prozesses des Abwägens.

Im Zentrum der hier vorgelegten Rekonstruktion steht die These, daß der Rawlssche Anspruch, rationale Parteien im Urzustand würden sich für seine zwei Prinzipien der Gerechtigkeit (oder etwas ähnliches) entscheiden, dann nachvollziehbar wird, wenn man mit Rawls annimmt, daß Rationalität verlangt, keine Entscheidungen zu treffen, die man später – in einem noch zu präzisierenden Sinn – bereut, und daß diese eventuelle Reue sich darauf bezieht, daß (gegeben bestimmte Vorstellungen vom Guten oder rationale Lebenspläne) Lebenssituationen bzw. -aussichten (im Urzustand charakterisiert durch einen Vektor von Grundgütern) – wieder in einem noch zu präzisierenden Sinn – „unerträglich“ sind.

Ich werde damit beginnen, im nächsten Abschnitt die grundlegende Intuition darzustellen. Abschnitt 6.3 modelliert die Entscheidungen eines einzelnen Individuums, das sich *nicht* hinter dem Schleier des Nichtwissens befindet, und dient dem Zweck, zu rationaler Reue „fähige“ Individuen und das Konzept inakzeptabler Positionen formal einzuführen und verständlich zu machen, weshalb wir annehmen werden, daß ein Individuum, das eine Position als inakzeptabel einstuft, dies auch für alle Positionen tun wird, die es als schlechter ansieht. Abschnitt 6.4 stellt dann zunächst meine Rekonstruktion der allgemeinen Rawlsschen Konzeption der Gerechtigkeit ausführlich dar und diskutiert die Alternativen des Maximinkriteriums, der zwei Prinzipien mit dem Vorrang der Freiheit und des Utilitarismus.

6.2 Entscheiden ohne Reue

Eine normative Theorie zur Entscheidung bei Unsicherheit enthält zumindest implizit eine Aussage über das richtige Verhältnis von ex ante und ex post Perspektive. In der orthodoxen Erwartungsnutzentheorie dominiert dabei in gewisser Weise die ex ante Perspektive. Die Tatsache, daß ein Mensch nach der Auflösung der Unsicherheit, unter den ihm dann bekannten Umständen, seine Entscheidung bereut und wünscht, anders gewählt zu haben, ist in jedem Fall völlig irrelevant, solange er seine Entscheidung entsprechend seiner subjektiven Wahrscheinlichkeitseinschätzungen und Präferenzen über Lotterien getroffen hat. Eine Entscheidung unter Unsicherheit von ihrem aktuellen Ergebnis her zu beurteilen, ist eine Dummheit, so wird uns gesagt – eine Dummheit, die allerdings oft begangen wird. [Binmore 1998, 224 f.]

Birger Priddat hat in einem bemerkenswerten Aufsatz [Priddat 1998] darauf hingewiesen, daß eine Analyse des menschlichen Subjekts, die die Semantik des Leides und Mißerfolgs, die in unseren subjektiven Lebenswelten und unserer Deutung von Entscheidungen und ihren Folgen eine große Rolle spielen, ernst nimmt, mit der Möglichkeit rechnen muß, daß die Rationalitätszuschreibung der Erwartungsnutzentheorie

zu einer Rationalitätszumutung wird, der menschliche Subjekte nicht entsprechen können. Die Folgen einer von außen als „rational“ analysierbaren Entscheidung können so – für das betroffenen Subjekt – unerträglich sein, daß sie selbst ihre Entscheidung nicht mehr als rational interpretieren können. Sie selbst können also (so sind wir Menschen gebaut) gar nicht anders, als ihre eigene Entscheidung zu bereuen, zu verfluchen und als einen verhängnisvollen Fehler (und damit das Gegenteil von rational) zu beurteilen.

Innerhalb der Erwartungsnutzentheorie gibt es keine Übersetzungs- oder Rekonstruktionsmöglichkeit, dieser Klagen. Es fehlt der Ökonomischen Theorie, wie es Pridat ausdrückt, die dafür notwendige Semantik. Es ist deshalb nur konsequent, dieses Klagen des Subjekts als bloßes Rauschen zu überhören, als irrelevant oder gar eben als eine Dummheit abzutun. Das beste, was die Erwartungsnutzentheorie einem solchen Subjekt anbieten kann, ist der Rat, seine Nutzenfunktion so zu ändern, daß für zukünftige Entscheidungen die Erfahrung des Mißerfolgs und Scheiterns (als rationales Subjekt) unwahrscheinlicher wird. Doch dieser Rat kann die Spannung zwischen ex ante und ex post Perspektive nicht lösen, da ja in keiner Weise auszuschließen ist, daß das klagende Subjekt für den (möglicherweise korrekt) als unwahrscheinlich eingestuften Fall des Mißerfolgs sein Leiden und Klagen antizipiert hat. Solange es der Erwartungsnutzentheorie folgt, ist die Möglichkeit des Eintretens dieses Falls nicht auszuschließen.

Es gibt aber einen Ratschlag, der die Spannung zwischen ex ante und ex post Perspektive zu lösen vermag: Wenn Du die Folgen deiner Entscheidungen nicht ertragen kannst, dann entscheide anders. Genauer: Wenn Du die Wahl hast zwischen Alternativen, die unerträgliche Folgen ausschließen, und solchen, die das nicht tun, wähle eine der ersteren. Diese Regel – auf passende, d. h. konsistente Weise mit Regeln für die Entscheidung zwischen verschiedenen Alternativen, die unerträgliche Folgen ausschließen und zwischen verschiedenen Alternativen, die unerträgliche Folgen nicht ausschließen ergänzt – kann für einen weiteren Bereich Kohärenz zwischen der ex ante und der ex post Perspektive herstellen als dies die traditionelle subjektive Erwartungsnutzentheorie kann.

Während Rawls in einer kurzen Bemerkung im ersten Teil der Theorie der Gerechtigkeit behauptet, daß die für den Urzustand vorausgesetzte Konzeption der Rationalität „mit Ausnahme einer Eigenschaft“ – der Freiheit von Neid – „der in der sozialwissenschaftlichen Theorie übliche Standardbegriff“ sei [Rawls 1975, 166f], machen seine ausführlicheren Ausführungen zum Begriff der abwägenden Vernunft im dritten Teil der Theorie der Gerechtigkeit deutlich, daß es zumindest eine weitere Ausnahme gibt: Für Rawls gehört die Maxime, so zu handeln, daß spätere Reue, in dem Sinn, daß man seine eigenen früheren Entscheidungen nicht mehr als rational akzeptieren kann, zum Begriff der Rationalität und ist auch im Urzustand anzuwenden:

„Now one feature of a rational plan is that in carrying it out the individual does not change his mind and wish that he had done something else instead. A rational person does not come to feel an aversion for the foreseen consequences so great that he regrets following the plan he has adopted. The absence of this sort of regret is not however sufficient to insure that a plan is rational. There may be another plan open to us such that were we to

consider it we would find it much better. Nevertheless, if our information is accurate and our understanding of the consequences complete in relevant respects, we do not regret following a rational plan, even if it is not a good one judged absolutely. In this instance the plan is objectively rational. We may of course, regret something else, for example that we have to live under such unfortunate circumstances that happy life is impossible. Conceivably we may wish that we had never been born. But we do not regret that, having been born, we followed the best plan as bad as it may be when judged by some ideal standard. A rational person may regret his pursuing a subjectively rational plan, but not because he thinks his choice is in any way open to criticism. For he does what seems best at the time, and if his beliefs later prove to be mistaken with untoward results, it through no fault of his own. There is no cause for self-reproach. There was no way of knowing which was the best or even a better plan.

Putting these reflections together, we have the guiding principle that a rational individual is always to act so that he need never blame himself no matter how his plans finally work out. Viewing himself as one continuing being over time, he can say that at each moment of his life he has done what the balance of reasons required, or at least permitted. Therefore any risks he assumes must be worthwhile, so that should the worst happen that he had any reason to foresee, he can still affirm that what he did was above criticism. He does not regret his choice, at least not in the sense that he later believes that at the time it would have been more rational to have done otherwise. This principle will not certainly prevent us from taking steps that lead to misadventure. Nothing can protect us from the ambiguities open to us. Acting with deliberative rationality can only insure that our conduct is above reproach, and that we are responsible to ourselves as one person over time. We should indeed be surprised if someone said that he did not care about how he will view his present actions later any more than he cares about the affairs of other people (which is not much, let us suppose). One who rejects equally the claims of his future self and the interests of others is not only irresponsible with respect to them but in regard to his own person as well. He does not see himself as one enduring individual.

Now looked at in this way, the principle of responsibility to self resembles a principle of right: the claims of the self at different times are to be so adjusted the self at each time can affirm the plan that has been and is being followed. The person at one time, so to speak, must not be able to complain about actions of the person at another time. This principle does not, of course, exclude the willing endurance of hardship and suffering; but it must be presently acceptable in view of the expected or achieved good. From the standpoint of the original position the relevance of responsibility to self seems clear enough. Since the notion of deliberative rationality applies there, it means that the parties cannot agree to a conception of justice if the consequences of applying it may lead to self-reproach should the least

happy possibilities be realized. They should strive to be free from such regrets. And the principles of justice as fairness seem to meet this requirement better than other conceptions, as we can see from the earlier discussion of the strains of commitment (§29).“ [Rawls 1999g, 370 f.]

Rawls verlangt also, daß ein rationale Entscheider (ein vernunftgeleiteter Mensch) so entscheiden soll, „daß er sich später keine Vorwürfe zu machen braucht, wie auch immer sich die Dinge schließlich entwickeln mögen“. [Rawls 1975, 461] In den hier zitierten Überlegungen Rawls' scheint dabei vorausgesetzt, daß die betrachteten Subjekte substantielle Präferenzen über die möglichen Ergebnisse ihrer Entscheidungen haben. Er kann deshalb präzisieren: „Demgemäß müssen alle von ihm eingegangenen Risiken gerechtfertigt sein, derart, daß er auch beim Eintritt des Schlimmsten, das er überhaupt begründet voraussehen konnte, noch sagen kann, was er getan habe, sei nicht kritisierbar.“ (ebd.) Entsprechendes gilt für die Entscheidung im Urzustand: „Da hier der Begriff der abwägenden Vernunft anwendbar ist, bedeutet sie [die Verantwortlichkeit vor sich selbst], daß die Beteiligten keiner Gerechtigkeitsvorstellung zustimmen können, deren Anwendung beim Eintritt der ungünstigsten Möglichkeiten zu Selbstvorwürfen führen könnte. Sie sollten solche Reue zu vermeiden trachten.“ [Rawls 1975, 462]

Rawls [1975, 462] schließt seine Überlegungen zur Reue mit dem Hinweis darauf ab, daß „die Grundsätze der Gerechtigkeit als Fairneß [...] diese Forderung besser erfüllen als andere Vorstellungen, wie sich aus der Diskussion der Vertragstreue ergab“ und verweist dabei auf den Abschnitt 29 der Theorie der Gerechtigkeit. Nun findet sich in dieser Diskussion der Vertragstreue kein Hinweis auf das Problem der Reue in dem hier vorausgesetzten Sinn, sodaß sich die Frage stellt, weshalb Rawls davon überzeugt ist, dort schon gezeigt zu haben, daß auch was die Vermeidung der Reue angeht, seine Grundsätze anderen Gerechtigkeitsvorstellung überlegen sind.

Diese Frage ist leicht zu beantworten, wenn wir zunächst bemerken, daß Rawls sowohl in der Diskussion des Problems der Reue wie in der Diskussion der Vertragstreue den Blick auf die schlechtestmöglichen Ergebnisse lenkt, und dann – aus verschiedenen Gründen – verlangt, daß diese Ergebnisse nicht – ex post – bestimmte Fähigkeiten des Menschen überfordert. Das ist im Fall der Vertragstreue die Fähigkeit eine Abmachung einhalten zu können. Rawls [1999g, 153 f.] geht also davon aus, daß es Konsequenzen aus geschlossenen Verträgen geben kann, die die Vertragsparteien nicht akzeptieren können, weshalb diese Verträge nicht einhaltbar sind. Diese Möglichkeit gehört für ihn zu den allgemeinen Tatsachen der Psychologie des Menschen, die den Parteien im Urzustand bekannt sind, und die sie bei ihrer Entscheidung beachten. Im Fall der möglichen „Reue“ ist es die Fähigkeit, sich selbst auch ex post als verantwortungsvolles rationales Wesen betrachten zu können, die angesichts besonders schlechter Folgen einer Entscheidung möglicherweise überfordert wird.

Beiden Argumentationslinien ist zum einen gemeinsam, daß sie davon ausgehen, daß es neben dem besser und schlechter der Präferenzordnungen über mögliche Ergebnisse einer Entscheidung bzw. der Konsequenzen eines Vertrags einen qualitativen Schnitt gibt, der die Situation der betroffenen Individuen gewissermaßen kategorial

ändert: Ergebnisse einer Entscheidungen oder eines Vertrags können so übel sein, daß Menschen, von denen an sich angenommen wird, daß sie fähig sind, die Konsequenzen ihrer eigenen Entscheidungen zu tragen, dies nicht mehr können, ohne ihre eigene Rationalität zu verneinen, bzw. daß Menschen, von denen an sich angenommen wird, daß sie fähig sind, Verträge auch wenn diese für sie mit Kosten verbunden sind einzuhalten (oder im konkreten Fall des Urzustands aus einem Sinn für Gerechtigkeit heraus zu handeln), in dieser Fähigkeit überfordert werden. Gemeinsam ist diesen Argumentationslinien auch, daß Rawls in beiden Fällen verlangt, die mögliche Überforderung dadurch zu vermeiden, daß entsprechend üble Folgen einer Entscheidung, eines Vertrags soweit möglich ex ante verhindert werden: Sei es um eine Rationalitätsform sicherzustellen, die sowohl die ex ante wie auch die ex post Perspektive umfaßt, sei es um sicherzustellen, daß ein Vertragsabschluß guten Glaubens geschieht (eine formale Bedingung eines gültigen Vertrags) und daß im besonderen Fall des Urzustands überhaupt eine wohlgeordnete, d. h. eine einer Gerechtigkeitsvorstellung entsprechend gestaltete Gesellschaft begründet wird.⁹

Ein ähnliche Argumentationsstruktur findet sich bei Rawls auch an anderen Stellen. Als ein Beispiel kann eines der Argumente in seiner „verbesserten“ Begründung für den Vorrang der Freiheit in „The basic liberties and Their Priority“ [Rawls 1993a, 310-312] dienen. Rawls macht hier darauf aufmerksam, daß es zumindest Menschen gibt, die von bestimmten gegebenen religiösen, philosophischen oder moralischen „Anschauungen von unserer Beziehung zur Welt“ so fest überzeugt sind, daß für sie „das Bekenntnis zu solchen Anschauungen und den Konzeptionen des Guten, die sie hervorbringen“ nicht verhandlungsfähig sind. Den Schutz dieser Überzeugungen und der entsprechenden Verhaltensweisen können Menschen, die solche Überzeugungen haben, nicht zugunsten von Grundgütern wie Einkommen und der Möglichkeit des Zugangs zu attraktiven gesellschaftlichen Positionen aufgeben oder gefährden. Weil der Schutz dieser Überzeugungen für die von ihnen möglicherweise vertretenen Personen unaufgebar, nicht akzeptabel ist, können auch die Parteien im Urzustand diesen Schutz nicht aufs Spiel setzen, weder zugunsten von Einkommen und Möglichkeiten des Zugangs zu gesellschaftlichen Positionen, noch auch aufgrund der möglicherweise geringen Wahrscheinlichkeit einer unterdrückten Minderheiten-Weltsicht anzuhängen:

⁹ Das ist meine Interpretation der *zwei* Gründe, die Rawls [1999g, 153] dafür anführt, daß die Parteien die „strains of commitment“, d. h. die Gefahr eines möglichen Vertragsbruches soweit möglich ex ante ausschließen werden.

Der erste (bei Rawls der zweite) ist eine moralische Bedingung für einen gültigen Vertrag, und gehört damit auf die Seite der Modellierung der Bedingungen des Urzustands, die u.a. die Beschreibung der Aufgabe der Parteien einschließt und die Dimension des *Vernünftigen* einfangen soll.

Der zweite kann auch als eine Konsequenz der *Rationalität* der Parteien verstanden werden: In der idealen Theorie haben sie die Gewißheit, daß eine Gerechtigkeitskonzeption, die die „strains of commitment“ berücksichtigt auch verwirklicht werden wird. Diese Gewißheit gibt es nicht, wenn der Rahmen einer idealen Theorie „wissentlich“ verlassen wird, indem eine Gerechtigkeitskonzeption beschlossen wird, deren Konsequenzen gemäß als bekannt vorausgesetzter Gesetze der moralischen Psychologie für einen Teil der Betroffenen inakzeptabel sind. Die Parteien würden mit einer solchen Entscheidung kein kalkuliertes Risiko eingehen, sondern sich auf die nicht kalkulierbaren Unwägbarkeiten einer „ungerechten“ (nicht wohlgeordneten) Gesellschaft einlassen. Für die Unterscheidung der Dimensionen des Vernünftigen und des Rationalen im Urzustand vgl. z. B. Rawls [1999c, 315-317].

„If the parties were to gamble in this way, they would show that they did not take the religious, philosophical, or moral convictions of persons seriously, and, in effect, did not know what a religious, philosophical, or moral conviction was.“ [Rawls 1993a, 311]

Auch in diesem Beispiel finden wir also wieder die strukturellen Elemente einer ex post nicht nur negativ bewerteten sondern unerträglichen, nicht akzeptablen Situation (hier die Unmöglichkeit, die eigenen weltanschaulichen Überzeugungen zu behalten und ihnen entsprechend leben zu können) und der Konsequenz, soweit möglich, solche von anderen möglichen Situationen kategorial verschiedene Situationen zu vermeiden, eben weil sie aus der ex post Perspektive unerträglich sein können.

6.3 Die wissende Partei: eine Vorübung

In einem ersten Schritt formulieren wir das Entscheidungsproblem einer Partei, die sowohl ihre Präferenzen, wie auch die möglicherweise für sie unerträglichen Bündel von Grundgütern kennt.

Sei X die Menge aller möglichen Vektoren $x = (x_1, \dots, x_m)$ von Mengen (amounts, jeweils ausgedrückt in passenden Einheiten) der Grundgüter $j \in M = 1, \dots, m$, die mit gesellschaftlichen Positionen verbunden sein können, und N die Menge $1, \dots, n$ aller Positionen i einer Gesellschaft. Wir nehmen an, daß die Partei über X eine (schwache) Präferenzordnung R besitzen. P und I seien respektive die auf Basis von R definierte strikte Präferenz und Indifferenz.

Für die Bündel von Grundgüter nehmen wir an, daß sie sich in geeigneter Weise numerisch darstellen lassen. Diese Voraussetzung dient vor allem dazu, die in der Rawlschen Schwachen Theorie des Guten und in den folgenden Axiomen vorausgesetzte Größer- bzw. Kleiner-Relation [vgl. Rawls 1999g, 79, 347-350, 381 f.] einzuführen.¹⁰

Axiom 6.1.

$$X = X_1 \times X_2 \dots \times X_m \wedge \forall i \in \{1, \dots, m\} : X_i \subseteq \mathbb{R}$$

Die Präferenzen über die Grundgüter sind, wie auch Rawls annimmt, monoton:

Axiom 6.2. Monotonie:

$$\forall x, x' \in X, i \in N : x \geq x' \Rightarrow xRx'$$

Wir nehmen an, die betrachtete Partei könne für mögliche Mengen $S \subseteq X^n$ von Listen (Vektoren) von in einer Gesellschaft vorhandenen Positionen $a = (a_1, \dots, a_n) \in X^n$ festlegen, welche Liste verwirklicht wird. Die Partei ist unsicher darüber, welche Position sie einnehmen wird. Gesucht ist eine der betrachteten Partei eigene Präferenzordnung \succeq über die Menge X^n auf deren Grundlage, die Partei bei gegebenen S entscheidet. Gegeben die Ordnung \succeq ist die Entscheidungsmenge $C(S)$ für jede Menge S von möglichen Alternativen definiert als:

¹⁰ Dabei sei $x \geq x' \leftrightarrow \forall i : x_i \geq x'_i$ und $x > x' \leftrightarrow x \geq x' \wedge \neg x' \geq x$.

Definition 6.1.

$$C(S) := \{a \in S : \forall b \in S : a \succeq b\}$$

Die Ordnung \succeq gehorche dabei dem Axiom der Dominanz, eine sehr weit verbreitete und einsichtige Annahme für das Verhalten unter Unsicherheit, die sich in einer dem dortigen Kontext angemessenen Variante auch bei Rawls [1975, 450 f.] findet.¹¹

Axiom 6.3.

$$\forall a, b \in X^n : [\forall i, j : a_i R b_j] \rightarrow a \succeq b$$

Zur Interpretation von $C(S)$ nehmen wir dabei an, daß alle Positionen, die Element eines der Vektoren in $C(S)$ sind, möglicherweise realisiert werden.

Für die gegebene Partei, können die „möglichen Welten“, in denen sie sich nach Auflösung der Unsicherheit wiederfinden wird, hinreichend jeweils durch ein Paar (S, x) beschrieben werden, wobei S , die Menge der Möglichkeiten ist, aus denen die Partei faktisch wählen konnte, und x das Bündel an Grundgütern darstellt, das mit der gesellschaftlichen Position, die das Individuum dann ex post faktisch einnimmt, gegeben ist.

Die oben beschriebene Grundidee der „Rawlsschen Reue“ ist es, daß – falls es für die Partei „unerträgliche“ Positionen gibt – diese dadurch charakterisiert sind, daß die Partei den Glauben an ihre eigene Rationalität nicht mehr aufrecht erhalten kann, wenn sie sich a) ex post in einer solchen Position wiederfindet und ihr b) Alternativen zur Wahl standen, mit denen sie hätte *sicher ausschließen* können, sich in einer solchen Position wiederzufinden. *Sicher ausschließen*, sich in einer unerträglichen Position wiederzufinden, kann die Partei genau dann, wenn es in der Menge von Alternativen S Vektoren a gibt, deren Elemente alle nicht unerträglich sind. Wenn wir für „ x ist unerträglich“ abkürzend „ $\text{rej}(x)$ “, für „glaubt in der Welt (S, x) , daß p “ „ $G_{(S,x)}(p)$ “ und für „ist irrational“ „ irr “ schreiben, können wir also formal festhalten:¹²

Axiom 6.4.

$$\forall x : \text{rej}(x) \leftrightarrow$$

$$\forall S \subseteq X^n : [\exists a \in C(S), i \in N : a_i = x] \wedge [\exists b \in S : \neg \exists i \in N : \text{rej}(b_i)] \rightarrow G_{(S,x)}(\text{irr})$$

Wenn ein Individuum in einer ex post möglichen Welt etwas für wahr hält, was falsch ist, dann ist es irrational:¹³

¹¹ Rawls untersucht die Rationalität von „Plänen“ und nicht die der Wahl von Güterbündeln. Seine Formulierung der Dominanz bezogen auf Wahrscheinlichkeiten der Verwirklichung wichtiger Ziele ist aber von direkter Relevanz für unseren Kontext hier, da es – gegeben die Vorstellung der Grundgüter als „Allzweckmittel“ – plausibel ist, daß Dominanz im hier definierten Sinn sich unmittelbar umsetzt in höhere Wahrscheinlichkeiten für die Erreichung (beliebiger) wichtiger Ziele.

¹² Der Geltungsbereich von Quantoren erstreckt sich jeweils – soweit er durch Klammern nicht eingeschränkt wird – bis zum Ende eines Ausdrucks oder bis zum nächsten Quantor über dieselbe Variable.

¹³ Ich möchte hier nicht entscheiden, ob der Ausschluß falscher Überzeugungen zu einem allgemeinen, philosophischen Begriff von Rationalität gehören würde. In jedem Fall schließt die übliche ökonomische Verwendung des Begriffs „Rationalität“ – der Rawls sich im großen und ganzen ja anschließen möchte – das Handeln auf der Basis falscher Überzeugungen aus. Für die Parteien im Urzustand schließt Rawls [1999g, 133] die Entscheidung auf Basis falscher Überzeugungen explizit aus.

Axiom 6.5.

$$\forall S, x : G_{(S,x)}(p) \wedge \neg p \rightarrow \text{irr}$$

Die Partei ist rational:

Axiom 6.6.

$$\neg \text{irr}$$

Wir zeigen nun:

Theorem 6.1. Gegeben die Axiome 6.4, 6.5 und 6.6 gilt:

$$\forall x \in X : \text{rej}(x) \rightarrow \forall S \subseteq X^n : \neg[[\exists a \in C(S), i \in N : a_i = x] \wedge \exists b \in S : \neg \exists i \in N : \text{rej}(b_i)]$$

Beweis. Betrachte ein beliebiges x , so daß $\text{rej}(x)$. Dann gilt wegen Axiom 6.4:

$$\forall S \subseteq X^n : [\exists a \in C(S), i \in N : a_i = x] \wedge [\exists b \in S : \neg \exists i \in N : \text{rej}(b_i)] \rightarrow G_{(S,x)}(\text{irr})$$

Betrachte weiter ein beliebiges S und nimm an, daß für das gegeben x :

$$[\exists a \in C(S), i \in N : a_i = x] \wedge [\exists b \in S : \neg \exists i \in N : \text{rej}(b_i)]$$

Damit würde gelten $G_{(S,x)}(\text{irr})$ und aufgrund Axiom 6.5 irr , in Widerspruch zu Axiom 6.6. Es muß also gelten:

$$\neg[[\exists a \in C(S), i \in N : a_i = x] \wedge [\exists b \in S : \neg \exists i \in N : \text{rej}(b_i)]]$$

Da x und S beliebig gewählt waren, gilt damit das Theorem. □

Theorem 6.2. Gegeben die Definition 6.1 und die Axiome 6.4, 6.5, 6.6 und 6.3 gilt:

$$\forall x, x' \in X : xRx' \wedge \text{rej}(x) \rightarrow \text{rej}(x')$$

Beweis. Betrachte beliebige x, x' mit $xRx' \wedge \text{rej}(x)$. Bestimme a, b , so daß $\forall i \in N : a_i = x \wedge b_i = x'$, und sei $S = \{a, b\}$. Dann gilt wegen Axiom 6.3 $a \succeq b$ und damit entsprechend Definition 6.1 $a \in C(S)$. Gälte nun $\neg \text{rej}(x')$, würde es also S und x geben, so daß

$$\text{rej}(x) \wedge [[\exists a \in C(S), i \in N : a_i = x] \wedge \exists b \in S : \neg \exists i \in N : \text{rej}(b_i)]$$

Ein Widerspruch zu Theorem 6.1. Für beliebige x, x' mit $xRx' \wedge \text{rej}(x)$ muß also gelten $\text{rej}(x')$. □

6.4 Die unwissende Partei

6.4.1 Das Grundmodell

Der Rawlssche Schleier des Nichtwissens schließt aus, daß die Parteien ihre Präferenzen kennen. Auch wenn Rawls nicht von „Präferenzen“ spricht, wird ein solches Wissen durch Beschränkungen wie die folgende ausgeschlossen: „Ferner kennt niemand seine Vorstellung vom Guten, die Einzelheiten seines vernünftigen Lebensplans, ja nicht einmal die Besonderheiten seiner Psyche wie seine Einstellung zum Risiko oder seine Neigung zu Optimismus oder Pessimismus.“ [Rawls 1975, 160]

Die Logik des Schleiers des Nichtwissens, wie ihn Rawls vorschlägt, verlangt auch, daß die Menge der *ihnen* „unerträglichen“ Situationen den Parteien im Urzustand unbekannt ist: Im Urzustand soll ja den Parteien überhaupt alles Wissen über individuelle Eigenschaften verwehrt sein, womit auch sichergestellt wird, daß sich alle Parteien in genau der gleichen Entscheidungssituation finden.¹⁴ Wenn wir also nicht annehmen wollen, daß alle Individuen dieselben Situationen für unerträglich halten, muß den Parteien die Menge der unerträglichen Situationen unbekannt sein.¹⁵

Wir betrachten nun also eine Partei, der nicht nur unbekannt ist, welche Position sie in einer Gesellschaft einnehmen wird, sondern auch, welche Präferenzen sie besitzt und welche Situationen sie für unerträglich empfindet. Wir können weiter auf eine Indizierung der Parteien im Urzustand verzichten, da ja jede Partei sich in genau der gleichen Entscheidungssituation befindet.

Ansonsten beschreiben wir die Entscheidungssituation und -aufgabe der Parteien im Urzustand im wesentlichen wie oben. D.h. wir nehmen weiter an, die betrachtete Partei könne für mögliche Mengen $S \subseteq X^n$ von Listen (Vektoren) von in einer Gesellschaft vorhandenen Positionen $a = (a_1, \dots, a_n) \in X^n$ festlegen, welche Liste verwirklicht wird. Die Partei ist unsicher darüber, welche Position sie einnehmen wird. Gesucht ist die Entscheidungsmenge $C(S)$ für jede Menge S von möglichen Alternativen, die wir wieder so interpretieren, daß alle Positionen, die Element eines der Vektoren in $C(S)$ sind, möglicherweise realisiert werden.

Können die Parteien ex ante die Menge der möglichen Mengen $S \subseteq X^n$ beschränken?

Zur Vereinfachung nehmen wir in Folgenden durchgehend an, daß sich die Parteien in den Entscheidungssituationen durchgängig nur endlichen Mengen von Möglichkeiten $S \subseteq X^n$ gegenübersehen. Wir definieren also deshalb zunächst:

¹⁴ Die *einzig* „besondere Tatsache“, die die Parteien kennen, ist die, daß in ihrer Gesellschaft die „circumstances of justice“ gegeben sind. [Rawls 1973, 137]. Diese eine besondere Tatsache bezieht sich auf die Gesellschaft, nicht auf ein Individuum. Den Parteien ist alles unbekannt, was sie unterscheidet, die Entscheidung im Urzustand entspricht deshalb derjenigen, einer jeden (zufällig ausgewählten) Person. [Rawls 1973, 139]

¹⁵ Rawls' wichtigstes Beispiel für „unerträgliche“, inakzeptable Positionen, ist die Situation eines Menschen mit „religiösen, philosophischen oder moralischen Überzeugungen“, dem verwehrt wird, diesen Überzeugungen gemäß zu leben. Rawls [1992b, 181 f.] betont aber, daß die „Parteien natürlich nicht sicher sein können, daß die von ihnen vertretenen Personen sich zu solchen Anschauungen bekennen“. Sie müssen allerdings mit dieser Möglichkeit rechnen, und Rawls nimmt an, daß sie normalerweise solche Anschauungen haben. In diesem für Rawls zentralen Beispielfall geht er also explizit davon aus, daß den Parteien unbekannt ist, was sie genau ex post als nicht akzeptabel betrachten werden.

Definition 6.2.

$$\Sigma = \{S : S \subseteq X^n \wedge S \text{ ist endlich}\}$$

Eine der von Rawls besonders betonten Voraussetzungen für seine Theorie ist die zu den Anwendungsverhältnisse der Gerechtigkeit gehörende Bedingung der „mäßigen Knappheit“. Ohne eine gewisse Knappheit gäbe es keine Notwendigkeit für die Zusammenarbeit und Regeln der Gerechtigkeit. Aber die Knappheit soll so sein, daß „allseitig vorteilhafte Regelungen möglich“ sind. [Rawls 1975, 149 f.] In der von mir vorgeschlagenen Interpretation und Rekonstruktion heißt das – und ist den Parteien bekannt –, daß es unter den zur Wahl stehenden, als Vektoren von Vektoren von Grundgütern dargestellten „Gesellschaftsmodellen“ solche gibt, die allen Parteien eine Situation sichern, die erträglich ist. Da uns und den Parteien allerdings unbekannt ist, welche Positionen *genau* unerträglich sein werden (dies ist hinter dem Schleier des Nichtwissens als unbekannt angenommen), können wir und die Parteien diese Bedingung – die wir dann allerdings den möglichen Welten, in denen sich die Parteien nach Auflösung der Unsicherheit wiederfinden, auferlegen werden¹⁶ – nicht dazu verwenden, ex ante die möglichen Mengen von Alternativen S zu beschränken.¹⁷

Den Parteien ist unbekannt, welche Präferenzen und strikten Abneigungen sie haben. Sie können die Menge der möglichen Mengen von Alternativen nur beschränken, wenn ihnen als Teil der „allgemeinen Tatsachen über die menschliche Gesellschaft [...]“, die sich aus dem Alltagsverstand und allgemein anerkannten Analysemethoden ergeben“ [Rawls 1975, 160 f.], obere oder untere Schranken dieser Menge bekannt sind.

Nun gehört Wissen über eine kleinste obere Schranke von Bündeln von Grundgütern, die Menschen für unerträglich halten können, kaum zu solchen allgemeinen Tatsachen. Das beste, auf was wir vielleicht hoffen können, ist allgemeines Wissen über Bündel von Grundgüter, die sicher von Menschen (nicht notwendiger allen) als unerträglich empfunden werden, und möglicherweise auch von Bündel von Grundgütern, die sicher von keinem Menschen als unerträglich empfunden werden.

Es ist klar, daß Rawls davon ausgeht, daß wir die erste Kategorie von allgemeinem Wissen besitzen, und daß auch den Parteien als allgemeine Tatsache gewisse Güterbündel bekannt sind, die für bestimmte Menschen nicht akzeptabel sind. Dies gilt in jedem Fall für Bündel, die nicht ein „völlig adäquatem System gleicher Grundfreiheiten“ beinhalten, indem sie z. B. die Gewissensfreiheit und die Unverletzlichkeit der Person ausschließen. Daß solche Bündel für Menschen inakzeptabel sind und dies den Parteien im Urzustand bekannt ist, ist zentraler Baustein des Rawlsschen Arguments für seine zwei Prinzipien und insbesondere den Vorrang der Freiheit.¹⁸ [Vgl. Rawls

¹⁶ Siehe unten Axiom 6.8.

¹⁷ Wir werden allerdings im allgemeinen annehmen, daß es Situation *gibt*, von denen bekannt ist, daß sie unerträglich sind. Vgl. unten S. 163 f.

¹⁸ Es ist nicht ganz klar, inwieweit Rawls auch für bestimmte „Mengen“ oder Mengenmischungen anderer Grundgüter davon ausgeht, daß es als allgemeine Tatsache gelten kann, daß sie für Menschen inakzeptabel sind. Mir ist in jedem Fall keine Stelle bekannt, wo Rawls dazu *eindeutig* Stellung nimmt.

Immerhin setzt er in der revidierten Fassung seines Arguments für das Differenzprinzip voraus, daß bezogen auch auf andere Grundgüter die Lasten der eingegangenen Verpflichtungen („strains of commitment“) so groß sein können, daß Bürger, denen das Minimum an Güter fehlt, um ein be-

1975, 176 u. 202 f.; Rawls 1992b, 181-183]

Anders verhält es sich mit Bündeln von Grundgütern, die sicher *nicht* als unerträglich empfunden werden. Meines Wissens nimmt Rawls nirgends Bezug auf solche Kombinationen von Grundgütern, sie scheinen in seiner Argumentation einfach keine Rolle zu spielen. Wenn man den geschichtlichen und gesellschaftlichen Charakter menschlicher Bedürfnisse beachtet, dann ist vielleicht sogar zu begründen, daß es ein solches allgemeines Wissen gar nicht geben kann. Was Menschen für akzeptabel ansehen scheint in hohem Maße davon bestimmt zu sein, in welcher sozialen Welt sie leben. [Vgl. Rawls 2001, 128 f.] Da wir aber die soziale Welt der jedenfalls fernerer Zukunft nicht kennen, scheint es vermessen, heute menschlichen Bedürfnissen Grenzen aufzuerlegen. Da den Parteien im Urzustand auch das Wissen um die geschichtliche Periode, in der sie sich wiederfinden, verwehrt ist [Rawls 1973, 137], können sie nicht Grenzen der Bedürfnisse, die heute bestehen, als gegeben voraussetzen. Als allgemeine, den Parteien zugängliche Tatsache kann nur gelten, was nach heutigem, anerkannten Wissensstand als anthropologische Konstante angesehen werden kann, weshalb wir Wissen um sicher akzeptable Bündel von Grundgütern nicht weiter betrachten werden.

Sei nun B die Menge aller Güterbündel x , für die aufgrund allgemeiner Tatsachen feststeht, daß es Individuen gibt, die diese für unerträglich halten. (B kann auch die leere Menge sein.) Da auch diese Individuen dem allgemeinen Typ entsprechen, den wir im vorangehenden Abschnitt modelliert haben, können wir aufgrund Theorem 6.2 und der Monotonie der Präferenzen der Individuen – an der wir festhalten – als Axiom einführen:

Axiom 6.7.

$$\forall x, x' : x \in B \wedge x \geq x' \rightarrow x' \in B$$

Nach unseren bisherigen Überlegungen ist das Wissen um Elemente von B zusammen-

scheidenes („decent“) Leben führen zu können und diejenigen Bedürfnisse zu erfüllen, die in einer Gesellschaft als wesentlich angesehen werden, die Gerechtigkeitskonzeption ihrer Gesellschaft ablehnen, sich selbst als unterdrückt ansehen und sogar zu Gewalt greifen werden, um ihrer Situation zu verbessern. [Rawls 2001, 128 f.] Die Sicherung eines entsprechenden sozialen Minimums soll deshalb sogar Verfassungsrang gewinnen, obwohl seine Gewährung auf der Ebenen der Prinzipien der Gerechtigkeit dem Differenzprinzip zugeordnet bleibt, das jedoch noch weitergehende Anforderungen stellt. [Rawls 2001, 47 f., 162]

Weitergehend erwähnt Rawls [2003, 80] in einer Anmerkung noch die *Möglichkeit* seinem ersten Prinzip ein weiteres Prinzip lexikalisch vorzuordnen, „das die Erfüllung der Grundbedürfnisse verlangt, zumindest insofern ihre Erfüllung eine notwendig Bedingung dafür ist, daß die Bürger dazu in der Lage sind, die Grundrechte und -freiheiten zu verstehen und fruchtbar wahrzunehmen“, ohne sich allerdings diese Möglichkeit direkt zu eigen zu machen.

Insofern als eines seiner Argumente für den Vorrang der Freiheit im Wesentlichen einfach darauf beruht, daß die Parteien wissen, daß etwa fehlende Gewissensfreiheit für bestimmte Menschen inakzeptabel ist, wäre seine Argumentation tatsächlich gefährdet, wenn er explizit etwa auch ein bestimmtes Niveau von Einkommen als sicher für bestimmte Menschen in der Weise als unerträglich anerkennen würde, wie dies das Fehlen bestimmter Grundfreiheiten sein kann. Ein sorgfältiges Argument dafür, daß ein Ernst nehmen der Gefahr, grundlegende sozioökonomische Notwendigkeiten nicht erfüllt zu bekommen (in der Sprache meiner Rekonstruktion: der Gefahr einer unerträglich sozioökonomischen Situation ausgesetzt zu sein), dazu führt, der Erfüllung dieser Grundbedürfnisse gleichen Vorrang wie der Gewährung der Grundfreiheiten zu geben, findet sich in Pogge [1989, 134-148].

men mit der Bedingung moderater Knappheit der einzige Grund, dessentwegen die Parteien die Menge möglicher Mengen von Alternativen einschränken können. Ihre Entscheidungsregel (das gesuchte Prinzip der Gerechtigkeit) muß also für alle $S \in \Sigma$ mit $\exists a \in S : \forall i : a_i \notin B$ gelten. Wir definieren:

Definition 6.3.

$$\Sigma_B := \{S \in \Sigma : \exists a \in S : \forall i : a_i \notin B\}$$

Wir nehmen nun an, daß die Situation („Welt“) einer Partei nach Auflösung der Unsicherheit hinreichend durch einen Vektor $w = (S, a, r, \rho, i^*, j^*)$ beschrieben werden kann, wobei $S \in \Sigma_B$, die Menge der Möglichkeiten ist, aus denen die Partei faktisch wählen konnte, $a \in C(S)$ der tatsächlich verwirklichte Vektor von möglichen Bündeln von Grundgütern, d. h. der Vektor der in der Gesellschaft vorhandenen gesellschaftlichen Positionen, r der Vektor der in der Gesellschaft vorhandenen Präferenzordnungen r_i , ρ der Vektor der in der Gesellschaft vorhandenen Mengen unerträglicher Situationen ρ_j , $i^* \in N$ der Index der Position, die das Individuum dann ex post faktisch einnimmt, und schließlich j^* der Index der Präferenzordnung und gleichzeitig der Menge von unerträglichen Situationen, die die Partei charakterisieren.

Wir beschreiben im folgenden die Menge möglicher Welten $W(S, a)$ mit denen die Parteien – gegeben eine bestimmte Entscheidungssituation S und dem realisierten Element a der Entscheidungsmenge $C(S)$ – rechnen müssen:

Axiom 6.8.

$$\forall S \in \Sigma_B, a \in S : W(S, a) = \{(S, a, r, \rho, i^*, j^*) : (\forall i \in N : r_i \text{ ist eine Ordnung über } X \wedge (\forall x, x' \in X : x \geq x' \rightarrow xr_i x'))\} \quad (6.1)$$

$$\wedge (\forall i \in N, x, x' \in X : x \in \rho_i \wedge xr_i x' \rightarrow x' \in \rho_i) \quad (6.2)$$

$$\wedge (i^*, j^* \in N) \quad (6.3)$$

$$\wedge (\exists b \in S : \forall i, j \in N : b_i \notin \rho_j) \quad (6.4)$$

$$\wedge (\forall x \in B : \exists j \in N : x \in \rho_j) \quad (6.5)$$

Zu den einzelnen Teilen des Axioms:

- Wie zuvor nehmen wir also an, daß die Präferenzen der Parteien monoton aber sonst keinen Beschränkungen unterworfen sein sollen. (Axiom 6.8-6.1)
- Da wir annehmen, daß die es sich bei den Parteien um Parteien von dem Typ handelt, den wir im vorhergehenden Abschnitt behandelt haben, mit dem einzigen Unterschied, daß die Präferenzen und die Menge der für die Partei unerträglichen Situationen unbekannt sind, können wir das Ergebnis des Theorems 6.2 (das ja für alle möglichen Parteien dieses Typs gilt) als Voraussetzung übernehmen. Die Parteien kennen zwar ihre Präferenzen und die für sie unerträglichen Situationen nicht, aber es wird in jedem Fall der in Axiom 6.8-6.2 beschriebene Zusammenhang bestehen.

6 Der Urzustand: Ein Rekonstruktion

- Wenn aufgrund des beschränkten Wissens der Parteien diese – wie oben ausgeführt – die Menge der Mengen von Alternativen nicht weiter einschränken können, aber trotzdem versichert sein sollen, daß in der Welt, für die sie entscheiden nur moderate Knappheit herrscht, d. h. sie sicher sein können, daß es für alle S , denen sie sich gegenüberfinden können $a \in S$ gibt, so daß alle Positionen für alle möglichen Mengen von unerträglichen Situationen erträglich sind, dann muß diese Bedingung den möglichen Mengen von unerträglichen Situationen *von der jeweiligen Entscheidungssituation abhängige Beschränkungen* auferlegen, wenn Unsicherheit über diese Mengen erhalten bleiben soll. (Axiom 6.8-6.4)
- Bedingung 6.5 von Axiom 6.8 hält die Beschränkung auf $S \in \Sigma_B$ und unser Wissen um die Existenz von Individuen, für die Elemente von B inakzeptabel sind, axiomatisch fest.

Die Parteien „werden“ zu Individuen vom in Abschnitt 6.3 betrachteten Typ, die den Glauben an ihre eigene Rationalität nicht mehr aufrecht erhalten können, wenn sie sich a) ex post in einer unerträglichen Position wiederfinden, und ihnen b) Alternativen zur Wahl standen, mit denen sie hätten *sicher ausschließen* können, sich in einer solchen Position wiederzufinden. *Sicher ausschließen*, sich in einer unerträglichen Position wiederzufinden, können die Parteien unter den Voraussetzungen, die wir bis jetzt gemacht haben, genau dann, wenn es in der Entscheidungssituation S Vektoren b gibt, deren Elemente in keiner der möglichen Mengen von unerträglichen Situationen enthalten sein können. Formal:

Axiom 6.9.

$$\begin{aligned} \forall S \in \Sigma_B, a \in S, w \in W(S, a) : a_{i^*} \in \rho_{j^*} \rightarrow \\ [a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j] \rightarrow \\ G_w(\text{irr})] \end{aligned}$$

Wieder gilt: Wenn ein Individuum in einer ex post möglichen Welt etwas für wahr hält, was falsch ist, dann ist es irrational:

Axiom 6.10.

$$\forall S \in \Sigma_B, a \in S, w \in W(S, a) : G_w(p) \wedge \neg p \rightarrow \text{irr}$$

Und die Parteien sind rational:

Axiom 6.11.

$$\neg \text{irr}$$

Wir zeigen:

Theorem 6.3. Gegeben die Definitionen 6.2 und 6.3 und die Axiome 6.9, 6.10 und 6.11 gilt:

$$\begin{aligned} \forall S \in \Sigma_B, a \in S, w \in W(S, a) : a_{i^*} \in \rho_{j^*} \rightarrow \\ \neg[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]] \end{aligned}$$

Beweis. Betrachte S und $w \in W(S, a)$, so daß $a_{i^*} \in \rho_{j^*}$. Dann gilt wegen Axiom 6.9:

$$[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]] \rightarrow G_w(\text{irr})$$

Nimm an, daß für das gegebene w :

$$a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]$$

Damit würde gelten $G_w(\text{irr})$ und aufgrund Axiom 6.10 irr , in Widerspruch zu Axiom 6.11. Es muß also gelten:

$$\neg[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]]$$

□

Wir führen noch folgende Definitionen ein:

Zunächst definieren wir inspiriert durch Theorem 6.3 Rawls-Rationalität bzw. das Prädikat rrat , für „die Parteien sind rawlsrational“. Diese Definition wird im folgenden Zielpunkt sowohl unserer *positiven* Bemühungen sein, eine Entscheidungsregel als akzeptabel auszuweisen, wie auch unserer *negativen* Bemühungen, eine Entscheidungsregel als mit der Rationalität der Parteien unvereinbar auszuweisen. Wir haben keine Definition von Rationalität im Allgemeinen vorgelegt, sondern nur eine notwendige Bedingung: Eine Entscheidungsregel, die zur Rawls-Rationalität der Parteien führt, erfüllt diese in Theorem 6.3 vorliegende notwendige Bedingung, eine Entscheidungsregel, die mit der Rawls-Rationalität der Parteien unvereinbar ist, ist auch mit der (angenommenen) Rationalität der Parteien unvereinbar.

Definition 6.4. *rawlsrational*

$$\text{rrat} \leftrightarrow \forall S \in \Sigma_B, a \in S, w \in W(S, a) : a_{i^*} \in \rho_{j^*} \rightarrow \neg[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]]$$

Zwei weitere Definitionen, die wir im folgenden brauchen werden:

Definition 6.5. *Bessermenge*

$$\forall a \in X^n : B(a) := \{x : \exists i \in N : x \geq a_i\}$$

Definition 6.6. *Minimale Selektion*

$$\forall a \in X^n : MS(a) := \{x : \exists i \in N : x = a_i \wedge \neg \exists i \in N : x > a_i\}$$

Theorem 6.4. *Gegeben die Definitionen 6.2, 6.3 und 6.5 und die Axiome 6.7 und 6.8, gilt:*

$$\forall S \in \Sigma_B : \forall b \in S : [\forall w \in W(S, b), i, j \in N : b_i \notin \rho_j \leftrightarrow [\forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a)]]$$

6 Der Urzustand: Ein Rekonstruktion

Beweis. 1. Wir beweisen zunächst die Implikation von rechts nach links. Wir nehmen also für beliebige S und $b \in S$ an:

$$[\forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a)]$$

Entsprechend der Definition der Bessermengen ist $b_i \in B(b)$ für alle $i \in N$ und damit auch für alle $a \in \{a \in S : \forall i : a_i \notin B\} : b_i \in B(a)$. Nimm an: $\exists w \in W(S), i, j \in N : b_i \in \rho_j$. Betrachte entsprechende w, i und j . Gemäß der Bedingungen 6.4 und 6.5 von Axiom 6.8 gibt es zumindest ein $a \in \{a \in S : \forall i : a_i \notin B\}$, so daß $\neg \exists i' \in N : a_{i'} \in \rho_j$. Wegen $b_i \in B(a)$ gibt es aber auch für dieses $a \in \{a \in S : \forall i : a_i \notin B\}$ ein $i' \in N$, so daß $b_i \geq a_{i'}$ und aufgrund Axiom 6.8-6.1 und 6.8-6.2 $a_{i'} \in \rho_j$. Ein Widerspruch, d. h. $\neg \exists w \in W(S, b), i, j \in N : b_i \in \rho_j$.

2. Wir beweisen jetzt die Implikation von links nach rechts. Wir nehmen also für beliebige S und $b \in S$ an:

$$\forall w \in W(S, b), i, j \in N : b_i \notin \rho_j$$

Nimm an, es gäbe $a \in \{a \in S : \forall i : a_i \notin B\}$, so daß $\neg B(b) \subseteq B(a)$, d.h. auch $\exists i : b_i \notin B(a)$. Für ein solches a konstruieren wir jetzt ein $w \in W(S, b)$, so daß $\exists j \exists i \in N : b_i \in \rho_j$, ein Widerspruch zu unserer Annahme, womit die Implikation dann gezeigt ist.

Definiere für 0 und mit dem Einheitsvektor e für alle $u \in \mathbb{R} : u \neq 0$:¹⁹

$$U(0) = \{x : x \in B(a) \wedge \neg \exists i \in N : x \gg a_i\}$$

$$U(u) = \{x : \exists x' \in U(0) : x = u \cdot e + x'\}$$

Wir definieren eine Präferenzordnung R und eine Menge q :

$$xRx' :\leftrightarrow \exists u, u' : x \in U(u) \wedge x' \in U(u') \wedge u \geq u'$$

$$q = \{x \in X : \exists u < 0 : x \in U(u)\}$$

Für $r = (R, R, \dots, R)$ und $\rho = (q, q, \dots, q)$ zeigen wir, daß für alle $i^*, j^* \in N$ gilt $w = (S, a, r, \rho, i^*, j^*) \in W(S, b)$ und außerdem $\exists j \exists i \in N : b_i \in \rho_j$.

a) Wir zeigen zunächst, daß R eine monotone Ordnung über X ist und deshalb für $r = (R, R, \dots, R)$ Axiom 6.8-6.1 erfüllt ist:

R ist *transitiv*, was direkt und trivialerweise aus der Transitivität der Größer-Gleich-Relation folgt.

¹⁹ Dabei gelte: $x \gg y \leftrightarrow \forall j : x_j > y_j$.

R ist *vollständig*, da auf \mathbb{R} die Größer-Gleich-Relation vollständig ist und für jedes $x \in X$ genau ein u existiert, so daß $x \in U(u)$: Für jedes $i \in N$ bestimme u_i als das minimale $u \in \mathbb{R}$, so daß $\exists j \in M : x_j = u + a_{ij}$. Ein solches u_i existiert, da $x_j = u + a_{ij}$ jeweils genau eine Lösung hat und M endlich ist. Es gilt dann für alle i : $x - u_i \cdot e \geq a_i \wedge \neg x - u_i \cdot e \gg a_i$. Für alle i gilt offensichtlich: $x - u_i \cdot e \in B(a)$. Wähle (wieder aus einer endlichen Menge) das maximale u_i als u . Es muß dann weiter gelten $\forall i : \neg x - u \cdot e \gg a_i$, d. h. zusammengenommen $x - u \cdot e \in U(0)$ und $x \in U(u)$.

R ist *monoton*: Sei $x \geq x'$. Nimm an, $\neg xRx'$. Da für jedes $x \in X$ ein u existiert, so daß $x \in U(u)$, müßte es u und u' geben, für die gilt: $u' > u$, $x \in U(u)$ und $x' \in U(u')$, d. h. $\exists x'' \in U(0) : x = u \cdot e + x''$ und $\exists x''' \in U(0) : x' = u' \cdot e + x'''$. Außerdem $x''' = x' - u' \cdot e \ll x' - u \cdot e \leq x - u \cdot e = x''$, d. h. $x'' \gg x'''$. Das ist unmöglich, da wegen $x''' \in U(0)$ und $x'' \in U(0)$ gleichzeitig auch gelten müßte $x''' \in B(a)$, d. h. $\exists i : x''' \geq a_i$, und $\neg \exists i : x'' \gg a_i$, ein Widerspruch wegen der Transitivität von \gg und \geq .

- b) Betrachte nun $\rho = (q, q, \dots, q)$. Offensichtlich ist wegen der Definition von q mit $r = (R, R, \dots, R)$ zunächst Axiom 6.8-6.2 erfüllt.

Weiterhin gilt: $\forall x \in B(a) : x \in U(u) \rightarrow u \geq 0$. Nimm nämlich an $x \in B(a)$, d. h. $\exists i : x \geq a_i$, $x \in U(u)$, und $u < 0$. Es müßte dann ein $x' \in B(a)$ geben, so daß $\neg \exists i : x' \gg a_i$ und $x = u \cdot e + x'$, d. h. $x' \gg x$, ein Widerspruch, weil dann auch gelten müßte $\exists i : x' \gg a_i$.

Für alle i gilt also $a_i \notin q$, womit es ein $a \in S$ gibt, so daß $\forall i : a_i \notin q$. Damit ist für ρ auch 6.8-6.4 erfüllt.

Ebenso gilt: $\forall x \notin B(a) : x \in U(u) \rightarrow u < 0$. Nimm nämlich an $x \notin B(a)$, d. h. $\neg \exists i : x \geq a_i$, $x \in U(u)$ und $u \geq 0$. Es müßte dann ein $x' \in B(a)$ geben, so daß $x = u \cdot e + x'$, d. h. $\exists i : x' \geq a_i$ und $x \geq x'$, ein Widerspruch, weil dann auch gelten müßte $\exists i : x \geq a_i$.

Zum einen ergibt sich daraus $\forall x \in B : \exists j \in N : x \in \rho_j$, womit auch 6.8-6.5 erfüllt ist. Annahmegemäß ist nämlich $a \in \{a \in S : \forall i : a_i \notin B\}$, wegen Axiom 6.7 gilt also $\forall x \in B, i \in N : a_i > x$, bzw. $\forall x : x \in B \rightarrow x \notin B(a)$. Wie gerade gezeigt muß dann auch gelten $\forall x \in B : x \in U(u) \rightarrow u < 0$, bzw. $x \in q$.

Außerdem folgt $\exists i : b_i \in q$, und es ergibt sich ein Widerspruch zur ursprünglichen Annahme über b .

□

Theorem 6.5. Gegeben Definitionen 6.2, 6.3 und 6.5 und die Axiome 6.7 und 6.8, gilt:

$$\begin{aligned} \forall S \in \Sigma_B : [\forall a \in S, w \in W(S, a) : a_{i^*} \in \rho_{j^*} \rightarrow \\ \neg[a \in C(S) \wedge \exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]] \leftrightarrow \\ [\exists b \in S : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a)] \rightarrow \\ \forall c \in C(S) : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(c) \subseteq B(a) \end{aligned}$$

6 Der Urzustand: Ein Rekonstruktion

Beweis. 1. Wir beginnen mit der Implikation von rechts nach links und betrachten ein beliebiges S . Für den Fall, daß $\neg \exists b \in S : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a)$ gilt wegen Theorem 6.4 auch

$$\neg \exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j$$

Antezedens und Konklusion sind also wahr. Bleibt der Fall, daß $\exists b \in S : \forall c \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(c)$. Gemäß dem Antezedens gilt dann $\forall a \in C(S) : \forall c \in \{a \in S : \forall i : a_i \notin B\} : B(a) \subseteq B(c)$ und wegen Theorem 6.4 auch

$$\forall a \in C(S) : \forall w' \in W(S, a), i, j \in N : a_i \notin \rho_j$$

Es kann also kein $a \in S$ geben, so daß $a \in C(S)$ und für eine Kombination von i und j : $a_i \in \rho_j$. Da es dann auch für $a \in C(S)$ keine i^* und j^* geben kann mit $a_{i^*} \in \rho_{j^*}$ ist auch damit wieder die Konklusion der betrachteten Implikation wahr.

2. Wir beweisen die Implikation von links nach rechts und betrachten dazu wieder zwei Unterfälle.

a) Es gelte:

$$\neg \exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j$$

Das Antezedens der betrachteten Implikation ist damit wahr. Wegen Theorem 6.4 gilt dann auch $\neg \exists b \in S : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a)$. Auch die Konklusion der betrachteten Implikation ist also wahr.

b) Es gelte:

$$\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j$$

Das Antezedens der betrachteten Implikation ist also falsch, wenn $\exists a, i, j : a_i \in \rho_j \wedge a \in C(S)$, da es sonst auch entsprechende i^* und j^* gäbe. Die Konklusion der betrachteten Implikation kann nur dann falsch sein, wenn $\neg \forall c \in C(S) : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(c) \subseteq B(a)$ und damit wegen Theorem 6.4 auch

$$\neg \forall c \in C(S) : \forall w' \in W(S, c), i, j \in N : c_i \notin \rho_j$$

bzw.

$$\exists c \in C(S) : \exists w' \in W(S, c), i, j \in N : c_i \in \rho_j$$

d. h. auch:

$$\exists a, i, j : a_i \in \rho_j \wedge a \in C(S)$$

Das Antezedens ist also falsch, die Implikation deshalb wahr. □

Theorem 6.6. Sei \succeq eine beliebige Ordnung über X^n und $C(S)$ für alle S gleich $\{a \in \{a \in S : \forall i : a_i \notin B\} : \forall b \in \{a \in S : \forall i : a_i \notin B\} : a \succeq b\}$. Gegeben die Definition 6.5 und Axiom 6.7 gilt dann:

$$\begin{aligned} [\forall a, b \in S : B(a) \subseteq B(b) \rightarrow a \succ b] \rightarrow \\ [\exists b \in S : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a) \rightarrow \\ \forall c \in C(S) : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(c) \subseteq B(a)] \quad (6.6) \end{aligned}$$

Beweis. Nimm an $\exists b \in S : \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b) \subseteq B(a)$. Auch für ein solches b muß dann wegen Axiom 6.7 und der Definition der Bessermenge gelten $b \in \{a \in S : \forall i : a_i \notin B\}$. Betrachte ein solches b und ein $c \in C(S)$, so daß $\neg \forall a \in \{a \in S : \forall i : a_i \notin B\} : B(c) \subseteq B(a)$. Dann wäre $B(c) \not\subseteq B(b)$, bzw., da $c \in \{a \in S : \forall i : a_i \notin B\}$, $B(b) \subset B(c)$ und wegen der Geltung des Antezedens $b \succ c$, ein Widerspruch aufgrund der Definition von $C(S)$. \square

Lemma 6.1. Gegeben die Definitionen 6.5 und 6.6 gilt:

$$\forall a \in X^n, x \in B(a) : \exists x' \in MS(a) : x \geq x'$$

Beweis. Betrachte ein x (im folgenden x^*), so daß $x \in B(a) = \{x : \exists i \in N : x \geq a_i\}$. Es folgt unmittelbar, daß $\{x : x^* \geq x \wedge \exists i : x = a_i\}$ nicht leer und endlich ist. Damit existieren bezüglich der Teilordnung \geq minimale Elemente dieser Menge und es gilt also $\exists x' : x^* \geq x' \wedge (\exists i : x' = a_i) \wedge \neg \exists x : (x^* \geq x \wedge \exists i \in N : x = a_i \wedge x' > a_i)$, bzw. $\exists x' : x^* \geq x' \wedge (\exists i : x' = a_i) \wedge \neg \exists i : (x^* \geq a_i \wedge x' > a_i)$. Nimm an $x' \notin MS(a) = \{x : \exists i \in N : x = a_i \wedge \neg \exists i \in N : x > a_i\}$, d. h. $\exists i : x' > a_i$. Wegen der Transitivität von \geq und $>$ gilt für ein solches a_i dann auch $x^* \geq a_i$, d. h. für x' gilt $\exists i : (x^* \geq a_i \wedge x' > a_i)$, ein Widerspruch. Es muß also gelten $x' \in MS(a)$ und damit $\exists x' \in MS(a) : x^* \geq x'$. \square

Theorem 6.7. Gegeben die Definitionen 6.5 und 6.6 gilt:

$$\forall a, b \in X^n : B(a) \subseteq B(b) \leftrightarrow \forall x \in MS(a) : \exists y \in MS(b) : x \geq y$$

Beweis. 1. Wir zeigen zunächst die Implikation von rechts nach links. Es gelte also für irgendwelche a und b : $\forall x \in MS(a) : \exists y \in MS(b) : x \geq y$. Betrachte ein $x \in B(a)$. Es gibt dann gemäß Lemma 6.1 ein $x' \in MS(a)$, so daß $x \geq x'$. Annahmegemäß gibt es dann ein $y \in MS(b)$, so daß $x' \geq y$ und aufgrund der Transitivität von \geq auch $x \geq y$. Es gibt also ein i , so daß $x \geq y = b_i$ bzw. $x \geq b_i$. x ist also entsprechend der Definition von $B(b)$ Element von $B(b)$.

2. Wir zeigen nun die Implikation von links nach rechts. Es gelte also für irgendwelche a und b : $B(a) \subseteq B(b)$. Betrachte ein $x \in MS(a) = \{x : \exists i \in N : x = a_i \wedge \neg \exists i \in N : x > a_i\}$. x ist dann auch Element von $B(a)$ und deshalb annahmegemäß von $B(b)$. Wegen Lemma 6.1 gilt dann aber auch $\exists y \in MS(b) : x \geq y$. \square

6 Der Urzustand: Ein Rekonstruktion

Wir definieren jetzt eine Ordnung \succ_α , die wir „strikte beschränkte Leximin-Ordnung für α “ nennen wollen.²⁰

Sei vorbereitend $\alpha \in \mathbb{R}_+^n$, d. h. $\forall j \in M : \alpha_j > 0$. Für jedes $a \in X^n$ sei g_a eine eindeutige Funktion von $MS(a)$ auf $N_a = \{1, \dots, \#MS(a)\}$, so daß $\forall x, x' \in MS(a) : \sum_j \alpha_j x_j > \sum_j \alpha_j x'_j \rightarrow g_a(x) > g_a(x')$. Darauf aufbauend definieren wir:

Definition 6.7. *strikte beschränkte Leximin-Ordnung für α*

$$\begin{aligned} \forall a, b \in X^n : a \succ_\alpha b \leftrightarrow_{def} \exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : \\ [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \\ \wedge [[g_a(x') = g_b(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j] \\ \vee [g_a(x) + 1 > \#MS(a) \wedge \#MS(b) > g_a(x)]] \end{aligned}$$

Auf der Basis der eben definierten strikten Leximin-Ordnung für α , definieren wir jetzt die beschränkte Leximin-Indifferenz für α :

Definition 6.8. *Beschränkte Leximin-Indifferenz für α*

$$\forall a, b \in X^n : a \sim_\alpha b \leftrightarrow_{def} \neg(a \succ_\alpha b) \wedge \neg(b \succ_\alpha a)$$

Zuletzt definieren wir noch die schwache, beschränkte Leximin-Ordnung für α :

Definition 6.9.

$$\forall a, b \in X^n : a \succeq_\alpha b \leftrightarrow_{def} a \succ_\alpha b \vee a \sim_\alpha b$$

Die so definierte schwache, beschränkte Leximin-Ordnung für α ist trivialerweise vollständig auf X^n . Um sicherzustellen, daß es sich bei \succeq_α bzw. \succ_α in der Tat um Ordnungen handelt, bleibt noch die Transitivität von \succeq_α zu zeigen. Den Beweis bereiten wir durch folgendes Lemma vor.

Lemma 6.2. *Gegeben die Definitionen 6.6, 6.7 und 6.8 gilt:*

$$\forall a, b \in X^n : a \sim_\alpha b \leftrightarrow \#MS(a) = \#MS(b) \wedge \forall x, y : g_a(x) = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$$

Beweis. Die Implikation von rechts nach links ergibt sich direkt aus den Definitionen von \succ_α und \sim_α .

Die Implikation von links nach rechts beweisen wir mittels vollständiger Induktion. Sei $a \sim_\alpha b$.

²⁰ Die beschränkte Leximinordnung hat mit der Leximinordnung gemein, daß ihre formale Darstellung komplizierter aussieht, als eine verbale Beschreibung vermuten läßt. Wenn wir die Idee einer Leximinordnung als bekannt voraussetzen, können wir einfach sagen: Bei der beschränkten Leximinordnung handelt es sich um die Ordnung von Vektoren entsprechend der Leximinordnung ihrer minimalen Elemente.

1. Da $MS(a)$ und $MS(b)$ nicht leer sind, und g_a und g_b eindeutige Abbildungen von $MS(a)$ bzw. $MS(b)$ in N_a bzw. N_b sind, gibt es x und y , sodaß $g_a(x) = 1$ und $g_b(x) = 1$. Sei $\sum_j \alpha_j x_j > \sum_j \alpha_j y_j$. Dann wäre gemäß der Definition von \succ_α auch $a \succ_\alpha b$, ein Widerspruch zur Ausgangsannahme. Entsprechend für $\sum_j \alpha_j y_j > \sum_j \alpha_j x_j$. Es muß also gelten $\sum_j \alpha_j x_j = \sum_j \alpha_j y_j$.
2. Betrachte x , so daß $\forall x', y : g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$. Sei $\#MS(a) = g_a(x)$. Nimm an $\neg \#MS(b) = g_a(x)$. Da wegen der Induktionsannahme $\#MS(b) \geq g_a(x)$, wäre dann in jedem Fall gemäß Definition 6.7 $a \succ_\alpha b$ im Widerspruch zur Ausgangsannahme. Es gilt also $\#MS(a) = g_a(x) \rightarrow \#MS(b) = g_a(x)$ und wie sich auf gleiche Weise zeigen läßt $\#MS(b) = g_a(x) \rightarrow \#MS(a) = g_a(x)$. Sei $\#MS(a) > g_a(x)$. Dann muß, wie gerade gezeigt, auch gelten $\#MS(b) > g_a(x)$, d. h. $\exists x', y : g_a(x') = g_b(y) = g_a(x) + 1$. Aus $\neg \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ folgt entweder $a \succ_\alpha b$ oder $b \succ_\alpha a$ im Widerspruch zur Ausgangsannahme. Also $\sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und aufgrund vollständiger Induktion $\forall x, y : g_a(x) = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$. □

Theorem 6.8. Transitivität von \succeq_α

Gegeben die Definitionen 6.6, 6.7, 6.8 und 6.9 gilt:

$$\forall a, b \in X^n : a \succeq_\alpha b \wedge b \succeq_\alpha c \rightarrow a \succeq_\alpha c$$

Beweis. Gegeben $a \succeq_\alpha b \wedge b \succeq_\alpha c$ können wir vier mögliche Fälle unterscheiden, und wir zeigen für jeden Fall, daß $a \succeq_\alpha c$:

1. Sei $a \succ_\alpha b$ und $b \succ_\alpha c$. Es gibt vier mögliche Unterfälle, und wir zeigen für jeden Unterfall $a \succeq_\alpha c$:
 - a) Es gelte $\exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge [g_a(x') = g_b(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j]$ (sei x_a das entsprechende $x \in MS(a)$) und $\exists x \in MS(b) : \forall x' \in MS(b), y \in MS(c) : [g_b(x) \geq g_b(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge [g_b(x') = g_c(y) = g_b(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j]$ (sei x_b das entsprechende $x \in MS(b)$). Sei $g_a(x_a) \geq g_b(x_b)$. Es gibt dann wegen der Transitivität von $=$ und $>$ und wegen $x = y > z \rightarrow x > z$ ein $x \in M(a)$, nämlich $g_a^{-1}(g_b(x_b))$, so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x') = g_c(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$. Sei $g_a(x_a) < g_b(x_b)$. Es gibt dann wegen der Transitivität von $=$ und wegen $x > y = z \rightarrow x > z$ ein x , nämlich x_a , so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x') = g_c(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.
 - b) Es gelte $\exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge [g_a(x') = g_b(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j]$ (sei x_a das entsprechende $x \in MS(a)$) und $\exists x \in MS(b) : \forall x' \in MS(b), y \in MS(c) :$

$[g_b(x) \geq g_b(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge g_b(x) + 1 > \#MS(b) \wedge \#MS(c) > g_c(x)$ (sei x_b das entsprechende $x \in MS(b)$). Es muß dann gelten $g_a(x_a) < g_b(x_b)$. Es gibt dann wegen der Transitivität von $=$ und wegen $x > y = z \rightarrow x > z$ ein x , nämlich x_a , so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x') = g_c(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.

c) Es gelte $\exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge g_a(x) + 1 > \#MS(a) \wedge \#MS(b) > g_a(x)$ (sei x_a das entsprechende $x \in MS(a)$) und $\exists x \in MS(b) : \forall x' \in MS(b), y \in MS(c) : [g_b(x) \geq g_b(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge [g_b(x') = g_c(y) = g_b(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j]$ (sei x_b das entsprechende $x \in MS(b)$). Sei $g_a(x_a) > g_b(x_b)$. Es gibt dann wegen der Transitivität von $=$ und wegen $x = y > z \rightarrow x > z$ ein x , nämlich $g_a^{-1}(g_b(x_b))$, so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x') = g_c(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$. Sei $g_a(x_a) \leq g_b(x_b)$. Da $\#MS(c) > g_a(x_a)$ und wegen der Transitivität von $=$ gibt es dann ein x , nämlich x_a , so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x) + 1 > \#MS(a) \wedge \#MS(c) > g_a(x)$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.

d) Es gelte $\exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge g_a(x) + 1 > \#MS(a) \wedge \#MS(b) > g_a(x)$ (sei x_a das entsprechende $x \in MS(a)$) und $\exists x \in MS(b) : \forall x' \in MS(b), y \in MS(c) : [g_b(x) \geq g_b(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge g_b(x) + 1 > \#MS(b) \wedge \#MS(c) > g_b(x)$ (sei x_b das entsprechende $x \in MS(b)$). Voraussetzungsgemäß gilt $\#MS(c) > g_b(x) = \#MS(b) > g_a(x)$ bzw. $\#MS(c) > g_a(x)$, und wegen der Transitivität von $=$ gibt es also ein x , nämlich x_a , so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x) + 1 > \#MS(a) \wedge \#MS(c) > g_a(x)$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.

2. Sei $a \succ_\alpha b$ und $b \sim_\alpha c$. Es gibt zwei mögliche Unterfälle, und wir zeigen für jeden Unterfall $a \succeq_\alpha c$:

a) Es gelte $\exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge [g_a(x') = g_b(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j]$ (sei x_a das entsprechende $x \in MS(a)$) und entsprechend Lemma 6.2 $\#MS(b) = \#MS(c) \wedge \forall x, y : g_b(x) = g_c(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$. Es muß dann gelten $\#MS(c) > g_a(x_a)$ und wegen der Transitivität von $=$ und wegen $x > y = z \rightarrow x > z$ gibt es ein x , nämlich x_a , so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x') = g_c(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.

b) Es gelte $\exists x \in MS(a) : \forall x' \in MS(a), y \in MS(b) : [g_a(x) \geq g_a(x') = g_b(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge g_a(x) + 1 > \#MS(a) \wedge \#MS(b) > g_a(x)$ (sei x_a das ent-

sprechende $x \in MS(a)$ und entsprechend Lemma 6.2 $\#MS(b) = \#MS(c) \wedge \forall x, y : g_b(x) = g_c(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$. Wegen $\#MS(c) = \#MS(b) > g_a(x)$ bzw. $\#MS(c) > g_a(x)$ und wegen der Transitivität von $=$ gibt es ein x , nämlich x_a , so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x) + 1 > \#MS(a) \wedge \#MS(c) > g_a(x)$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.

3. Sei $a \sim_\alpha b$ und $b \succ_\alpha c$. Es gibt zwei mögliche Unterfälle, und wir zeigen für jeden Unterfall $a \succeq_\alpha c$:

- a) Entsprechend Lemma 6.2 gilt $\#MS(a) = \#MS(b) \wedge \forall x, y : g_a(x) = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$ und es gelte $\exists x \in MS(b) : \forall x' \in MS(b), y \in MS(c) : [g_b(x) \geq g_b(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge [g_b(x') = g_c(y) = g_b(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j]$ (sei x_b das entsprechende $x \in MS(b)$). Es gibt dann wegen der Transitivität von $=$ und wegen $x = y > z \rightarrow x > z$ ein x , nämlich $g_a^{-1}(g_b(x_b))$, so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x') = g_c(y) = g_a(x) + 1 \rightarrow \sum_j \alpha_j x'_j > \sum_j \alpha_j y_j$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.
- b) Entsprechend Lemma 6.2 gilt $\#MS(a) = \#MS(b) \wedge \forall x, y : g_a(x) = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$ und es gelte $\exists x \in MS(b) : \forall x' \in MS(b), y \in MS(c) : [g_b(x) \geq g_b(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j] \wedge g_b(x) + 1 > \#MS(b) \wedge \#MS(c) > g_b(x)$ (sei x_b das entsprechende $x \in MS(b)$). Da $g_a(g_a^{-1}(g_b(x_b))) = g_b(x_b)$ und wegen der Transitivität von $=$ gibt es dann ein x , nämlich $g_a^{-1}(g_b(x_b))$, so daß für alle $x' \in MS(a), y \in MS(c)$ gilt $g_a(x) \geq g_a(x') = g_c(y) \rightarrow \sum_j \alpha_j x'_j = \sum_j \alpha_j y_j$ und $g_a(x) + 1 > \#MS(a) \wedge \#MS(c) > g_a(x)$, d. h. $a \succ_\alpha c$ bzw. $a \succeq_\alpha c$.

4. Sei schließlich $a \sim_\alpha b$ und $b \sim_\alpha c$. Entsprechend Lemma 6.2 gilt dann $\#MS(a) = \#MS(b) \wedge \forall x, y : g_a(x) = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$ und $\#MS(b) = \#MS(c) \wedge \forall x, y : g_b(x) = g_c(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$. Aufgrund der Transitivität von $=$ folgt unmittelbar $\#MS(a) = \#MS(c) \wedge \forall x, y : g_a(x) = g_c(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j$, d. h. $a \sim_\alpha c$ bzw. $a \succeq_\alpha c$.

□

Die folgenden zwei Lemmata und drei Theoreme bereiten das Haupttheorem dieses Abschnittes vor, das Entscheidungen auf Grundlage der beschränkten Leximin-Ordnung als rawlsrational ausweist.

Lemma 6.3. Gegeben die Definition 6.6 und die oben gegebene Bestimmung der Funktionen g_a bzw. g_b gilt:

6 Der Urzustand: Ein Rekonstruktion

$$\begin{aligned} \forall a, b \in X^n : \forall l \in N, z \in MS(b) : & [\forall i \leq l : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \wedge g_a^{-1}(i) \in MS(b)] \\ & \wedge [\forall y \in MS(b) : \sum_j \alpha_j y_j = \sum_j \alpha_j z \rightarrow g_b(y) \leq l] \\ & \rightarrow z \in MS(a) \wedge g_a(z) \leq l \end{aligned}$$

Beweis. Es gelte für gewisse $l \in N$ und $z \in MS(b)$:

$$[\forall i \leq l : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \wedge g_a^{-1}(i) \in MS(b)] \quad (6.7)$$

$$\wedge [\forall y \in MS(b) : \sum_j \alpha_j y_j = \sum_j \alpha_j z \rightarrow g_b(y) \leq l] \quad (6.8)$$

Sei $H =_{def} \{y \in MS(b) : \sum_j \alpha_j y_j = \sum_j \alpha_j z\}$ und $K =_{def} \{x \in MS(a) : \sum_j \alpha_j x_j = \sum_j \alpha_j z \wedge g_a(x) \leq l\}$. Wegen (6.8) gilt dann:

$$H = \{y \in MS(b) : \sum_j \alpha_j y_j = \sum_j \alpha_j z \wedge g_b(y) \leq l\}$$

Die Funktion $h =_{def} g_b^{-1}(g_a(x))$ ist eineindeutig und wegen 6.7 definiert für alle $x \in MS(a)$ mit $g_a(x) \leq l$, also auch für alle $x \in K$. Entsprechend ist h^{-1} , d. h. $g_a^{-1}(g_b(y))$, wegen 6.7 und der Definition von H definiert für alle $y \in H$. h ist also eine Bijektion von K auf H . K und H sind deshalb gleichmächtig.

Sei $x \in K$, dann gilt wegen der Definition von K zusammen mit (6.7) $x \in MS(b)$ und $x \in H$. Es gilt also $K \subseteq H$. Da aber K und H endlich und – wie gerade gezeigt – gleichmächtig sind, müssen sie dann auch identisch sein. Es gilt also $y \in H \rightarrow y \in K$. z ist trivialerweise Element von H , also auch von K , womit aufgrund der Definition von K unmittelbar die Konklusion folgt. \square

Lemma 6.4. Gegeben die Definition 6.6 und die oben gegebene Bestimmung der Funktionen g_a gilt:

$$\begin{aligned} \forall a, b \in X^n : & [\forall x \in MS(a) : \exists y \in MS(b) : x \geq y] \\ & \rightarrow [\forall l \in N_a : [\forall i \leq l : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j] \rightarrow g_a^{-1}(l) \in MS(b)] \end{aligned}$$

Beweis. Es gelte

$$\forall x \in MS(a) : \exists y \in MS(b) : x \geq y \quad (6.9)$$

und für ein gewisses l :

$$\forall i \leq l : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \quad (6.10)$$

Aufgrund von Annahme 6.9 gibt es ein $y \in MS(b)$, so daß $g_a^{-1}(l) \geq y$. Wir betrachten ein solches y und nennen es y^* .

Mit $g_a^{-1}(l) = y^*$ folgt unmittelbar $g_a^{-1}(l) \in MS(b)$.

Nimm also an, $g_a^{-1}(l) > y^*$ (*). Es kann dann nicht gelten: $g_b(y^*) \geq l$. Aufgrund der Konstruktion von g_b und Annahme 6.10 würde dann nämlich gelten: $\sum_j \alpha_j y_j^* \geq \sum_j \alpha_j g_b(l)_j = \sum_j \alpha_j g_a(l)_j$, ein Widerspruch zu (*).

Da aber in jedem Fall $g_b(y^*) \geq 1$ folgt für $l = 1$: $g_a^{-1}(l) = y^*$. Damit gilt für $l = 1$ schon unmittelbar: $\forall i \leq l : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \rightarrow g_a^{-1}(i) \in MS(b)$.

Für $l > 1$ nimm an, es gelte $\forall i \leq l - 1 : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \rightarrow g_a^{-1}(l - 1) \in MS(b)$. Wegen Annahme 6.10 würde dann gelten $\forall i \leq l - 1 : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \wedge g_a^{-1}(l - 1) \in MS(b)$. Da wegen Annahme 6.10 $\sum_j \alpha_j g_a^{-1}(l)_j = \sum_j \alpha_j g_b^{-1}(l)_j$, würde für alle $y \in MS(b)$ aufgrund von (*) und der Konstruktion von g_b weiter gelten: $\sum_j \alpha_j y_j = \sum_j \alpha_j y_j^* \rightarrow g_b(y) \leq l - 1$. Für $l - 1$ und $z = y^*$ wären damit die Antezedensbedingung von Lemma 6.3 erfüllt. Damit wäre $y^* \in MS(a)$, zusammen mit (*) ein Widerspruch zur Definition von $MS(a)$. Es muß also gelten $g_a^{-1}(l) = y^*$ und damit $g_a^{-1}(l) \in MS(b)$, womit aufgrund vollständiger Induktion 6.4 folgt. \square

Theorem 6.9. Gegeben die Definitionen 6.6, 6.7, 6.8 und 6.9 gilt:

$$\forall a, b \in X^n : [\forall x \in MS(a) : \exists y \in MS(b) : x \geq y] \rightarrow a \succeq_\alpha b$$

Beweis. Betrachte a und b , so daß $\forall x \in MS(a) : \exists y \in MS(b) : x \geq y$, und eindeutige Funktionen g_a und g_b von $MS(a)$ auf $N_a = \{1, \dots, \#MS(a)\}$ bzw. von $MS(b)$ auf $N_b = \{1, \dots, \#MS(b)\}$, so daß $\forall x, x' \in MS(a) : \sum_j \alpha_j x_j > \sum_j \alpha_j x'_j \rightarrow g_a(x) > g_a(x')$ bzw. $\forall x, x' \in MS(b) : \sum_j \alpha_j x_j > \sum_j \alpha_j x'_j \rightarrow g_b(x) > g_b(x')$. Wir müssen zeigen $\neg b \succ_\alpha a$, bzw. auf der Basis von Definition 6.9:

$$\begin{aligned} & \neg \exists y \in MS(b) : \forall y' \in MS(b), x \in MS(a) : \\ & \quad [g_b(y) \geq g_b(y') = g_a(x) \rightarrow \sum_j \alpha_j y'_j = \sum_j \alpha_j x_j] \\ & \quad \wedge [[g_b(y') = g_a(x) = g_b(y) + 1 \rightarrow \sum_j \alpha_j y'_j > \sum_j \alpha_j x_j] \\ & \quad \vee [g_b(y) + 1 > \#MS(b) \wedge \#MS(a) > g_b(y)]] \end{aligned}$$

Da g_a und g_b eindeutig sind, können wir die Umkehrfunktionen g_a^{-1} und g_b^{-1} bilden und stattdessen zeigen, daß

$$\begin{aligned} & \neg \exists i \in N_b : \forall i' \in N_b : \\ & \quad [i \geq i' \rightarrow \sum_j \alpha_j g_b^{-1}(i')_j = \sum_j \alpha_j g_a^{-1}(i')_j] \\ & \quad \wedge [\sum_j \alpha_j g_b^{-1}(i + 1)_j > \sum_j \alpha_j g_a^{-1}(i + 1)_j \vee i + 1 \in N_a \wedge i + 1 \notin N_b] \end{aligned}$$

Betrachte ein i^* , so daß:

$$\begin{aligned} & \forall i' \in N_b : \\ & \quad [i^* \geq i' \rightarrow \sum_j \alpha_j g_b^{-1}(i')_j = \sum_j \alpha_j g_a^{-1}(i')_j] \\ & \quad \wedge [\sum_j \alpha_j g_b^{-1}(i^* + 1)_j > \sum_j \alpha_j g_a^{-1}(i^* + 1)_j \vee i^* + 1 \in N_a \wedge i^* + 1 \notin N_b] \end{aligned} \tag{6.11}$$

6 Der Urzustand: Ein Rekonstruktion

Nimm an, für ein $i \leq i^*$ gelte $\exists y \in MS(b) : g_a^{-1}(i) > y$. Da aber wegen (6.11), Lemma 6.4 und des Antezedens unseres Theorems $g_a^{-1}(i) \in MS(b)$, wäre das ein Widerspruch zur Konstruktion von $MS(b)$. Das Antezedens muß also für alle $x \in MS(a)$ mit $g_a(x) \leq i^*$ als Gleichheit gelten, d. h. $\forall x \in MS(a) : g_a(x) \leq i^* \rightarrow \exists y \in MS(b) : x = y$ oder, da aufgrund der Konstruktion von g_a gilt: $\forall x \in MS(a) : g_a(x) \leq i^* \rightarrow \sum_j \alpha_j x_j \leq \sum_j \alpha_j g_a^{-1}(i^* + 1)_j$, auch $\{x \in MS(a) : g_a(x) \leq i^*\} \subseteq \{y \in MS(b) : \sum_j \alpha_j y_j \leq \sum_j \alpha_j g_a^{-1}(i^* + 1)_j\}$. Aufgrund der Konstruktion von g_b und (6.11) gilt $\#\{y \in MS(b) : g_b(y) \leq i^*\} = \#\{y \in MS(b) : \sum_j \alpha_j y_j \leq \sum_j \alpha_j g_a^{-1}(i^* + 1)_j\} = i^*$. Da auch $\#\{x \in MS(a) : g_a(x) \leq i^*\} = i^*$ gilt schließlich wegen der Teilmengenbeziehung $\{x \in MS(a) : g_a(x) \leq i^*\} = \{y \in MS(b) : g_b(y) \leq i^*\}$ (*).

Wegen des Antezedens unseres Theorems und 6.11 gilt: $\exists y \in MS(b) : g_a^{-1}(i^* + 1) \geq y$. Sei y^* ein solches y . Aufgrund von (6.11) und der Konstruktion von g_b müßte gelten $g_b(y^*) \leq i^*$. Das führt sowohl für $g_a^{-1}(i^* + 1) > y^*$ wie für $g_a^{-1}(i^* + 1) = y^*$ zu einem Widerspruch. Im ersten Fall gäbe es dann wegen (*) $x, x' \in MS(a)$, so daß $x > x'$ im Widerspruch zur Konstruktion von $MS(a)$, im zweiten Fall wieder wegen (*) $x \in MS(a)$, so daß neben $g_a(x) = i^* + 1$ auch $g_a(x) = i'$ mit $i' \leq i^*$ im Widerspruch zur Konstruktion von g_a . □

Theorem 6.10. Gegeben die Definitionen 6.6, 6.7, 6.8 und 6.9 gilt:

$$\forall a, b \in X^n : [\forall x \in MS(a) : \exists y \in MS(b) : x \geq y] \wedge \neg[\forall y \in MS(b) : \exists x \in MS(a) : y \geq x] \\ \rightarrow a \succ_\alpha b$$

Beweis. Gegeben das Antezedens folgt aus Theorem 6.9 unmittelbar $a \succ_\alpha b$. Bleibt zu zeigen, daß gegeben das Antezedens des Theorems auch $\neg a \sim_\alpha b$, bzw. wegen Lemma 6.2, daß für eineindeutige Funktionen g_a und g_b von $MS(a)$ auf $N_a = \{1, \dots, \#MS(a)\}$ bzw. $MS(b)$ auf $N_b = \{1, \dots, \#MS(b)\}$, so daß $\forall x, x' \in MS(a) : \sum_j \alpha_j x_j > \sum_j \alpha_j x'_j \rightarrow g_a(x) > g_a(x')$ und $\forall x, x' \in MS(b) : \sum_j \alpha_j x_j > \sum_j \alpha_j x'_j \rightarrow g_b(x) > g_b(x')$, gilt:

$$\neg[\#MS(a) = \#MS(b) \wedge \forall x, y : g_a(x) = g_b(y) \rightarrow \sum_j \alpha_j x_j = \sum_j \alpha_j y_j] \quad (6.12)$$

Da g_a und g_b eineindeutig sind, können wir die Umkehrfunktionen g_a^{-1} und g_b^{-1} bilden und stattdessen zeigen, daß

$$\neg[\#N_a = \#N_b \wedge \forall i \in N_a : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j]$$

Nimm an:

$$\#N_a = \#N_b \wedge \forall i \in N_a : \sum_j \alpha_j g_a^{-1}(i)_j = \sum_j \alpha_j g_b^{-1}(i)_j \quad (6.13)$$

Wegen Lemma 6.4 und dem ersten Teil des Antezedens gilt dann $\forall i \in N_a : g_a^{-1}(i) \in MS(b)$ bzw., da es sich bei g_a um eine Bijektion handelt, $MS(a) \subseteq MS(b)$. Außerdem wegen $\#N_a = \#N_b$ bzw. $\#MS(a) = \#MS(b)$ aus (6.13) auch $MS(a) = MS(b)$.

Nimm weiter an, der erste Teil des Antezedens gelte für alle $x \in MS(a)$ als Gleichheit. Wegen $MS(a) = MS(b)$ würde dann im Widerspruch zum zweiten Teil des Antezedens auch gelten $\forall y \in MS(b) : \exists x \in MS(a) : y \geq x$. Für mindestens ein $x \in MS(a)$ muß also vielmehr gelten $\exists y \in MS(b) : x > y$ bzw. wegen $MS(a) = MS(b)$ auch $\exists x, y \in MS(b) : x > y$ – ein Widerspruch zur Definition von $MS(b)$. \square

Theorem 6.11. Gegeben die Definitionen 6.5, 6.6, 6.7, 6.8 und 6.9 gilt:

$$\forall S \in \Sigma_B : a, b \in S : B(a) \subset B(b) \rightarrow a \succ_\alpha b$$

Beweis. Aus dem Antezedens folgt aufgrund Theorem 6.7:

$$\forall x \in MS(a) : \exists y \in MS(b) : x \geq y \wedge \neg[\forall y \in MS(b) : \exists x \in MS(a) : y \geq x]$$

Mit Theorem 6.10 ergibt sich unmittelbar die Konklusion. \square

Theorem 6.12. Gegeben die Definitionen 6.2, 6.3, 6.5, 6.6, 6.7, 6.8 und 6.9 und die Axiome 6.7 und 6.8 gilt:

$$[\forall S \in \Sigma_B : C(S) = \{a \in \{a \in S : \forall i : a_i \notin B\} : \forall b \in \{a \in S : \forall i : a_i \notin B\} : a \succeq_\alpha b\}] \rightarrow \text{rrat}$$

Beweis. Nimm an $\forall S \in \Sigma_B : C(S) = \{a \in \{a \in S : \forall i : a_i \notin B\} : \forall b \in \{a \in S : \forall i : a_i \notin B\} : a \succeq_\alpha b\}$. Betrachte ein $S \in \Sigma_B$ und ein $a \in C(S)$ und nimm an $\exists b \in S : \forall w \in W(S, b), i, j \in N : b_i \notin \rho_j$. Sei b^* ein solches b , für das dann wegen der Definition 6.3 auch gelten muß $\forall i : b_i^* \notin B$. Wegen Theorem 6.4 gilt $\forall a \in \{a \in S : \forall i : a_i \notin B\} : B(b^*) \subseteq B(a)$ (*) und wegen Theorem 6.7 zusammen mit Theorem 6.9 $\forall a \in \{a \in S : \forall i : a_i \notin B\} : b^* \succeq_\alpha a$. b^* muß also ein Element von $C(S)$ sein. Für jedes $c \in C(S)$ muß dann weiter gelten $B(b^*) = B(c)$, da ansonsten wegen (*) und der Theoreme 6.7 und 6.10 entgegen der Definition von $C(S)$ im Antezedens gelten würde: $b^* \succ c$. Wegen $B(b^*) = B(c)$ und unseren Annahmen über b^* gilt also zuletzt für jedes $c \in C(S)$ auch $\forall w \in W(S, c), i, j \in N : c_i \notin \rho_j$ und zusammen mit Definition 6.4 folgt unmittelbar rrat. \square

6.4.2 Abweichungen von Rawls

In zwei wichtigen Aspekten weicht die Ordnung \succeq_α , für die gezeigt wurde, daß sie in der Tat im Urzustand als Entscheidungskriterium gewählt werden könnte, von der Rawlsschen Lösung ab: Zum einen ersetzt sie die Maximinordnung durch eine etwas kompliziertere Ordnung. Zum anderen wird das Grundgut „Freiheit“ anders als bei Rawls nicht besonders ausgezeichnet. Es gibt keinen Vorrang der Freiheit. Im folgenden möchte ich diese beiden Punkte kurz diskutieren.

6.4.2.1 Maximin

Das Besondere der hier vorgeschlagenen Ordnung ist es, daß durch ihre Wahl die Parteien *unabhängig* von α , d. h. völlig unabhängig von der spezifischen Gewichtung der

einzelnen Grundgüter im Index der Grundgüter, „rationale Reue“ ausschließen. Wann immer sie – gegeben ihr Wissen – vermeiden können, in einer Situation zu enden, die unerträglich für sie wäre, vermeiden sie dies auch in der Tat, egal welche Gewichtung sie gewählt haben.

Dies gilt nicht für das Maximinkriterium: Es gibt *keine* Gewichtung des Index von Grundgütern, die für alle Mengen von Alternativen, in denen unerträgliche Situationen ausgeschlossen werden können, diese auch in der Tat ausschließen. Würden die Parteien das Maximinkriterium befolgen, wären sie nicht immer rawlsrational. Rationale Parteien sind aber immer rawlsrational, sie werden also das Maximinkriterium nicht befolgen.

Das folgende Theorem illustriert das Problem. Wir beschränken uns dabei Einfachheit halber, und um das Problem des Maximinkriteriums zu isolieren, auf den Fall $B = \emptyset$. Es ergibt sich damit $\Sigma_B = \Sigma$ und $\{a \in S : \forall i : a_i \notin B\} = S$, was wir im folgenden schon berücksichtigen.

Wir definieren zunächst für ein gegebenes $\alpha \in \mathbb{R}_+^n$ die Maximinordnung $\succeq_{m(\alpha)}$:

Definition 6.10.

$$\forall a, b \in X^n : a \succeq_{m(\alpha)} b \leftrightarrow \exists i : \forall i' : \sum_j \alpha_j a_{i'j} \geq \sum_j \alpha_j b_{i'j}$$

Theorem 6.13. *Nimm an $B = \emptyset$ und $\forall \alpha : \exists x, y \in X : \sum_j \alpha_j x_j = \sum_j \alpha_j y_j \wedge x \neq y$.²¹ Gegeben die Definitionen 6.2, 6.3 und 6.5 und die Axiome 6.7 und 6.8 gilt dann:*

$$\forall \alpha \in \mathbb{R}_+^n : \exists S \in \Sigma : C(S) = \{a \in S : \forall b \in S : a \succeq_{m(\alpha)} b\} \rightarrow \neg \text{rrat}$$

Beweis. Betrachte $x, y \in X$ mit $\sum_j \alpha_j x_j = \sum_j \alpha_j y_j \wedge x \neq y$. Seien x^*, y^* entsprechende x, y . Nimm – ohne Beschränkung der Allgemeinheit – an $x_1^* > y_1^*$.

Sei $a^* = (x^*, x^*, \dots, x^*)$ und $b^* = (x^*, y^*, y^*, \dots, y^*)$. Betrachte $S = \{a^*, b^*\}$. Mit $\alpha' = (2\alpha_1, \alpha_2, \dots, \alpha_m)$ definiere eine Ordnung R über X mit $xRy \leftrightarrow_{\text{def}} \sum_j \alpha'_j x_j \geq \sum_j \alpha'_j y_j$, eine Menge unerträglicher Situationen $q = \{x \in X : \sum_j \alpha'_j x_j \leq \sum_j \alpha'_j y^*\}$ und $r = (R, R, \dots, R)$ sowie $\rho = (q, q, \dots, q)$. Dann sind für $(S, b^*, r, \rho, 2, 2)$ alle Bedingungen von Axiom 6.8 erfüllt. Insbesondere ist (wie man leicht sieht) r eine monotone Ordnung über X , ($x \in q \wedge xrx' \rightarrow x' \in q$) und $\forall i, j : a_i^* \notin \rho_j$, da $\sum_j \alpha'_j x_j^* > \sum_j \alpha'_j y^*$. Außerdem ist $B(a^*) \subseteq B(b^*)$. Wegen Theorem 6.4 gibt es also $b \in S$, so daß $\forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j$. Weiter wäre mit $C(S) = \{a \in S : \forall b \in S : a \succeq_{m(\alpha)} b\}$ auch $b^* \in C(S)$, da $\forall i : \sum_j \alpha_j b_{ij}^* \geq \sum_j \alpha_j a_{ij}^*$ und zuletzt ist (konstruktionsgemäß) $b_2^* \in q$. $\forall x \in B : \exists j \in N : x \in \rho_j$ gilt trivialerweise. Zusammengenommen gilt also – entgegen der Definition 6.4 von Rawls-Rationalität:

$$\exists S \in \Sigma, a \in S, w \in W(S, a) :$$

$$a_{i^*} \in \rho_{j^*} \wedge [a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]]$$

□

²¹ Diese Bedingung ist erfüllt, wenn es für $i, j \in M$ mit $i \neq j$ und irgendwelche $x \in X$ Intervalle I und J gibt, so daß für jedes x' mit $x'_h = x_h$ für alle $h \neq i \wedge h \neq j$ und $x'_i \in I \wedge x'_j \in J$ gilt $x' \in X$.

6.4.2.2 Der Vorrang der Freiheit

Die Rawlssche Lösung der Entscheidungsproblems im Urzustand besteht bekannterweise aus zwei Prinzipien, deren erstes dem zweiten vorgeordnet ist und ein bestimmtes, gleiches Maß an Grundfreiheiten (in der Form einer Liste) für alle verlangt. Es gibt wenig Zweifel, daß Rawls, obwohl er das Differenzprinzip – als Teil des zweiten Prinzips – bis zum Ende seines Schaffens aufrechterhalten hat, dem ersten Prinzip eine viel größere Bedeutung einräumte.²²

Umso wichtiger ist es, den Zusammenhang des Vorrangs der Freiheit mit dem hier behandelten Vorschlag einer Rekonstruktion darzustellen.

Wir nehmen dazu an, daß es sich bei dem ersten Gut um das Gut der Freiheit handelt, von dem wir voraussetzen, daß es sich – auch wenn es sich um eine Zusammenfassung verschiedener *Freiheiten* zu einem System von Freiheiten handelt – in geeigneter Weise numerisch repräsentieren lasse.

Wie die zwei Prinzipien das erste Prinzip als vorgeordnet mit einer Anwendung des Differenzprinzips auf die übrigen Grundgüter verknüpfen²³, so wollen auch wir im folgenden unsere Versuche, eine Vorrangstellung des ersten Gutes zu rekonstruieren, jeweils mit der Anwendung der beschränkten Leximinordnung statt der als inadäquat aufgezeigten Maximinordnung verbinden.

Einer der entscheidenden Gründe für das erste Prinzip und seinen Vorrang ist die schon mehrmals erwähnte Überzeugung Rawls', daß uns bekannt ist, daß zumindest für eine *Gruppe* von Menschen Situationen, die ein gewisses Grundmaß an Freiheiten (das z. B. die Gewissensfreiheit einschließt) nicht gewähren, unabhängig von dem Maß anderer Grundgüter nicht akzeptabel sind. Diese Überzeugung können wir – gegeben die Annahme des vorletzten Absatzes – bei entsprechender Interpretation der Menge *B* der bekannterweise inakzeptabler Situationen leicht formal rekonstruieren, indem wir ohne Beschränkung der Allgemeinheit für ein beliebiges zahlenmäßig bestimmtes Niveau an Grundfreiheiten – sagen wir 0 – annehmen, daß genau die Situationen bekannterweise inakzeptabel sind, die dieses oder ein geringeres Niveau an Grundfreiheiten bieten.²⁴

Axiom 6.12.

$$B = \{x : x_1 \leq 0\}$$

Mit diesem Axiom und der vorangestellten Interpretation kann auch die von uns vorgeschlagene Lösung als eine mögliche Rekonstruktion des Vorrangs der Freiheit

²² Das wird z. B. deutlich in Rawls [2001, 94-97]. Rawls geht davon aus, daß, was er den ersten Vergleich nennt, die Gründe für das erste Prinzip gibt, und hält diesen Vergleich für „quite conclusive“. Der zweite Vergleich gebe die Gründe für das Differenzprinzip und sei „less conclusive“ (S. 95). Weiterhin hält er den ersten Vergleich für „more fundamental because the aim of justice as fairness is to work out an alternative conception of political justice to those found in utilitarianism, perfectionism, and intuitionism [. . .], while at the same time finding a more appropriate moral basis for the institutions of a modern democratic society. Should the two principles win in the first comparison, this aim is already in good part achieved; but should they lose, all is lost.“

²³ Von dem ersten Teil des zweiten Prinzips sehen wir dabei wie auch sonst ab.

²⁴ Das im folgenden festgelegte *B* erfüllt Axiom 6.7.

verstanden werden. Intuitiv verlangt sie, sich einfach auf alle die Vektoren gesellschaftlicher Positionen zu beschränken, die nicht bekannterweise für einige Individuen zu wenig Grundfreiheiten bieten, und auf diese dann die beschränkte Leximin-Ordnung anzuwenden.

In vor allem zwei Punkten unterscheidet sich diese Rekonstruktion des Vorrangs der Freiheit allerdings substantiell von den Rawlsschen zwei Prinzipien.

Zum einen nimmt die beschränkte Leximin-Ordnung nach wie vor *alle* Grundgüter – inklusive der Freiheit – in Betracht, während Rawls sein Differenzprinzip nur noch auf die übrigen Grundgüter anwenden möchte. Diese weitere Beachtung der Freiheit auch nach der Vorselektion auf der Basis der Freiheit ist notwendig, da (wie sich leicht auch formal zeigen ließe) nicht auszuschließen ist, daß (nur) ein höheres als das notwendige Mindestmaß an Freiheit für alle sicherstellen kann, daß alle Positionen für alle akzeptabel sind. Und dies selbst dann, wenn – in einem gleich noch zu klärenden Sinn²⁵ – dieses Mindestmaß gleichzeitig ausreichend ist: Dazu müssen wir nur Individuen zulassen, die in der relevanten Region des Raumes der Grundgüter niedrigere Niveaus anderer Güter durch ein höheres Niveau an Grundfreiheiten auszugleichen bereit sind.

Wichtiger – vor allem auch für ein kritisches Verständnis der Rawlsschen Position – ist die Tatsache, daß Rawls in seiner Formulierung des Vorrangs der Freiheit ein *gleiches* Niveau an Freiheit für alle verlangt.²⁶

Rawls hat sein erstes Prinzip in verschiedenen Fassungen vorgelegt, und auch in der Rekonstruktion sind verschiedene Interpretationen möglich, die zum Teil auch verschiedene Anforderungen an die Welten stellen, denen sich die Parteien möglicherweise gegenüber sehen. In allen Versionen des ersten Prinzips hat Rawls – für die ideale Theorie – eine Gleichverteilung der Freiheit gefordert, Unterschiede ergeben sich in dem Niveau, auf dem diese Gleichverteilung stattfinden sollte. In der ursprünglichen Version der Theory of Justice hat Rawls im ersten Prinzip eine Art Maximierung von Freiheit verlangt:

First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others. [Rawls 1973, 60]

Rawls hat diesen Maximierungsgedanken später vollkommen aufgegeben, und statt dessen gefordert, daß allen Personen einer Gesellschaft ein „völlig adäquates System gleicher Grundfreiheiten“ gewährt wird. [Rawls 1992b, 160] Da wir grundsätzlich davon ausgehen, daß Personen ein Interesse an allen Grundgütern haben, und nicht notwendigerweise Präferenzen haben, die sich wie bei lexikographischen Präferenzen auf die einzelnen Güter hin „separieren“ lassen, ist es nicht evident, was es für die Menge eines einzelnen Gutes bedeutet, daß sie für jede Person ausreichend (adäquat) ist. Abgesehen von dem Fall, wo alle Personen Präferenzen haben, die – wenigstens im

²⁵ Siehe unten S. 183.

²⁶ Bei dieser Formulierung setzen wir – wie oben S. 181 ausgeführt – voraus, daß sich die verschiedenen Systeme von Grundfreiheiten numerisch repräsentieren lassen. Der späte Rawls verwendet in der Regel den Plural und spricht etwa von „equal basic liberties“ oder von der „equality of basic liberties“ [Rawls 2001, 105, 111].

relevanten Bereich – Freiheit den übrigen Gütern lexikographisch vorordnen, macht es keinen Sinn, davon zu sprechen, daß ein bestimmtes Maß von Freiheit unabhängig vom Maß anderer Güter adäquat wäre: Adäquat sind im allgemeinen (d. h. ohne radikale Einschränkung der betrachteten Präferenzen und Mengen inakzeptable Positionen) eben Bündel von Grundgüter und nicht Mengen einzelner Grundgüter. Die Frage, ob ein bestimmtes Maß von Grundfreiheiten adäquat ist, kann deshalb davon abhängen, welches Niveau der übrigen Grundgüter zur Verfügung steht. Selbst Rawls hat wenigstens indirekt verschiedenemale eingestanden, daß, solange eine Basisniveau anderer Grundgüter nicht erreicht ist, Personen nicht primär an Freiheit interessiert sind [vgl. etwa Rawls 1973, 542], was als Konsequenz dann auch bedeuten würde, daß dieses Basisniveau notwendig dafür wäre, daß ein Gesamtbündel als adäquat erachtet werden könnte. Ohne weitere Einschränkung der Präferenzen kann also, solange man das Prädikat auf das Maß der Grundfreiheiten anwendet, „adäquat“ nur heißen, daß es Niveaus der übrigen Güter gibt, die mit der entsprechenden Menge der Grundfreiheiten kombiniert ein insgesamt adäquates Gesamtbündel ergeben. Daß es Grundfreiheiten in einem solchen (*unbestimmten*) Maß für alle Individuen geben muß, ergibt sich schon einfach aus Axiom 6.8. Rawls geht aber nicht nur davon aus, daß es entsprechende allgemein verwirklichtbare und adäquate Niveaus von Grundfreiheiten gibt, sondern außerdem, daß wir ein solches Niveau kennen und dann für alle verwirklichen sollen. Das Niveau ist durch eine aus der Tradition übernommene²⁷ Liste von Grundfreiheiten und ihre Interpretation gegeben. Formal können wir das dadurch darstellen, daß wir ein beliebiges, durch eine Zahl repräsentiertes Niveau als adäquat und für alle verwirklichtbar annehmen und in einem Rawlsschen Gerechtigkeitsprinzip seine Verwirklichung verlangen.²⁸ Sei 1 das in der gegebenen Interpretation adäquate Niveau der Freiheit. Damit die Forderung des erstens Prinzips erfüllbar ist, muß es dann solche $a \in S$ geben, so daß für alle Individuen i $a_{i1} = 1$. Unser Axiom 6.8 wird modifiziert zu:²⁹

Axiom 6.13.

$$\forall S \in \Sigma_B, a \in S : W(S, a) = \{(S, a, r, \rho, i^*, j^*) : \\ (\forall i \in N : r_i \text{ ist eine Ordnung über } X \wedge \forall x, x' \in X : x \geq \wedge x' \rightarrow xr_i x')\} \quad (6.14)$$

$$\wedge (\forall i \in N, x, x' \in X : x \in \rho_i \wedge xr_i x' \rightarrow x' \in \rho_i) \quad (6.15)$$

$$\wedge (i^*, j^* \in N) \quad (6.16)$$

$$\wedge (\exists b \in S : \forall i \in N : b_{i1} = 1 \wedge \forall j : b_j \notin \rho_j) \quad (6.17)$$

$$\wedge (\forall x \in B : \exists j \in N : x \in \rho_j) \quad (6.18)$$

Selbst für diese Version unseres Axioms 6.8, die versucht, Rawls entgegenzukommen, werden wir nun zeigen, daß ein Vorrang der Freiheit, der Gleichheit verlangt, problematisch ist.

²⁷ Faktisch. Aber siehe 139.

²⁸ In der Interpretation ist die gewählte Zahl dann der Index für das adäquate System der Grundfreiheiten, andere Zusammenstellungen von Grundfreiheiten sind – wie schon oben festgestellt – ebenfalls als auf angemessene Weise indiziert vorzustellen.

²⁹ Bedingung 6.4 von Axiom 6.8 folgt unmittelbar aus Bedingung 6.17 und ist deshalb verzichtbar.

6 Der Urzustand: Ein Rekonstruktion

In der gegebenen Definition von „adäquat“ wären die zwei Prinzipien der Gerechtigkeit folgendermaßen formulierbar:³⁰

Definition 6.11. *Zwei Prinzipien*

$$\forall S \in \Sigma_B : C(S) = \{a \in S : [\forall i : a_{i1} = 1] \wedge [\forall b : [\forall i : b_{i1} = 1] \rightarrow a \succeq_\alpha b]\}$$

Wir zeigen nun:

Theorem 6.14. *Betrachte beliebige $X \setminus B$, so daß:*

$$\begin{aligned} & \exists z, z', y, y' \in X \setminus B : \\ z_1 = z'_1 = 1 \wedge \neg(y_1 = 1 \wedge y'_1 = 1) \wedge \neg(z \geq z' \vee z' \geq z) \wedge y > z \wedge y' > z \wedge y > z' \wedge y' > z' \end{aligned} \quad (6.19)$$

Gegeben die Axiome 6.7 und 6.13 und Definitionen 6.2, 6.3, 6.5, 6.4 und 6.11 gilt:

$$\forall \alpha \in \mathbb{R}_+^m : \exists S \in \Sigma_B : \neg \text{rrat}$$

Beweis. Betrachte beliebige α und $S = \{a, b, c\} \in \Sigma_B$, so daß $a = (z, \dots, z)$, $b = (z', \dots, z')$ und $c = (y, y', \dots, y')$ mit z, y, z' und y' , die Bedingung 6.19 erfüllen. Offensichtlich muß dann wegen Axiom 6.13 gelten $\forall w' \in W(S, c), i, j \in N : c_i \notin \rho_j$, d. h.:

$$\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j$$

Außerdem ist wegen Definition 6.11 entweder $a \in C(S)$ oder $b \in C(S)$. Ohne Beschränkung der Allgemeinheit nimm an: $a \in C(S)$. Bleibt zu zeigen, daß $\exists w \in W(S, a) : a_{i^*} \in \rho_{j^*}$. Betrachte wieder die im Beweis zu Theorem 6.4 konstruierte Ordnung R mit $U(0) = \{x : x \in B(b) \wedge \neg \exists i \in N : x \gg b_i\}$, und die Menge inakzeptabler Situationen $q = \{x \in X : \exists u < 0 : x \in U(u)\}$, $r = (R, \dots, R)$ und $\rho = (q, \dots, q)$. Wie wir wissen, sind für die so definierten r und ρ Bedingungen 6.1 und 6.2 von Axiom 6.8 und damit natürlich auch die identischen Bedingungen 6.14 und 6.15 von Axiom 6.13 erfüllt. Da $\forall x \in B : x \notin B(b)$ und $\forall i : a_i = z \wedge z \notin B(b)$ folgen wieder sowohl $\forall x \in B : x \in q$ (mit Axiom 6.7), und damit Bedingung 6.18, als auch $\exists i : a_i \in q$. Zuletzt gilt wieder $\forall i : b_i \notin q$ und, da $\forall i : b_{i1} = 1$, Bedingung 6.17. Es ist dann $w = (S, a, r, \rho, 1, 1) \in W(S, a)$ und zusammengenommen gilt also entgegen der Definition 6.4 von rrat :

$$\begin{aligned} & \exists S \in \Sigma_B, a \in S, w \in W(S, a) : \\ & a_{i^*} \in \rho_{j^*} \wedge [a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]] \end{aligned} \quad (6.20)$$

□

³⁰ Näher an Rawls müßte im folgenden statt \succeq_α eine entsprechende Ordnung auf der Basis eines Index nur über Vektoren (a_{i2}, \dots, a_{i3}) verwendet werden. Für das im Folgenden zuerst vorgetragene *negative* Resultat ist dieser Unterschied zu Rawls aber irrelevant, da es nur auf die Gleichheitsforderung ankommt. Für das spätere *positive* Resultat ist der Unterschied irrelevant, da ohnehin nur $a \in S$ verglichen werden, bei denen für alle i a_{i1} identisch gleich 1 ist.

Man sieht leicht, daß dieser Beweis sich auch direkt auf den Fall der ursprünglichen – ein Maximum verlangenden – Rawlschen Formulierung der zwei Prinzipien übertragen ließe, da im konstruierten Beispiel alle Vektoren gesellschaftlicher Positionen, die allen Individuen dasselbe Niveau an Grundfreiheiten bieten, dies auf gleichem Niveau tun, und deshalb auch gleichzeitig das Maximum bieten, das mit Gleichheit vereinbar ist. Außerdem wäre für Fälle der Existenz von entsprechenden Vektoren von Positionen, in denen alle das gleiche Niveau an Freiheiten genießen, der Beweis auch für das ursprüngliche Axiom 6.8 anwendbar, da ja aus der Erfüllung des Axioms 6.13 die Erfüllung des Axioms 6.8 direkt folgt.

Das Problem ist allemal, daß die Festlegung auf die Gleichheit in den Grundfreiheiten dazu zwingen kann, ein niedrigeres – und deshalb möglicherweise unzureichendes – Niveau der übrigen Grundgüter zu wählen, obwohl es möglich wäre, allen Individuen mehr oder zumindest gleich viel aller Grundgüter (inklusive der Grundfreiheiten) zu bieten.

Bis jetzt haben wir Bedingung 6.17 nur deshalb in Axiom 6.13 aufgenommen, um den zwei Prinzipien 6.11 eine Chance geben. Die Bedingung ist notwendig, um eine Lösung wie die zwei Prinzipien überhaupt in Betracht ziehen zu können. Wie sich herausstellte, ist sie allerdings nicht ausreichend.

Wollen wir Probleme wie in Theorem 6.14 vermeiden, ohne jedoch X so zu verarmen, daß Konstellationen wie in Bedingung 6.19 unmöglich werden, und ohne die möglichen Präferenzen und Mengen inakzeptabler Positionen explizit weiter zu beschränken, dann müssen wir den möglichen Mengen von Alternativen S Beschränkungen auferlegen.

Meines Wissens hat Rawls nirgends *explizit* eine der Bedingung 6.17 von Axiom 6.13 entsprechende Voraussetzung seiner Lösung formuliert, obwohl er sie natürlich implizit annehmen muß, wenn er für seine Lösung beansprucht, anders als der Utilitarismus sichere sie – gegeben seine Voraussetzungen – eine für alle akzeptable Lösung. Wir können also nur spekulieren, wie er sie begründen würde.

Eine Vermutung könnte sein, daß Rawls – ohne es explizit zu machen – von etwas ausging, was ich „free disposal of freedom“ nennen möchte: die Vorstellung, daß es immer möglich ist, ohne Einbußen an anderen Grundgütern die Grundfreiheiten für alle auf das gleiche adäquate Niveau zu reduzieren. Eine entsprechende Annahme könnte zum einen *einen Teil* der Bedingung 6.17 in Axiom 6.13 begründen,³¹ und wäre zum anderen hinreichend, um die zwei Prinzipien als rawlsrational zu erweisen.

Definiere für jedes $a \in X^n$ das „egalisierte“ $e(a)$ so, daß $\forall i : e(a)_{i1} = 1$ und $\forall i : \forall j \neq 1 : e(a)_{ij} = a_{ij}$.

Wir nehmen zunächst an:

Axiom 6.14.

$$\forall a \in X^n : \forall i : a_{i1} \geq 1 \rightarrow e(a) \in X^n$$

³¹ Und zwar der, der sich auf die bloße Existenz entsprechender – nicht notwendig allgemein akzeptabler – Vektoren bezieht. Die korrespondierende Annahme einer kostenlosen *Produktion* von Grundfreiheiten wäre übrigens hinreichend, um die Maximierungsversion der zwei Prinzipien als rawlsrational zu erweisen und würde zugleich die Existenz entsprechender akzeptabler Vektoren sichern.

Und wichtiger:

Axiom 6.15.

$$\forall S \in \Sigma_B : \forall a \in S : [\forall i : a_{i1} \geq 1] \rightarrow e(a) \in S$$

Theorem 6.15. Gegeben Axiome 6.15, 6.14, 6.13, 6.12 und Definitionen 6.2, 6.3, 6.5, 6.4 und 6.11 gilt rrat.

Beweisskizze. • Ganz analog zum Beweis von Theorem 6.4 läßt sich zunächst zeigen:

$$\begin{aligned} \forall S \in \Sigma_B : \forall b \in S : [\forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j \leftrightarrow \\ [\forall a \in S : [\forall i : a_{i1} = 1 \rightarrow B(b) \subseteq B(a)]] \end{aligned}$$

Beim Beweis der Implikation von rechts nach links ist dabei zu beachten, daß der Beweis nur davon abhängt, daß wir für die Menge, von deren Elemente a angenommen wird $B(b) \subseteq B(a)$, wissen, daß sie ein Element a enthält, so daß $\forall i, j : a_i \notin \rho_j$. Das ist für $\{a : \forall i : a_{i1} = 1\}$ über Bedingung 6.17 in Axiom 6.13 sichergestellt.

Beim Beweis der Implikation von links nach rechts ist Folgendes zu beachten: Solange sichergestellt ist, daß für alle Elemente der Menge, für deren Elemente gezeigt werden soll $B(b) \subseteq B(a)$, auch gilt $\forall i : a_i \notin B$ (hier aufgrund der Spezifikation von B in Axiom 6.12 der Fall), ist für $\neg B(b) \subseteq B(a)$ immer (für alle $b \in S$) die vorgenommene Konstruktion einer Präferenzordnung und einer Menge q möglich, die die Bedingungen von Axiom 6.8 erfüllt, so daß $\exists i : b_i \in q$. Für $\{a : \forall i : a_{i1} = 1\}$ ist damit automatisch auch Bedingung 6.17 in Axiom 6.13 erfüllt.

- Auf der Basis des zu Theorem 6.4 parallelen Satzes läßt sich dann analog zu Theorem 6.5 leicht zeigen:

$$\begin{aligned} \forall S \in \Sigma_B : [\forall a \in S, w \in W(S, a) : a_{i^*} \in \rho_{j^*} \rightarrow \\ \neg[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i, j \in N : b_i \notin \rho_j]]] \leftrightarrow \\ \exists b \in S : \forall a \in \{a : \forall i : a_{i1} = 1\} : B(b) \subseteq B(a) \rightarrow \\ \forall c \in C(S) : \forall a \in \{a : \forall i : a_{i1} = 1\} : B(c) \subseteq B(a) \end{aligned}$$

- Nimm an $\exists b \in S : \forall a \in \{a : \forall i : a_{i1} = 1\} : B(b) \subseteq B(a)$ (*). Zu zeigen ist $\forall c \in C(S) : \forall a \in \{a : \forall i : a_{i1} = 1\} : B(c) \subseteq B(a)$.

Wir zeigen zunächst, daß $\forall a \in \{a : \forall i : a_{i1} = 1\} : B(e(b)) \subseteq B(a)$. Wegen (*) gilt: $\forall a \in \{a : \forall i : a_{i1} = 1\}, j \in N : \exists i : b_j \geq a_i$. Aber wenn $x \geq y$ und $y_1 = 1$, dann ist auch $(1, x_2, \dots, x_m) \geq y$, es muß also auch gelten $\forall a \in \{a : \forall i : a_{i1} = 1\}, j \in N : \exists i : e(b)_j \geq a_i$ bzw. $\forall a \in \{a : \forall i : a_{i1} = 1\} : B(e(b)) \subseteq B(a)$ und bei der in Definition 6.11 gegebenen Bestimmung von $C(S)$ auch $\forall c \in C(S) : B(e(b)) \subseteq B(c)$. Gelte nun für ein $c \in C(S) : \neg \forall a \in \{a : \forall i : a_{i1} = 1\} : B(c) \subseteq B(a)$. Dann

wäre $B(e(b)) \subset B(c)$ und wegen Theorem 6.11 $e(b) \succ_{\alpha} c$, ein Widerspruch zur Bestimmung von $C(S)$. □

Wenn wir zu Axiom 6.8 zurückkehren – die Version 6.13 sollte ja zunächst nur den zwei Prinzipien 6.11 eine Chance geben –, dann können wir festhalten, daß im Urzustand in einem Vergleich der zwei Prinzipien in der Formulierung 6.11 und der Entscheidung auf der Basis der beschränkten Leximin-Ordnung die letztere gewählt und die ersteren abgelehnt würden.

Für den von uns zunächst gewählten Rahmen ist das, was Rawls eine Spezifizierung der allgemeinen Gerechtigkeitsprinzips nennt, und für ihn – auch wenn Vorrang der Freiheit genannt – immer die Forderung der Gleichheit in den Grundfreiheiten beinhaltet, nicht zu begründen.

Anders verhält es sich, wenn man die Kombination von Bedingung 6.17 von Axiom 6.13 mit Axiom 6.15 annimmt. Die Frage ist also, ob man diese Voraussetzungen für die Konstruktion des Urzustandes für angemessen hält: Ob wir also annehmen in einer Welt zu leben, in der wir es nicht nötig haben, eine mangelnde Ausstattung mit den übrigen Grundgütern durch ein höheres Maß an Freiheit auszugleichen, und in der Freiheit über das als angemessen erkannte Maß hinaus keine Anreizeffekte für die „Produktion“ der übrigen Grundgüter besitzt.

6.4.3 Und der Utilitarismus?

Eines der wichtigsten Anliegen Rawls' war es, eine Alternative zum Utilitarismus zu entwickeln. Es mußte ihm deshalb auch darum gehen, zu zeigen, daß im Urzustand der Utilitarismus nicht gewählt würde, und in der Tat versucht er in „A Theory of Justice“, das für einige Varianten – den seines Erachtens wichtigsten – des Utilitarismus zu zeigen.

Damit stellt sich die Frage, ob auch in unserer Rekonstruktion des Urzustands der Utilitarismus abgelehnt würde.

Es gibt zwei Gründe, weshalb der Utilitarismus gewissermaßen von vornherein als Lösung nicht in Frage kommt: Wir haben nicht angenommen, daß die Präferenzen der Individuen, als die sich die Parteien möglicherweise wiederfinden, notwendigerweise durch Nutzenfunktionen darstellbar sind und auch Rawls macht diese Voraussetzung nicht.

Zwar könnte man natürlich auch für Personen mit Präferenzen, die nicht durch eine Nutzenfunktion repräsentierbar sind, irgendeine Nutzenfunktion stipulieren. Aber neben der dann ungeklärten Frage, wie eine solche Nutzenfunktion aussehen sollte, verliert der Utilitarismus ein wesentliches Stück seiner Attraktivität, wenn seine Nutzenfunktionen nicht den Präferenzen der Individuen entsprechen. Außerdem wird einem Begründungsprogramm, das den Utilitarismus à la Harsanyi aus einer Variante des Urzustands ableiten möchte und für die Individuen nicht nur ordinale sondern sogar kardinale Nutzenfunktionen voraussetzt (z. B. Harsanyi [1977b, 54]), der Boden

entzogen.³²

Des weiteren haben wir das Entscheidungsproblem der Parteien im Urzustand in einer Weise formuliert, die den Utilitarismus als Lösung nicht zuläßt. Denn auch in den eigentlichen Entscheidungssituationen ist den Parteien nicht bekannt, wer welche Position einnehmen wird, oder anders gesagt, sie können ihre Entscheidung nicht von einer Zuordnung von Positionen zu Präferenzen (oder dann Nutzenfunktionen) abhängig machen. Die Entscheidungsmenge $C(S)$ muß unabhängig davon sein, wer – mit welcher Nutzenfunktion – die jeweiligen Positionen einnimmt, während im Utilitarismus bei hinreichend unterschiedlichen Nutzenfunktionen diese Ordnungen sehr verschieden sein werden, abhängig davon, wer die Positionen einnimmt.

Beide Gründe artikulieren meines Erachtens beachtliche Einwände gegen den Utilitarismus. So spricht einiges dafür, zumindest in Bereichen lexikographische (oder ähnliche) Ordnungen nicht auszuschließen, die sich unter der Voraussetzung, daß sie über einen Bereich gelten, in dem sowohl das lexikographisch vorgeordnete Gut wie die lexikographisch nachgeordnete Gütergruppe überabzählbar unendlich sind, bekanntlicherweise nicht durch eine Nutzenfunktion darstellen lassen. Insbesondere beschreibt Rawls die Interessen von Menschen, für die zumindest die Gewissensfreiheit – ein integraler Bestandteil des Grundgutes der Freiheit – nicht „kommensurabel“ mit anderen menschlichen Interessen ist, in einer Weise, die (zusammen mit der Annahme der Monotonie) – wenigstens regional und wenigstens bezogen auf das Gut der Freiheit – lexikographische Präferenzen zu meinen scheint: Für sie gibt es kein Austauschverhältnis, so daß es rational wäre, „den Schutz des einen gegen den Schutz des anderen abzuwägen“ [Rawls 1992b, 183].³³ Für sie gilt eben nicht, daß „für jeweils zwei beliebige Interessen bei gegebenem Grad der Befriedigung gilt: Es gibt stets einen Wechselkurs, an den sich rationale Parteien halten und dem entsprechend sie bereit sind, eine geringere Befriedigung des einen Interesses im Austausch für eine höhere Befriedigung des anderen Interesses zu akzeptieren (und umgekehrt).“ [Rawls 2003, 167 f.]³⁴ Vielmehr gilt für religiöse, philosophische und moralische Anschauungen, „daß das Bekenntnis zu solchen Anschauungen [...] als sozusagen nicht verhandlungsfähig anerkannt wird. Sie werden als Formen des Glaubens und Handelns verstanden, deren Schutz [durch u.a. die Gewissensfreiheit, A.G.] wir richtigerweise nicht wegen der vom zweiten Gerechtigkeitsgrundsatz abgedeckten Überlegungen aufgeben können“. [Rawls 1992b, 182] Ja, Rawls geht so weit, denjenigen, die nicht verstehen, daß für jemanden, der etwa einer bestimmten Religion anhängt, die Freiheit, dieser Religion entsprechend zu leben, nicht verhandelbar ist, und gegen z. B. kein Einkommen, in welcher Höhe auch immer, eintauschbar ist, schlichtweg ein Verständnis davon ab, was es heißt einer Religion anzuhängen. [Vgl. Rawls 1992b, S. 183]³⁵

³² Zur Interpretation des Harsanyi-Programms als einer Version des Urzustands vgl. Harsanyi [1976a, 44-46]

³³ Zum Verständnis der Stelle sollte man allerdings auch auf den englischen Text [Rawls 1993a, 312] zurückgreifen, da ein nicht zitierter Teil der Übersetzung mißverständlich ist.

³⁴ Im Original Rawls [2001, 105].

³⁵ Rawls formuliert sein Argument allgemeiner und abstrakter, als ich es hier wiedergeben. Wie dem auch sei, es gibt genügend Menschen in der Geschichte der Religionen, die durch Verhalten gezeigt haben,

Für eine Form der Aufgabenstellung, die von vornherein eine Abhängigkeit die ethische Beurteilung gesellschaftlicher Zustände von individuellen Eigenschaften der Individuen, die bestimmte gesellschaftliche Positionen einnehmen, ausschließt, scheinen mir ein zunächst praktisches und ein im strikten Sinne und unmittelbar ethisches Argument zu sprechen.³⁶

Das erste Argument stellt die praktischen Anwendbarkeit einer Theorie in Frage, die die Zuordnung gesellschaftlicher Positionen auf der Basis detaillierter individueller Eigenschaften verlangt. Rawls selbst hat ja seine Theorie nicht als theoretische Spielerei verstanden, sondern als einen Beitrag zur Lösung einer praktischen Problemstellung unserer modernen Gesellschaften. Eine Theorie, die verlangen würde, die Grundinstitutionen einer Gesellschaft auf Wissen um z. B. die Nutzenfunktionen aller einzelnen Individuen hin zu gestalten und es als Aufgabe gesellschaftlicher Planung betrachtet, einzelnen Individuen aufgrund dieses Wissens bestimmten Positionen zuzuordnen, kann für die modernen großen Gesellschaften nur (schlechte) Utopie sein. Um es konkreter am Beispiel zu sagen, ein Utilitarismus, der tatsächlich verlangen würde, die Gesellschaft so zu gestalten, daß jedes einzelne Individuum den individuellen Nutzen erreicht, der gerade zu eine Maximierung des Gesamtnutzens führt, hätte sich eben damit als praktisch anwendbare Theorie der Gerechtigkeit desavouiert. Die Vorstellung ist einfach lächerlich und in der Tat sind uns nicht nur für die große Mehrheit der Menschen etwa Deutschlands Nutzenfunktionen über die Grundgüter unbekannt sondern für alle.³⁷

Das direkt ethische Argument hat sich schon in der Qualifikation des utopischen Charakters einer individuellen Zuordnung von gesellschaftlichen Positionen zu Individuen aufgrund des Wissens um individuelle Eigenschaften, wie z. B. ihren Nutzenfunktionen, als einer *schlechten* Utopie angekündigt. Es scheint unseren (liberalen) sozial-ethischen Grundvorstellungen entgegenzustehen, die verlangen, daß die gesellschaftli-

daß sie nicht bereit waren, ihre religiösen Überzeugungen gegen den Schutz ihres Einkommens, ja ihres Lebens einzutauschen.

Da alle diese Menschen nur endliches Einkommen, oder endliche Mengen anderer (Grund-)Güter verloren haben, könnte man rein formal auch ihr Verhalten mit Nutzenfunktionen modellieren, die bei einem höheren Verlust zu einer Aufgabe ihrer Überzeugungen geführt hätten. Die Frage ist allerdings, ob eine solche Vorgehensweise den Motivationen von z. B. Märtyrern gerecht würde. Mit Rawls wird man sagen müssen, daß wer meint, bei einem höheren Verlust wäre jeder bereit, seine Religion zu verraten, einfach nicht versteht, was es heißt, aufgrund seiner Überzeugungen das Martyrium auf sich zu nehmen.

³⁶ Das zweite Argument beruht direkt auf einer ethischen Überzeugung, das erste nicht. Trotzdem kann auch das erste innerhalb einer ethischen Argumentation höchste ethische Relevanz gewinnen.

³⁷ Wenn es zu praktischen Fragestellungen kommt, aggregieren deshalb auch Utilitaristen großzügig. Sehr großzügig z.T., wie wenn für ein ganzes Land oder gar die ganze Welt eine Nutzenfunktion verwandt wird, so etwa Nordhaus [1991] und [1996].

Eine Alternative würde noch die Behauptung bilden, bestimmte Institutionen (der Marktmechanismus z. B.) würden automatisch, ohne weiter Intervention und Gestaltung, zu dem – hier utilitaristisch – erwünschten Ergebnis der Zuordnung von Individuen zu bestimmten Positionen führen. Die hierzu verlangten unrealistischen Voraussetzungen – die im gegebenen Fall auch Ausgangsbedingungen betreffen würden – machen einen solchen Anspruch aber ebenfalls für praktische Zwecke hinfällig. Und selbst wenn man hier etwas optimistischer wäre: Ein entsprechender Nachweis wäre ein höchst anspruchsvolles Unternehmen, das von niemand bisher durchgeführt wurde.

chen Positionen, die wir einnehmen, in geeigneter Weise zumindest auch von unseren freien Entscheidungen abhängen – Grundlage der Forderung einer zumindest formalen Chancengleichheit wie auch etwa der Freiheit der Berufswahl. Eine Gesellschaft, die die endgültige Zuordnung von Individuen zu den Positionen, die sie einnehmen, aufgrund irgendeines – sei es auch ethischen Kriteriums – zu einer gesellschaftlichen, politischen Aufgabe machen würde, würde den Freiheitsraum ihrer Mitglieder in einer inakzeptablen Weise einschränken.³⁸

Das praktische Argument kritisiert die zu hohen informationellen Anforderungen des Utilitarismus. Auch wenn Rawls durchaus die geringeren informationellen Anforderungen seiner Prinzipien gegenüber dem Utilitarismus betont,³⁹ so baut er seine Widerlegung letztlich nicht darauf auf. Hier gilt wohl Ähnliches, wie für die Frage der Möglichkeit von Nutzenvergleichen:

Rawls ist zwar ohne Zweifel skeptisch, was die Möglichkeiten interpersoneller Nutzenvergleiche angeht: Sie sind entweder viel zu vage, um einen klaren Maßstab zur Beurteilung von Fragen der Gerechtigkeit geben zu können – dieselbe utilitaristische Formel kann zu diametral entgegengesetzter Beurteilung in bestimmten Fragen führen –, oder, um der Unbestimmtheit abzuweichen, werden nach moralischem Gesichtspunkt willkürliche – und auch sonst zweifelhafte – Annahmen gemacht, wie die, daß Nutzenabstände zwischen zwei Alternativen mit der individuell wahrgenommenen Zahl von Nutzenniveaus zwischen ihnen identifiziert werden sollten, oder die Null-Eins Regel angewandt auf Erwartungsnutzenfunktionen. [Vgl Rawls 1973, 90 f., 320-324] Rawls sieht deshalb – zum gegenwärtigen Zeitpunkt wenigstens – „no satisfactory answer to these difficulties from a utilitarian point of view.“ [Rawls 1973, 91] Aber Rawls nimmt nicht an, daß eine zufriedenstellende Lösung *unmöglich* ist, und möchte seine Kritik am Utilitarismus nicht auf diese Schwierigkeiten bauen: „The controversy about interpersonal comparisons tends to obscure the real question, namely, whether the total (or average) happiness is to be maximized in the first place.“ [Rawls 1973, 90] Ich verstehe diesen Satz so, daß Rawls sagen möchte, daß selbst dann, wenn die Frage interpersoneller Nutzenvergleiche zufriedenstellend gelöst werden könnte, keine Nutzensumme maximiert werden sollte, und daß dies die Hauptstoßrichtung seiner Kritik ist.

Diese Hauptstoßrichtung besteht – wie schon ausgeführt – nun eben darin, daß er meint zeigen zu können, daß im direkten Vergleich seiner Prinzipien mit dem Utilitarismus dieser im Urzustand deshalb abgelehnt würde, weil seine Prinzipien besser gegen die Gefahr schützen können, sich in einer Situation wiederzufinden, die nicht zu ertragen, inakzeptabel ist.

Um dieses Argument zu rekonstruieren, müssen wir also die formale Konstruktion des Urzustandes so ändern, daß die Parteien die Zuordnung bestimmter Individuen zu

³⁸ Es sei daran erinnert, daß es im Rawlsschen Selbstverständnis keineswegs notwendigerweise ein Problem ist, wenn über die Konstruktion des Urzustandes, inklusive der Formulierung der Aufgabenstellung, direkt ethische Überzeugungen umgesetzt werden. [Vgl. Rawls 1975, 634 f.] Die Konstruktion des Urzustandes dient gerade dazu, ethische Überzeugungen einzufangen, und z. B. für die Wahl eines geeigneten Spargrundsatzes geht eine ethische Überzeugung schon direkt in die Aufgabenstellung ein.

³⁹ Vgl. etwa die Ausführungen Rawls [1975, 111-115].

bestimmten Positionen bei ihrer Beurteilung möglicher Listen von in einer Gesellschaft vorhandenen Positionen berücksichtigen können. Außerdem müssen ihnen in der Entscheidungssituation auch schon Informationen zu den jeweils vorliegenden Nutzenfunktionen bzw. Präferenzen der Individuen vorliegen. Zur Vereinfachung nehmen wir an, daß die Präferenzen aller Individuen durch eine Nutzenfunktion repräsentiert werden können⁴⁰, und außerdem, daß gleichen Präferenzen auch die gleichen Nutzenfunktionen zugeordnet werden, d. h. wir nehmen an, es gebe eine Funktion U aus der Menge aller monoton steigenden Ordnungen R in die Menge der Funktionen von X in \mathbb{R} , so daß für jede Ordnung R $U(R)$ die für den utilitaristischen Kalkül relevante, R repräsentierende Nutzenfunktion ist. Diese letztere Annahme ist keineswegs typisch für den Utilitarismus, der über die Annahme der Kardinalität des Nutzens im Gegenteil Raum dafür schafft, Individuen mit gleichen Präferenzen über die Ergebnisse der Auflösung einer unsicheren Situation (die sicheren „Auszahlungen“) verschiedene Nutzenfunktionen zuzuordnen. Die vereinfachende Annahme ist aber für unsere Zwecke unproblematisch, da die Parteien auch bei einem weiteren Ansatz damit rechnen müssen, daß sie „zufälligerweise“ in einer Welt enden, in der entweder sowieso alle Individuen verschiedene Präferenzen haben oder alle Individuen mit gleichen Präferenzen die gleiche Nutzenfunktion. Wenn für den beschränkten Rahmen, wo dies immer der Fall sein wird, gezeigt werden kann, daß der Utilitarismus mit unseren Axiomen unvereinbar ist, dann ist er das auch für den weiteren.⁴¹

Gegeben diese vereinfachenden Annahmen sind nur wenige Änderungen im formalen Aufbau vorzunehmen:

Die Ordnung der Positionen in den Vektoren $a \in S$, r und ρ sind im folgenden relevant in dem Sinne, daß für alle $i \in N$ die Position a_i von dem Individuum eingenommen wird, das die Präferenzen r_i hat und die Menge der inakzeptablen Positionen ρ_i . Um die relevanten Daten über ein Individuum nach Auflösung der Unsicherheit zu kennzeichnen, reicht nun also ein Index i^* .

Wir nehmen jetzt also an, daß die Situation („Welt“) einer Partei nach Auflösung der Unsicherheit hinreichend durch einen Vektor $w = (S, a, r, \rho, i^*)$ beschrieben werden kann, wobei S , a , r , r_i und ρ wie bisher definiert sind, und $i^* \in N$ der Index sowohl der Position, die das Individuum dann ex post faktisch einnimmt, als auch der Präferenzordnung und der Menge von unerträglichen Situationen, die die Partei charakterisieren.

Für die Menge möglicher Welten $W(S, a)$, mit denen die Parteien – gegeben eine bestimmte Entscheidungssituation S und Entscheidung a – rechnen müssen, gilt nun alternativ:

⁴⁰ Eine Vereinfachung, die dem Utilitarismus entgegenkommt, indem sie die oben S. 187 erwähnte erste Schwierigkeit per Annahme aus dem Weg räumt.

⁴¹ Vorausgesetzt ist dabei, daß wie die Präferenzen so auch die Nutzenfunktionen – wie immer sie bestimmt werden mögen – ein Faktum über die Individuen darstellen.

Axiom 6.16.

$$\forall S \in \Sigma_B, a \in S : W(S, a) = \{(S, a, r, \rho, i^*) : \\ (\forall i \in N : r_i \text{ ist eine Ordnung über } X \wedge \forall x, x' \in X : x \geq x' \rightarrow xr_i x')\} \quad (6.21)$$

$$\wedge (\forall i \in N, x, x' \in X : x \in \rho_i \wedge xr_i x' \rightarrow x' \in \rho_i) \quad (6.22)$$

$$\wedge (i^* \in N) \quad (6.23)$$

$$\wedge (\exists b \in S : \forall i, j \in N : b_i \notin \rho_j) \quad (6.24)$$

$$\wedge (\forall x \in B : \exists j \in N : x \in \rho_j) \quad (6.25)$$

Axiom 6.9 wird zu:

Axiom 6.17.

$$\forall S \in \Sigma_B, a \in S, w \in W(S, a) : a_{i^*} \in \rho_{i^*} \rightarrow \\ [a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i \in N : b_i \notin \rho_i] \rightarrow \\ G_w(\text{irr})$$

Entsprechend wird Definition 6.4 zu:

Definition 6.12. *rawlsrational*

$\text{rrat} \leftrightarrow$

$$\forall S \in \Sigma_b, a \in S, w \in W(S, a) : a_{i^*} \in \rho_{i^*} \rightarrow \\ \neg[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i \in N : b_i \notin \rho_i]]$$

Es ist unmittelbar einsichtig, daß alternativ zu Theorem 6.3 gilt:

Theorem 6.16. *Gegeben die Definitionen 6.2 und 6.3 und die Axiome 6.17, 6.11 und 6.10 gilt:*

$$\forall S \in \Sigma_b, a \in S, w \in W(S, a) : a_{i^*} \in \rho_{i^*} \rightarrow \\ \neg[a \in C(S) \wedge [\exists b \in S : \forall w' \in W(S, b), i \in N : b_i \notin \rho_i]]$$

Wir definieren als nächstes die utilitaristische Ordnung $\succeq_{u(r)}$.⁴²

Definition 6.13.

$$\forall a, b \in X^n : a \succeq_{u(r)} b \leftrightarrow \sum_i U(r_i)(a_i) \geq \sum_i U(r_i)(b_i)$$

Von folgendem Axiom, das festlegt, daß es für zumindest ein Gut und gegebene Mengen der übrigen Güter ein Intervall gibt, aus dem alle Werte für Entscheidungssituationen möglich und nicht bekannterweise für bestimmte Individuen unerträglich sind, hängt das dann bewiesene negative Resultat ab:

⁴² Diese Ordnung wird – wie oben S. 191 diskutiert – von der Zuordnung von Positionen zu Individuen und deren Nutzenfunktionen bzw. Präferenzen abhängig sein.

Axiom 6.18.

$$\exists j \in M, s, s' \in \mathbb{R}, x \in X : [\forall t \in \mathbb{R} : s' \geq t \geq s \rightarrow \\ [\exists x' \in X \setminus B : \forall j' \in M : [j' \neq j \rightarrow x_{j'} = x_j] \wedge x_j = t]]$$

Theorem 6.17. Sei U eine Funktion aus der Menge aller monoton wachsenden Ordnungen R über X in die Menge aller reellwertigen Funktionen über X mit $U(R)(x) \geq U(R)(x') \leftrightarrow xRx'$. Gegeben Definitionen 6.2, 6.3 und 6.12 und Axiome 6.16 und 6.18 gilt:

$$[\forall S \in \Sigma_B : C(S) = \{a \in S : \forall b \in S : a \succeq_{u(r)} b\}] \rightarrow \neg \text{rrat}$$

Beweis. Betrachte j, s, s' und x für die der in Axiom 6.18 angenommene Zusammenhang gilt. Ohne Beschränkung der Allgemeinheit sei $j = 1$. Betrachte irgendeine streng monoton wachsende Ordnung R und $r = (R, \dots, R)$. Es muß dann auch $t, t' \in [s, s']$ geben, so daß für $x' = (s', x_2, \dots, x_m) \in X \setminus B$, $y = (t, x_2, \dots, x_m) \in X \setminus B$ und $y' = (t', x_2, \dots, x_m) \in X \setminus B$ gilt:⁴³

$$0 < U(r_1)(y') - U(r_1)(y) < \sum_{i=2}^n [U(r_i)(x') - U(r_i)(y')] = (n-1)(U(R)(x') - U(R)(y'))$$

Betrachte nun $S = \{a, b\} \in \Sigma = \Sigma_B$ mit $a = (y, x', \dots, x')$ und $b = (y', \dots, y')$ und nimm an $\forall S : C(S) = \{a \in S : \forall b \in S : a \succeq_{u(r)} b\}$. Dann ist konstruktionsgemäß $a \succeq_{u(r)} b$, d. h. $a \in C(S)$. Betrachte weiter $w^* = (S, a, r, \rho, 1) \in W(S, a)$ mit $\rho_i = \{x : U(R)(y) \geq U(R)(x)\}$ für alle $i \in N$, wobei dann $a_1 \in \rho_1$. w^* ist in der Tat Element von $W(S, a)$: Die Bedingungen 6.21, 6.22, 6.23 und 6.25 von Axiom 6.16 sind trivialerweise erfüllt. Außerdem gilt $\forall i, j : b_i \notin \rho_j$, so daß auch 6.24 erfüllt ist.

Weiterhin gilt wegen $\forall i : a_1 < b_i$ und Bedingung 6.24 von Axiom 6.16 für alle $w = (S, b, r, \rho, i^*) \in W(S, b) : \forall i, j : b_i \notin \rho_j$, d. h. insbesondere auch $\forall i : b_i \notin \rho_i$.

⁴³ Sei x^* ein x , für das der in Axiom 6.18 angenommene Zusammenhang für $j = 1$ gilt. Wenn wir uns auf $x \in X$ beschränken, für die $\forall i \neq 1 : x_i = x_i^*$ und $x_1 \in [s, s']$, dann können wir $U(R)(x)$ als eine monoton wachsende Funktion von x_1 betrachten. Sei also f eine im Intervall $[s, s']$ überall definierte streng monoton steigende reellwertige Funktion. Zu zeigen ist, daß es $t, t' \in [s, s']$ gibt, so daß $0 < f(t') - f(t) < (n-1)(f(s') - f(t'))$. Betrachte ein beliebiges $z \in (s, s')$. Falls $f(z) - f(s) < (n-1)(f(s') - f(z))$ wären s und z schon die gesuchten t, t' . Nimm also an $f(z) - f(s) \geq (n-1)(f(s') - f(z))$. Betrachte ein $n' \in \mathbb{N}$, so daß

$$n' > \frac{f(z) - f(s)}{(n-1)(f(s') - f(z))}$$

Da sonst die Summe über die einzelnen Abschnitte größer als der Gesamtabschnitt wäre, muß es dann zumindest ein $i \in \{1, \dots, n'\}$ geben, so daß

$$f(i \frac{z-s}{n'} + s) - f((i-1) \frac{z-s}{n'} + s) \leq \frac{f(z) - f(s)}{n'} < (n-1)(f(s') - f(z))$$

Mit $t' = i \frac{z-s}{n'} + s$ und $t = (i-1) \frac{z-s}{n'} + s$ für ein solches i gilt dann schließlich:

$$f(t') - f(t) < (n-1)(f(s') - f(z)) \leq (n-1)(f(s') - f(t'))$$

6 Der Urzustand: Ein Rekonstruktion

Zusammengenommen gilt also entgegen der Definition 6.12 von *rrat*:

$$\begin{aligned} \exists S \in \Sigma_B, a \in S, w \in W(S, a), i^* \in N : \\ a_{i^*} \in \rho_{i^*} \wedge a \in C(S) \wedge \exists b \in S : \forall w' \in W(S, b), i \in N : b_i \notin \rho_i \end{aligned}$$

□

Dieses Unmöglichkeitstheorem hängt, wie schon gesagt, von den Annahmen in Axiom 6.18 über die Eigenschaften von $X \setminus B$ ab. Allerdings handelt es sich um kein Charakterisierungsergebnis, d. h. es stellt nicht fest, was die hinreichenden und notwendigen Bedingungen dafür sind, daß Nutzenfunktionen möglich sind, die nicht dazu führen, daß über den utilitaristischen Kalkül bestimmte Entscheidungsmengen mit Rawlsrationalität unvereinbar sind. Es ist aber klar, daß es $X \setminus B$ gibt, für die das möglich ist, und so stellt sich die Frage, inwieweit diese Tatsache den Utilitarismus in einem besseren Licht erscheinen läßt.

Nimm also an, für ein bestimmtes X lasse sich für jede mit Axiom 6.16 vereinbare Kombination von Präferenzen r und Mengen inakzeptabler Positionen ρ ein System von Nutzenfunktionen $U(\cdot)$ finden, so daß $\forall S \in \Sigma : C(S) = \{a \in S : \forall b \in S : a \succeq_{u(r)} b\}$ mit *rrat* vereinbar ist. Betrachte nun eine beliebige Folge $\{x^h\}^{44}$ in $X \setminus B$ mit $\forall h : x^{h+1} \geq x^h$. Dann muß gelten:⁴⁵

$$\begin{aligned} \forall i, j \in N, h, h', h'' \in \mathbb{N} : h'' > h' > h \wedge x^h p_i x^{h'} \rightarrow \\ U(r_i)(x^{h'}) - U(r_i)(x^h) > (n-1)[U(r_j)(x^{h''}) - U(r_j)(x^{h'})] \quad (6.26) \end{aligned}$$

Betrachte nämlich beliebige $i, j, h, h', h'', r, \rho$, so daß $U(r_i)(x^{h'}) - U(r_i)(x^h) \leq (n-1)[U(r_j)(x^{h''}) - U(r_j)(x^{h'})]$ (*). Sei für entsprechende $i, j, h, h', h'', r, \rho$: $R = r_i, R' = r_j, r' = (R, R', \dots, R'), q = \{x \in X : x^{h'} P x\}, q' = \{x \in X : x^{h'} P' x\}, \rho' = (q, q', \dots, q'), a = (x^h, x^{h'}, \dots, x^{h''}), b = (x^{h'}, \dots, x^{h'})$ und $S = \{a, b\}$. Offensichtlich erfüllen dann mit $i^* \in N$ alle $w = (S, a, r', \rho', i^*)$ annahme- oder konstruktionsgemäß die Bedingungen von Axiom 6.16, insbesondere $\exists b \in S : \forall i, j : b_i \notin \rho'_i$ (Bedingung 6.24), da $\forall j : x^{h'} \notin b_j$. Wegen (*) wäre aber $a \in C(S)$ und gleichzeitig gilt konstruktionsgemäß $\forall j : a_1 \in \rho'$, d. h. für $i^* = 1$ gilt $a_{i^*} \in \rho_{i^*}$ bzw. zusammengenommen gegen die Definition von *rrat*:

$$\begin{aligned} \exists S \in \Sigma, a \in S, w \in W(S, a), i^* \in N : \\ a_{i^*} \in \rho_{i^*} \wedge a \in C(S) \wedge \exists b \in S : \forall w' \in W(S, b), i \in N : b_i \notin \rho_i \end{aligned}$$

Die Rawlsrationalität legt also dem System der Nutzenfunktionen ein enges und eigentümliches Netz von Bedingungen auf, das den Utilitarismus in einem fraglichen Licht erscheinen läßt. Abgesehen davon, daß, wie Theorem 6.17 zeigt, nur für spezielle

⁴⁴ Statt der üblichen tiefgestellten Indizes benutze ich hier um der Konsistenz mit dem restlichen Teil des Kapitels willen hochgestellte.

⁴⁵ p_i sei die auf übliche Weise definierte zu r_i gehörige strikte Präferenz.

$X \setminus B$ sich ein solches System der Nutzenfunktion konstruieren läßt, und damit der Utilitarismus abhängig davon wäre, daß wir den Raum möglicher Entscheidungssituationen geeignet einschränken, wäre der Utilitarismus in jedem Fall nicht vereinbar mit der Idee, daß wir die Nutzenfunktionen in einer von den ethischen Entscheidungen gewissermaßen unabhängigen Weise vorgängig zu diesen Entscheidungen – etwa aufgrund einer substantiellen Nutzenkonzeption – bestimmen. Im Gegenteil müßte das System der Nutzenfunktionen sorgfältig darauf hin abgestimmt werden, daß es die Verurteilung von Individuen zu möglicherweise inakzeptablen Positionen da vermeidet, wo es vermeidbar ist.

Ohne ein substantielles Nutzenkonzept oder jedenfalls ein den ethischen Entscheidungen vorausliegendes konstruktives Verfahren zur Bestimmung von Nutzenfunktionen ist es aber fraglich, ob man überhaupt sinnvollerweise von Utilitarismus sprechen kann, und man nicht vielmehr nur von der bloßen Repräsentierbarkeit von ethischen Entscheidungen in einer utilitaristischen Formel sprechen sollte.⁴⁶

Rawls selbst möchte jedenfalls nicht ausschließen, daß selbst seine Theorie der Gerechtigkeit mit entsprechend gewählten Nutzenfunktionen *ausgedrückt* werden kann – ohne daß dies jedoch die Gerechtigkeit als Fairneß zu einer utilitaristischen Theorie machen würde:

[...] justice as fairness does not deny that the idea of a utility function can be used to formulate justice as fairness. Indeed, I suppose any conception of justice can be expressed as maximizing a suitably doctored utility function. Surely the debate between utilitarian and nonutilitarian views is not about that trivial formal question! Such a utility function is but a mathematical representation that encodes certain basic features of our normative assumption. That there is a representation of this kind [...] is [not] the substantive idea of utility in the tradition of utilitarianism. [Rawls 2001, 107]

Zur Illustration dessen, *was* dem System von Nutzenfunktionen durch Bedingung 6.26 auferlegt wird, noch ein Zahlenbeispiel. Nehmen wir ein Land von der ungefähren Größe der Bundesrepublik Deutschland, genauer sei $n = 80000001$, und betrachten wir 10 Bündel von Grundgütern, die alle eins größer als das andere sind, formal eine (endliche) Folge $[x^h]_1^{10}$ mit $x^{h+1} > x^h$ für alle h . Sei R irgendeine über dies Folge streng monoton wachsende Präferenzordnung und $u_h = U(R)(x^h)$. Setze $u_1 = 0$ und $u_2 = 1$. Es muß dann gelten $u_{10} < 1,0000000125$, oder mit anderen Worten, der Nutzenabstand des (zehnt-)besten Bündels zum zweitebesten Bündel kann höchstens $0,0000125$

⁴⁶ In der Auseinandersetzung mit Harsanyis Verteidigung des Utilitarismus argumentieren Amartya Sen und in der Folge Autoren wie John Weymark und John Roemer (meineserachtens überzeugend), daß man ohne ein unabhängiges, substantielles Nutzenkonzept in der Tat nicht von Utilitarismus sprechen kann und es sich bei den von Harsanyi 1953, 1955 zuerst vorgelegten Theoremen nicht wirklich um axiomatische Begründungen des Utilitarismus handelt. Insbesondere handele es sich bei dem Theorem, das aus dem Paretoprinzip und daraus, daß sowohl individuelle wie soziale Präferenzen den Axiomen der Erwartungsnutzentheorien gehorchen, den Utilitarismus herleiten möchte, nur um ein Repräsentationstheorem. Vgl. etwa Harsanyi [1976b, 3-23] (für einen Neuabdruck Harsanyis Artikel von 1953 und 1955), Sen [1986, 1122-1124], Weymark [1991] und Roemer [1996, S. 138-150].

Promille betragen. Der Nutzenabstand des zehnten vom neunten Bündel dagegen, muß kleiner sein als $5,9605 \cdot 10^{-64}$.⁴⁷ Eine Version des Utilitarismus, die Nutzenfunktionen dieser Form vorschlägt, ist mir bisjetzt unbekannt.

Ob nun allerdings eine Version des Utilitarismus, die sorgfältig darauf abgestimmt wäre, Bedingung 6.26 zu erfüllen, noch gerechtfertigter Weise als Utilitarismus zu bezeichnen wäre oder nicht: Die Parteien im Urzustand werden in jedem Fall keine Version des Utilitarismus akzeptieren, der diese Bedingung nicht erfüllt. Vor die Wahl gestellt zwischen einem konstruktiven Verfahren, das Rawlsrationalität sicherstellt (wie etwa der Entscheidung auf der Basis der beschränkten Leximinordnung) und einem Utilitarismus, der entweder auf die Angabe eines konstruktiven Verfahrens zur Bestimmung der Nutzenfunktionen überhaupt verzichtet (und sich etwa mit der *Möglichkeit* von passenden Nutzenfunktionen zufriedengibt), oder ein Verfahren angibt, das nicht die Erfüllung der Bedingung 6.26 sicherstellt, werden die Parteien im Urzustand sich für ersteres entscheiden.

Für das obige Unmöglichkeitstheorem ist Bedingung 6.24 von Axiom 6.16 essentiell. Wenn wir sie so abschwächen, daß sie nur noch fordert $\exists b \in S : \forall i : b_i \notin \rho_i$, dann ist der Beweis von Theorem 6.17 so allgemein nicht mehr durchführbar, da nur noch dann bekannt ist, daß ein Vektor von Positionen für alle akzeptabel ist, wenn er alle anderen Vektoren aus $X \setminus B$ dominiert. Aber zumindest, wenn er auch alle anderen Vektoren aus X dominiert, wird er dann auch vom utilitaristischen Kriterium ausgewählt werden.

Im Geist der Abschwächung von Bedingung 6.24, wie überhaupt der utilitaristischen Überzeugung von der Möglichkeit gesellschaftliche Positionen den Trägern bestimmter Nutzenfunktionen zuzuordnen, wäre es wohl, dann auch als bekannt vorauszusetzen, zu welchen Nutzenfunktionen bekannterweise inakzeptable Positionen gehören.⁴⁸

Auch wenn wir das tun und weiter annehmen, daß um die obere Schranke von B herum in geeigneter Weise ein Gebiet zu X gehört, können wir nicht in zu Theorem 6.17 paralleler Weise allgemein ausschließen, daß es für alle entsprechend unserer Annahmen möglichen r Nutzenfunktionen gibt, die verhindern, daß trotz der Existenz von Vektoren $a \in (X \setminus B)$ Vektoren gewählt werden, die *bekannterweise* Individuen zu inakzeptablen Positionen verurteilen.

Allerdings legt auch hier die Bedingung der Rawlsrationalität den möglichen Nutzenfunktionen erhebliche und keineswegs – für Utilitaristen – selbstverständliche Beschränkungen auf, wie ich nur an einem Beispiel zeigen möchte.

Betrachte ein beliebiges $\alpha \in \mathbb{R}_+^m$ mit $\sum_j \alpha_j = 1$ und sei $X = \mathbb{R}_+^m$ und $B = \{x : \prod_j x_j^{\alpha_j} \leq 1\}$. Betrachte weiter die Präferenzordnungen $xR_1y \leftrightarrow \prod_j x_j^{\alpha_j} \geq \prod_j y_j^{\alpha_j}$ mit zugeordneter Nutzenfunktion U_1 und $xR_2y \leftrightarrow \sum_j \alpha_j x_j \geq \sum_j \alpha_j y_j$ mit zugeordneter Nutzenfunktion U_2 und $r^* = (R_1, R_2, R_2, \dots, R_2)$. Für jedes $a \in S$ und jedes $w \in W(S, a)$ mit $r = r^*$ muß offensichtlich gelten $B \subseteq \rho_1$. Nimm an $\exists x, y \in X \setminus B, z \in B : (n-1)(U_2(y) - U_2(x)) > U_1(x) - U_1(z)$. Konstruiere $a = (z, y, y, \dots, y)$, $b = (x, \dots, x)$, $S = \{a, b\}$ und ρ mit $\rho_1 = B$ und $\forall i \neq 1 : \rho_i = \emptyset$. Es gilt dann konstruktionsgemäß $a \in C(S) = \{a \in$

⁴⁷ Das ergibt sich aus sukzessiver Anwendung von Bedingung 6.26.

⁴⁸ Auch wenn wir das nicht annehmen wollen, ist zu bemerken, daß für bestimmte Konstellationen – wie etwa im unten ausgeführten Beispiel – von den Präferenzen und B aus dies erschlossen werden kann.

$S : \forall b \in S : a \succeq_{u(r)} b$. Für $w = (S, a, r^*, \rho, 1)$ sind außerdem alle Bedingungen von Axiom 6.16 (mit der abgeschwächten Version von 6.24) erfüllt, $a_1 \in \rho_1$ und wegen der abgeschwächten Version von 6.24 zusammen mit $a_1 \in \rho_1$ muß für alle $w \in W(S, b)$ mit $w = (S, b, r, \rho, i^*)$ gelten: $\forall i : b_i \notin \rho_i$. Zusammengenommen gilt also entgegen der Definition 6.12 von rrat :

$$\exists S \in \Sigma_B, a \in S, w \in W(S, a), i^* \in N : \\ a_{i^*} \in \rho_{i^*} \wedge a \in C(S) \wedge \exists b \in S : \forall w' \in W(S, b), i \in N : b_i \notin \rho_i$$

Das Beispiel zeigt, wie auch hier die Bedingung der Rawlsrationalität dem System der *möglichen* Nutzenfunktionen ein – allerdings weniger enges – Netz von Bedingungen auferlegt, wie insbesondere die Abstände (in Nutzeneinheiten) zwischen bekannterweise inakzeptablen Positionen zu Positionen, die nicht bekannterweise inakzeptable sind, eine obere Schranke für Nutzen etablieren. Dieses Netz betrifft alle Nutzenfunktionen, wenn wir anders als in unserem Beispiel davon ausgehen, daß die Zuordnung von Nutzenfunktionen und bekannterweise inakzeptabler Positionen ex ante bekannt ist. In analoger Form bleiben damit die oben S. 194 angeführten Argumente gegen den Utilitarismus gültig. Insbesondere werden die Parteien im Urzustand sich gegen den Utilitarismus aussprechen, da ihnen ein konstruktives Verfahren zur Verfügung steht, rawlsrational zu entscheiden, während der Utilitarismus bestenfalls das – von niemand unternommene – Projekt, ein der Bedingung der Rawlsrationalität entsprechendes System von Nutzenfunktionen zu entwickeln, in Aussicht stellt.

6 *Der Urzustand: Ein Rekonstruktion*

7 Das RICE-Modell

7.1 Einleitung

Im folgenden Kapitel möchte ich an einem Beispiel zeigen, zu welche unterschiedlichen normative Ergebnissen verschiedene ethische Ansätze führen können.

Als Beispiel verwende ich das RICE-Modell, das William Nordhaus (zusammen mit seinen Mitarbeitern) beginnend mit Nordhaus u. Yang [1996] in verschiedenen Versionen vorgelegt hat. Genauer benutze ich die als Exceltabelle vorliegende RICE-2001 Version [Nordhaus u. Boyer 2001], allerdings mit dem Datenbestand der Tabelle Nordhaus u. Boyer [2000a], die die letzten ausführlichen Darstellungen des Modells Nordhaus u. Boyer [2000b] bzw. Nordhaus u. Boyer [2000c] begleitet.

Da es mir um eine Untersuchung der Konsequenzen der gewählten ethischen Ansätze geht, werde ich dabei die „deskriptiven“ Komponenten des Modells nicht kritisch hinterfragen, sondern sie im Gegenteil fast durchgängig voraussetzen.¹

Insbesondere werde ich das integrierte Klimamodell, das ökonomische Produktionsmodell, die Aufteilung der Welt in acht Regionen² und die ökonomischen Verhaltensannahmen unverändert lassen. Im ersten Unterabschnitt des Kapitels und in Teilen des zweiten Unterabschnitts werde ich auch die Modellierung der Klimaschäden unverändert übernehmen, dann werde ich aber die Schadensfunktion näher untersuchen, da sich darin, wie noch zu zeigen sein wird, erhebliche normative Annahmen verstecken.

7.2 Effizienz und Utilitarismus

Nordhaus legt seinen Ausführungen eine etwas eigenwillige Mischung verschiedener normativer Kriterien zugrunde: Eine gute Dosis Utilitarismus wird ergänzt durch Elemente der Kosten-Nutzen Literatur und ein gehöriges Maß paretianischer Rhetorik. Während Nordhaus es vermeidet, explizit auf den utilitaristischen Charakter seines Unternehmens hinzuweisen, verwendet er ausgiebig die Sprache der paretianischen Wohlfahrtsökonomie: Insgesamt 98mal spricht er von Effizienz oder nennt etwas effizient oder ineffizient, 17mal bezieht er sich auf die Paretooptimalität einer Politik. Da aber – wie so häufig – gegenläufige Verteilungskonsequenzen der verschiedenen beurteilten umweltpolitischen Strategien dazu führen, daß das Paretokriterium – trotz des

¹ Damit möchte ich allerdings nicht nahelegen, daß die deskriptiven Elemente des Modells über Kritik erhaben wären.

² USA, China, OECD-Europa, andere Länder mit hohem Einkommen, Russland und Osteuropa, Länder mit mittlerem Einkommen, OPEC-Länder mit hohem Einkommen, Länder mit niedrigem mittlerem Einkommen, Länder mit niedrigem Einkommen. [Nordhaus u. Boyer 2000c, 27-38]

hohen Aggregationsgrads – nur sehr wenige eindeutige normative Aussagen zulässt,³ ist Nordhaus aber in der Tat auf weit darüber hinausgehende normative Annahmen angewiesen.

Die zentrale Passage [Nordhaus u. Boyer 2000c, 14-16], in der Nordhaus explizit seine ethischen Voraussetzungen diskutiert, scheint zunächst einmal ganz in die Richtung eines utilitaristischen Ansatzes zu weisen:

A central organizing framework of the DICE/RICE models is that the purpose of economic and environmental policies is to improve the living standards or consumption of people now and in the future. The relevant economic variable is *generalized consumption*, which denotes a broad concept that includes not only traditional market purchases of goods and services like food and shelter but also non-market items such as leisure, cultural amenities, and enjoyment of the environment.

The fundamental assumption we adopt is that policies should be designed to optimize the flow of generalized consumption over time. This approach rests on the view that more consumption is preferred to less. Moreover, increments of consumption become less valuable as consumption levels increase. In technical terms, we model these assumptions by assuming that regions maximize a social welfare function that is the discounted sum of the population-weighted utility of per capita consumption. This social welfare function is a mathematical representation of three basic value judgments: (i) higher levels of consumption have higher worth; (ii) there is diminishing marginal valuation of consumption as consumption increases; and (iii) the marginal social utility of consumption is higher for the current generation than for a future generation of the same size with the same per capita consumption.

RICE adds a significant level of complexity to the original DICE model by incorporating the simultaneous growth paths of different regions. The exact objective function, or criterion to be maximized, for region J is:

$$W_J = \sum_t U[c_J(t), L_J(t)]R_J(t)$$

where W_J is the objective function of region J , $U_J[c_J(t), L_J(t)]$ is the utility of consumption for region J , $c_J(t)$ is the flow of consumption per capita during period t , $L_J(t)$ is the population at time t , and $R(t)$ is the pure time preference discount factor.

[...]

³ Unter Voraussetzung der von Nordhaus angenommenen regionalen Zielfunktionen ist nach den von ihm publizierten Zahlen z. B. von den elf alternativen Strategien, für die Nordhaus regional disaggregierte Daten anführt, nur die der Beschränkung des Temperaturanstiegs auf den aufgrund des Kyoto-Protokolls zu erwartenden eine Paretoverbesserung gegenüber dem Referenzfall (keine Klimapolitik). Davon kann man sich leicht anhand der Tabellen 7.1 und 7.2 [nach Nordhaus u. Boyer 2000c, 123-127, 131, 147-148 u. 160] überzeugen.

The next set of equations represents the different regions. The first equation is the definition of utility, which was described and motivated in the previous subsection.

Utility represents the current value of economic well-being and is assumed equal to the size of population $[L_J(t)]$ times the utility of per capita consumption $u[c_J(t)]$. Equation (2.3) [das ist die folgende, AG] uses the general case of a power function to represent the form of the utility function:

$$U_J[c_J(t), L_J(t)] = L_J(t) \{c_J(t)^{1-\alpha} - 1\} / (1 - \alpha)$$

Der utilitaristische Charakter dieser Ausführungen ist offensichtlich: Die Zielfunktion einer jeden Region ist die über die Zeit abdiskontierte Summe des mit der Zahl der Einwohner multiplizierten Nutzens des Durchschnittskonsums, die man als einen (aggregierenden) Proxy der Summe des (intertemporalen) Nutzens aller Einwohner verstehen kann.

Seltsamerweise bleibt Nordhaus aber hier stehen und geht weder in Kapitel 2, in dem er den Aufbau seines Modells entwickelt, noch in Anhang A, wo er die Gleichungen seines Modells auflistet, weiter zu einer weltweiten Zielfunktion, die er ja brauchen würde, um verschiedene Politiken, die nicht in einem Verhältnis der Paretodominanz stehen, zu vergleichen. Dies ist umso sonderbarer, als er in der zitierten Passage von *den* normativen Voraussetzungen bzw. Werturteilen spricht, die *er* (zusammen mit seinem Mitautor) annimmt – von denen man deshalb erwarten würde, daß sie helfen, die verschiedenen untersuchten Strategien zu beurteilen. Die regionalen Zielfunktionen als solche helfen dazu aber in keiner Weise, vielmehr wird ihre Maximierung – gegeben die jeweilige Umweltstrategie – durchgehend vorausgesetzt. Es handelt sich bei ihnen also nicht um Kriterien der Autoren, sondern vielmehr um (durchgängig erfüllte) *Verhaltensannahmen* für die Regionen. Diese setzen in jedem Szenario ihre Politikvariablen so, daß sie die ihnen zugeordnete Zielfunktion maximieren. Der Fall liegt nicht anders als sonst in der Social Choice Literatur: Von den handelnden Akteuren (hier eben Regionen) wird angenommen, daß sie aufgrund ihrer Rationalität eine Zielfunktion (ihren Nutzen oder hier ihre Wohlfahrtsfunktion) maximieren. Die Aufgabe der „sozialen Wahl“ besteht dagegen darin, verschiedene Konstellationen von den Akteuren entzogenen Variablen – etwa aufgrund einer sozialen Wohlfahrtsfunktion – zu beurteilen.⁴

Faktisch benutzt Nordhaus als weltweite Wohlfahrtsfunktion nicht die Summe der regionalen Zielfunktionen sondern die Summe der regionalen intertemporalen Summen des abdiskontierten Konsums, mit regional spezifischen und über die verschiedenen Strategien identischen Diskontfaktoren. Diese Summe ist es, mit der Nordhaus die „wirtschaftlichen Auswirkungen“ („economic impact“) verschiedener Strategien mißt, und die er allen seinen Vergleichen verschiedener Strategien zugrundelegt. Ihre Gewichte bestimmen es, ob Nordhaus die Auswirkungen auf die einzelnen Regionen als „groß“ oder „klein“ bezeichnet:

⁴ In einem etwas anderen Kontext wird diese Parallelität von den Autoren selbst erwähnt. [Nordhaus u. Boyer 2000c, 109]

The net economic impact of a policy is the sum across regions of the present value of consumption under that policy minus the present value of consumption in the base case. The present values are computed using the base case discount factors.

Formal benutzt Nordhaus also als Kriterium die – von ihm allerdings als Zielfunktion nicht explizit angegebene – Wohlfahrtsfunktion:⁵

$$W = \sum_J \sum_t C_J(t) DF_J(t) = \sum_J \sum_t L_J(t) c_J(t) DF_J(t)$$

Abhängig von den regionalen Diskontfaktoren geht also der Periodendurchschnittskonsum der verschiedenen Regionen mit unterschiedlichen Gewichten in die Wohlfahrtsfunktion ein: Faktisch – entgegen dem zunächst explizit aufgezähltem Werturteil (ii) (s. o. S. 200), demgemäß die marginale Gewichtung des Konsums mit wachsendem Konsum abnehmen sollte – werden bei Nordhaus die wesentlichen höheren Durchschnittskonsumwerte etwa der USA auch deutlich höher gewichtet als etwa die Durchschnittskonsumwerte der ärmsten betrachteten Region.⁶

Darüber, was Nordhaus zu diesem Bruch⁷ im normativen Aufbau seines Ansatzes führt, kann man nur spekulieren. Er selbst stellt ja in den von ihm veröffentlichten Exceltabellen eine utilitaristische Alternative zur Verfügung, die allerdings die Höhergewichtung der Konsums in der Anfangsperiode reicherer Regionen beibehält: Alternativ zu den Gegenwartswerten des Konsums berechnet er auch jeweils eine Nutzen-summe, die von ihm so kalibriert wurde, daß sie im Referenzfall genau mit seinem Gegenwartswert des Konsums übereinstimmt. Aufaddiert ergäbe das ein weltweite, utilitaristische Wohlfahrtsfunktion⁸:

$$W = \sum_J \psi_J W_J + \phi_J = \sum_J \psi_J \sum_t L_J(t) u[c_J(t)] R_J(t) + \phi_J$$

Als Gewichte ψ_J hat Nordhaus den Pro-Kopf-Konsum in der ersten Periode des Referenzfalls gewählt. Zusammen mit der gewählten Nutzenfunktion $ln(c_J(t))$ und der von Nordhaus vorgenommenen Bestimmung der Diskontfaktoren führt das dazu, daß die Differenzen der Werte dieser utilitaristischen Wohlfahrtsfunktion zu deren Werten im Referenzfall eine Annäherung an die entsprechenden Differenzen für die von Nordhaus verwendete Wohlfahrtsfunktion darstellen. Die beiden Ansätze führen also

⁵ Wir lassen dabei die von Nordhaus subtrahierte Konstante (der Wert der Wohlfahrtsfunktion im Referenzfall) als normativ irrelevant außer Acht.

⁶ 1,2fach in der zweiten Periode bis 6,8fach in späten Perioden.

⁷ Zwar handelt es sich bei den in die weltweite Zielfunktion eingehenden regionalen Terme um Näherungen der regionalen Zielfunktionen, aber eben nicht um diese selbst.

⁸ In einer früheren als Arbeitspapier veröffentlichten Version ihres Buches haben Nordhaus und Boyer diese Wohlfahrtsfunktion – ohne die irrelevanten additiven Terme – noch explizit als Zielfunktion aufgeführt. [Nordhaus 1999, 15] Über den Grund der späteren Auslassung kann man wieder nur spekulieren. Ein guter Grund wäre natürlich gewesen, daß sie faktisch nicht diese Funktion maximieren. Allerdings wäre es dann doch angemessener gewesen, nun die tatsächliche verwendete Zielfunktion anzugeben.

zu ähnlichen normativen Konsequenzen bzw. zu genau denselben in den betrachteten Fällen.⁹

Hätte Nordhaus diese utilitaristische Wohlfahrtsfunktion seinen Ausführungen zugrundegelegt, wären zwei Dinge offensichtlicher geworden: Daß es sich bei den von ihm in Dollar ausgewiesenen Summen nicht einfach um „harte“ und sozusagen wertfreie Größen handelt, sondern um den Ausdruck einer von Nordhaus vorgenommenen ethischen Bewertung des Konsums der Bewohner verschiedener Regionen, und daß er – anders als es seine expliziten ethischen Überlegungen vermuten lassen – Nutzeneinheiten aus verschiedenen Regionen nicht gleich sondern sehr verschieden (mit dem Pro-Kopf-Konsum in der ersten Periode des Referenzfalls) gewichtet. Der Nutzen des Pro-Kopf-Konsums einer Periode in den USA geht so etwa mit dem 56fachen Gewicht des entsprechenden Nutzens in der ärmsten Region in die Wohlfahrtssumme ein.

Eine Begründung für die von Nordhaus gewählte Zielfunktion könnte noch der Hinweise darauf sein, daß diese mögliche Kompensationen ausweise. Wenn damit gemeint ist, daß eine Situation, die es ermöglicht, die in ihr gegenüber einer anderen Situation schlechter gestellten zu entschädigen, besser ist als die letztere und zwar unabhängig von der Durchführung dieser Kompensationen, dann bleibt – unabhängig von den allgemeinen Problemen dieses Kompensationskriteriums – die Frage, weshalb dieser Wechsel in der normativen Zielsetzung gegenüber dem zunächst dargestellten ethischen Ansatz vorgenommen wird, unbeantwortet. Wenn damit nur gemeint ist, daß die Möglichkeit von Kompensationen auf mögliche Paretoverbesserungen hinweise, ohne in sich schon eine Ordnung verschiedener Strategien zu bedeuten, dann sind die von Nordhaus aus seiner weltweiten Zielfunktion abgeleiteten normativen Urteile mit dem Kompensationskriterium eben nicht begründet (das Argument weist also, soweit es um diese Urteile geht, ins Leere).

Außerdem ist diese Form des Kompensationsarguments insoweit irreführend, als sie wiederum eine Wertfreiheit suggeriert, die nicht gegeben ist: Selbst wenn man annähme, daß – im weiten Sinn – verlustfreie monetäre Transfers zwischen den Regionen möglich sind (eine sicher problematische Annahme), folgt (wegen der unterschiedlichen Gewichtung des Periodenkonsums in verschiedenen Regionen) daraus, daß die Zielfunktion in der Strategie A einen höheren Wert als in der Strategie B ausweist, noch nicht, daß Kompensationen in der Höhe der Differenz als *monetäre* Transfers auch möglich sind. Dazu müßte vielmehr ein Schema von monetären Transfers in den verschiedenen Perioden vorgelegt werden, so daß die Kombination von Strategie B mit diesem Transferschema von allen Regionen als besser bewertet werden, als Strategie A. Diese Überlegungen machen deutlich, daß das betrachtete Kompensationsargument als solches einfach die verlustfreie Übertragung der *intertemporalen Wohlfahrt* voraussetzt und damit einem einfachen Vergleich der weltweiten Wohlfahrt auch entsprechend einer utilitaristischen Wohlfahrtsfunktion nichts voraus hat.¹⁰

⁹ Wovon man sich durch einen Vergleich der Tabellen 7.3 und 7.4 leicht überzeugen kann.

¹⁰ Einem Teil dieses Arguments könnte man dadurch begegnen wollen, daß man die regionalen Terme der weltweiten Zielfunktion als Näherungen der eigentlichen regionalen Zielfunktionen versteht. Da Nordhaus den Wert dieser „eigentlichen“ Zielfunktion aber berechnet, bliebe offen, warum er sie nicht benützt. Außerdem würde damit die Annahme der Möglichkeit verlustfreier Wohlfahrtstransfers ganz

Der von Nordhaus u. Boyer [2000c, 14-16] entwickelte ethische Ansatz hätte seine konsequente Fortsetzung in einer weltweiten utilitaristischen Wohlfahrtsfunktion mit gleichen Gewichten gefunden, also etwa (ohne die irrelevanten additiven Konstanten):

$$W = \sum_J \psi_1 W_J = \sum_J \psi_1 \sum_t L_J(t) u[c_J(t)] R_J(t)$$

Wie zu erwarten, führt eine solche egalitäre Gewichtung der Nutzen zu einer sehr verschiedenen Bewertung der betrachteten Alternativen. In den Tabellen 7.3 und 7.5 finden sich für die vierzehn von Nordhaus betrachteten Strategien (Tabelle 7.1) der Zuzugewinn (oder Verlust) gegenüber dem Referenzfall („business as usual“, d. h. keine international koordinierte Klimapolitik) zum einen auf der Basis des abdiskontierten Konsums in eigener Berechnung¹¹, zum anderen auf der Basis der angegebenen egalitären utilitaristischen Wohlfahrtsfunktion.

0	Referenzfall
1	Optimum
2	Optimum ab 2005
3	1990-Emissionen
4	Doppelte CO ₂ -Konzentration
5	Temperaturlimit: 2,5 °C
6	Temperaturlimit: 1,5 °C
7	Geoengineering
8	Kyoto-CO ₂ -Konzentration
9	Kyototemperatur
10	Kyoto mit globalem Zertifikatshandel
11	Kyoto mit Zertifikatshandel zwischen Annex-I Regionen
12	Kyoto mit Zertifikatshandel zwischen OECD Regionen
13	Kyoto ohne Zertifikatshandel

Tabelle 7.1: Vierzehn Umweltstrategien nach Nordhaus u. Boyer [2000c].

Die wichtigste Veränderung beim Übergang zu einer egalitären utilitaristischen Wohlfahrtsfunktion ist zunächst, daß die – nach dem Sonderfall des Geoengineering¹² – beste Strategie nicht mehr das „Optimum“ ist, sondern das Kyoto-Protokoll in seiner aktuellen Fassung. Das „Optimum“ kommt nicht mehr an zweiter, sondern nur noch an sechster Stelle und wird von allen vier betrachteten Varianten des Kyoto-Protokolls

offensichtlich.

Auch unabhängig vom Kompensationsargument könnte man die von Nordhaus vorgenommenen Beurteilungen als auf Näherungswerten basierend verteidigen wollen. Wenn Nordhaus aber tatsächlich seine Zielfunktion nur als Näherung der eigentlichen, utilitaristischen Zielfunktion verstände, stellt sich von neuem die Frage, weshalb er diese Näherungswerte und nicht die Ergebnisse der eigentlichen Zielfunktion, die er ja berechnet, ausweist.

¹¹ In die Berechnung gehen auch pragmatische Entscheidungen ein, die etwa die Zahl der Iterationen, den gewählten Grad der Zielerreichung bei der Optimierung oder die Parameter von Hilfsschadensfunktionen betreffen. Meine Ergebnisse weichen deshalb leicht von den Ergebnissen Nordhaus' in Tabelle 7.2 ab.

¹² Hier wird das Umweltproblem einfach wegdefiniert, indem eine kostenlose Möglichkeit der Beseitigung aller Umweltschäden angenommen wird. Vgl. Nordhaus u. Boyer [2000c, 126 f.].

Strategie	1	2	3	4	5	6	7	8	9	10	11	12	13
USA	22	-	-946	-305	-885	-	82	12	10	-78	-313	-692	-833
OHI	26	-	-139	-6	-131	-	-391	10	9	3	-15	-2	-26
EU	126	-	258	162	121	-	1943	47	36	47	46	161	54
EE	-9	-	-359	-64	-191	-	-110	-1	0	29	113	3	3
MI	19	-	-300	-103	-304	-	620	8	7	11	11	17	18
LMI	5	-	-512	-122	-341	-	549	6	5	13	13	20	21
China	-10	-	-425	-74	-226	-	-21	0	1	5	4	6	6
LI	20	-	-597	-174	-458	-	1228	12	9	19	21	32	34
Summe	198	192	-3021	-684	-2414	-26555	3901	94	77	49	-121	-455	-723

Tabelle 7.2: Wohlfahrtsgewinn gegenüber Referenzfall: abdiskontierter Konsum. [Nach Nordhaus u. Boyer 2000c, 131 u. 160]

übertroffen. Viele der von Nordhaus angeführten normativen Ergebnisse werden ungültig.

Es gilt nicht mehr:

From bad to worst we would rank Kyoto (Annex I trading), Kyoto (OECD trading), limiting CO₂ concentrations to twice pre-industrial levels, Kyoto (no trading), limiting climate change to a 2.5 °C temperature rise, stabilizing global emissions at 1990 levels, and limiting climate change to a 1.5 °C temperature rise [...]. [Nordhaus u. Boyer 2000c, 175]

Sondern eher¹³:

From bad to worst we would rank the optimal policy from 2005, limiting the global temperature to that resulting from the Kyoto protocol, doing nothing (the base case), limiting CO₂ concentrations to twice pre-industrial levels, limiting climate change to a 2.5 °C temperature rise, stabilizing global emissions at 1990 levels, and limiting climate change to a 1.5 °C temperature rise.¹⁴

Während unter dem Nordhauschen Kriterium drei der Kyoto-Varianten eine Verschlechterung gegenüber dem Referenzfall bedeuten, führen jetzt alle Kyoto-Varianten zu einer Verbesserung. Unter Verwendung einer egalitären Nutzenfunktion bestehen sie also einen Kosten-Nutzen Test, anders als bei Nordhaus:

Policies which have near-term carbon taxes in the \$100 per ton range, such as those associated with the Kyoto-Protocol, are almost sure to fail a cost-benefit test because they impose excessive near-term abatement. [Nordhaus u. Boyer 2000c, 175]

¹³ Wenn wir beim gleichen Rang beginnen wie Nordhaus.

¹⁴ Die drei schlechtesten Fälle sind also gleichgeordnet.

Strategie	0	1	2	3	4	5	6	7	8	9	10	11	12	13
USA	0	22	21	-947	-305	-883	-8478	82	12	11	-79	-313	-694	-833
OHI	0	26	26	-140	-6	-131	-2518	-391	11	9	3	-15	-2	-25
EU	0	126	120	257	163	115	-2801	1942	48	36	47	46	161	54
EE	0	-9	-7	-359	-64	-187	-1889	-110	-1	-0	29	112	3	4
MI	0	19	17	-300	-103	-304	-2888	620	9	7	11	11	17	18
LMI	0	5	5	-512	-122	-337	-2786	549	6	5	13	13	20	21
China	0	-10	-8	-424	-74	-222	-2175	-21	-0	1	5	4	6	6
LI	0	20	18	-596	-173	-454	-3538	1228	12	9	19	21	32	33
Summe	0	198	191	-3021	-685	-2403	-27074	3898	96	78	49	-122	-458	-723

Tabelle 7.3: Wohlfahrtsgewinn gegenüber Referenzfall: abdiskontierter Konsum. Eigene Berechnung.

Strategie	0	1	2	3	4	5	6	7	8	9	10	11	12	13
USA	0	22	21	-958	-311	-910	-9061	80	12	11	-79	-313	-699	-840
OHI	0	26	26	-141	-6	-132	-2600	-394	11	9	3	-15	-2	-25
EU	0	126	120	253	161	113	-2871	1898	47	36	47	46	161	54
EE	0	-10	-8	-366	-65	-194	-2033	-110	-1	-0	29	112	3	4
MI	0	19	17	-304	-105	-313	-3080	613	9	7	11	11	17	18
LMI	0	5	4	-518	-124	-345	-2931	541	6	5	13	13	20	21
China	0	-10	-8	-434	-76	-232	-2383	-21	-1	1	5	4	6	6
LI	0	20	18	-606	-177	-470	-3796	1196	12	9	19	21	32	33
Summe	0	197	190	-3074	-703	-2482	-28755	3802	96	78	49	-123	-463	-731

Tabelle 7.4: Wohlfahrtsgewinn gegenüber Referenzfall: ungleich gewichteter Nutzen. Eigene Berechnung.

Strategie	0	1	2	3	4	5	6	7	8	9	10	11	12	13
USA	0	22	21	-958	-311	-910	-9061	80	12	10	-79	-313	-699	-840
OHI	0	25	26	-139	-7	-131	-2564	-388	10	9	3	-15	-2	-25
EU	0	147	141	297	187	133	-3364	2223	56	41	55	53	188	63
EE	0	-102	-80	-3898	-692	-2065	-21633	-1173	-7	-1	314	1191	35	38
MI	0	87	80	-1406	-486	-1449	-14266	2840	40	29	50	50	78	81
LMI	0	59	52	-5971	-1433	-3975	-33779	6233	68	55	148	150	229	240
China	0	-483	-393	-20579	-3592	-10968	-112866	-1000	-24	33	259	189	263	281
LI	0	1101	981	-33933	-9938	-26319	-212657	67031	683	500	1091	1152	1784	1867
Summe	0	857	827	-66587	-16272	-45683	-410190	75845	838	677	1841	2456	1876	1706

Tabelle 7.5: Wohlfahrtsgewinn gegenüber Referenzfall: gleichgewichteter Nutzen. Eigene Berechnung.

Die negative Beurteilung des Kyoto-Protokolls bei Nordhaus hat keine Basis mehr, so wenn er schreibt:

First, we conclude that the Kyoto Protocol has no economic or environmental rationale. The approach of freezing emissions for a subgroup of countries is not related to a particular goal for concentrations, temperature, or damages. Nor does it bear any relation to an economically oriented strategy that would balance the costs and benefits of greenhouse-gas reductions. [Nordhaus u. Boyer 2000c, 177]

Auch die Ordnung der Kyoto-Varianten verändert sich. Bei Nordhaus war die Variante mit globalem Zertifikatshandel die beste [Nordhaus u. Boyer 2000c, 177]. Jetzt wird sie vom vorliegenden Protokoll wie auch dem Fall des Handels nur unter OECD-Ländern übertroffen.¹⁵ Auch das Resümee aus dem Vergleich dieser Varianten gilt natürlich nicht mehr:

These calculations emphasize that efficient design of the policy should be the major concern of policymakers. [Nordhaus u. Boyer 2000c, 177]

Tabelle 7.6 zeigt die Wohlfahrtsgewinne gegenüber dem Referenzfall berechnet auf der Basis eines gleichgewichteten Konsums. Nordhaus hat ja, entgegen seinem normativen Bekenntnis zu einer Wohlfahrtsfunktion, die höheren Konsum geringer gewichtet, Konsum nicht einmal gleichgewichtet, sondern faktisch den höheren Konsum z. B. der Vereinigten Staaten höher gewichtet als z. B. den geringen Konsum der ärmsten Länder (s. o. S. 200, 202). Der Wechsel von ungleich gewichtetem zu gleich gewichtetem Konsum hat (erwartungsgemäß) weniger dramatische Folgen als der Wechsel zu einem gleichgewichteten Nutzen. Vor allem erscheinen die „optimale“ und die leicht verzögerte „optimale“ Strategien jetzt wieder in der Tat als die nach der Magie des Geoengineering besten Strategien.

Immerhin aber verbessert sich das Abschneiden der verschiedenen Versionen des Kyoto-Protokolls: Auch das Kyoto-Protokoll in seine aktuellen Fassung bedeutet nun eine Verbesserung gegenüber dem Referenzfall. Kyoto mit globalem Zertifikatshandel (10) erscheint nun besser als die „ökologisch bedeutungsvolleren“ Strategien 8 und 9, die doch 10 an „when efficiency“ übertreffen [Nordhaus u. Boyer 2000c, 149]. Kyoto ohne Zertifikatshandel schließlich erscheint nun deutlich besser als das Ziel einer Verdoppelung der CO₂ Temperatur.

¹⁵ Nordhaus u. Boyer [2000c, 148] betonen: “It should be emphasized that global trading [...] is actually a radical extension of the Kyoto Protocol and contains crucial and problematical assumptions about the behavior of non-Annex I countries. In principle, each non-Annex-I country will be better off by agreeing to this limit-and-trade procedure – it can do no worse than simply consuming its permits and can do better by reducing its low marginal-cost sources and selling the permits at the world price.” Im Kontext könnte man das so (miß-)verstehen, als wollten die Autoren behaupten, „im Prinzip“ wäre Kyoto mit globalem Handel für jedes „non-Annex-I country“ auch dem tatsächlichen Kyoto-Protokoll überlegen. Das ist nicht der Fall, wie ein Blick auf Tabelle 7.2 zeigt. Das Argument zeigt nur die Attraktivität des Handels gegenüber dem Referenzfall.

Ein ganz anderes Bild ergibt sich auch in Bezug auf die Frage, wer der größte Gewinner der verschiedenen Formen von Klimapolitik sind. Bewertet mit einem gleichgewichteten Nutzen ist das nicht mehr generell Westeuropa [Nordhaus u. Boyer 2000c, 131, 162]. Für die beiden Optimalstrategien, Geoengineering und alle Varianten des Kyoto-Protokolls übertrifft der Nettogewinn der ärmsten Länder bei weiten den Nettogewinn Westeuropas. Unter den genannten Strategien wird nur für das eigentliche Kyoto-Protokoll der Nettogewinn der ärmsten Länder durch den Nettogewinn Osteuropas noch übertroffen.

7.3 Eine rawlsianische Perspektive

Die Ausführungen im vorangehenden Abschnitt sollten klar gemacht haben, in welchem hohem Maß die normativen Ergebnisse des RICE-Modells von den sehr spezifischen Bewertungsmaßstäben abhängen, die Nordhaus zugrunde gelegt hat: Bewertungsmaßstäbe, die zu den ethischen Maßstäben, zu denen er sich bekennt, nicht gut passen.

In diesem Abschnitt möchte ich jetzt die Perspektive wechseln und untersuchen, wie eine rawlsianische Interpretation der Ergebnisse des Modells aussehen könnte.

Eine direkte und einfache „Anwendung“ der Rawls'schen Theorie der Gerechtigkeit auf die Fragestellungen des RICE-Modells ist leider nicht möglich. Das Modell wurde ja ganz auf seine utilitaristische Bewertbarkeit hin konstruiert, und nicht im Blick auf mögliche rawlsianische Maßstäbe. Ebenso hatte Rawls bestimmte Fragestellungen nicht im Blick, die durch das Problem des Klimawandels aufgeworfen werden.¹⁶ Es sind deshalb auch weniger eindeutige und einfache normative Ergebnisse zu erwarten, als man sich das vielleicht wünschen könnte. Es wird eher darum gehen, die Akzente deutlich zu machen, die von einer rawlsianischen Sicht her zu setzen sind; die Fragen zu formulieren, die aus einer rawlsianischen Sicht heraus ein Klimamodell beantworten sollte; und die Aufgaben herauszuarbeiten, die von den im Modell erfaßten Problemstellungen her sich für eine rawlsianische Theorie der Gerechtigkeit stellen.

Im Folgenden diskutiere ich kurz einige Aspekte des Modells, die für ein Anwendung der Rawls'schen Theorie der Gerechtigkeit relevant sind und z.T. eine „Übersetzung“ der Theorie auf das Modell hin verlangen.

7.3.1 Ein Ein-Gut-Modell

Beim RICE-Modell handelt es sich um ein Ein-Gut-Modell. Das kommt einer rawlsianischen Interpretation und der Anwendung der zwei Prinzipien insoweit entgegen, als damit ein Indexproblem von vornherein ausgeschlossen wird.

¹⁶ Ganz allgemein finden sich etwa zu Fragen der Umweltpolitik nur einige wenige kursorische Bemerkungen, wie etwa Rawls [1975, 304 f.]. Einige weitere solcher „Lücken“ werden im Laufe der Darstellung noch deutlich werden.

Strategie	0	1	2	3	4	5	6	7	8	9	10	11	12	13
USA	0	22	21	-947	-306	-883	-8478	82	12	10	-79	-313	-694	-833
OHI	0	27	28	-135	-6	-132	-2504	-381	11	9	3	-15	-1	-23
EU	0	119	114	241	151	110	-2689	1846	45	33	45	44	157	51
EE	0	-11	-13	-803	-228	-591	-3821	-153	-1	-0	42	162	11	14
MI	0	37	32	-609	-264	-716	-5136	1089	15	12	20	23	35	37
LMI	0	21	14	-1133	-383	-962	-5757	1325	14	12	28	37	56	58
China	0	-1	-9	-1566	-522	-1342	-7135	79	3	7	22	32	45	47
LI	0	87	69	-1829	-819	-1903	-9744	3803	41	33	62	83	123	129
Summe	0	302	255	-6781	-2375	-6419	-45265	7690	141	117	143	54	-268	-521

Tabelle 7.6: Wohlfahrtsgewinn gegenüber Referenzfall: abdiskontierter, gleichgewichteter Konsum. Eigene Berechnung.

7.3.2 Die gerechte Sparrate

Beim RICE-Modell handelt es sich um ein dynamische Modell. Damit gibt es keine direkte Anwendbarkeit der zwei Prinzipien der Gerechtigkeit *allein*, die sich auf einen Zeitpunkt bzw. auf eine zu einem (Zeit-)Punkt zusammengefaßte Zeitspanne (bei Rawls ein ganzes Leben) beziehen.

Rawls behandelt intertemporale (ökonomische) Aspekte als Fragen der Sparrate, die im Urzustand so zu bestimmen ist, daß eine jede Generation wollen kann, daß alle früheren Generationen dieser Sparrate gefolgt sind.¹⁷ Nach Rawls würde auf diese Weise ein Schema beschlossen, das jedem Niveau des gesellschaftlichen Reichtums eine bestimmte Sparrate, d. h. den zu sparenden prozentualen Anteil des Sozialprodukts zuordnet. Über das genauere Aussehen dieses Schemas macht Rawls keine Angaben, zwei Eigenschaften aber stellt er fest: Die aus Gründen der Gerechtigkeit geforderte reale Nettosparrate ist positiv bis die (materiellen) Voraussetzungen einer gerechten Grundstruktur gegeben sind, danach verlangt die Gerechtigkeit keine positive Sparrate mehr. (Eine Gesellschaft kann aber aus anderen Gründen weiter sparen.) Außerdem kann man vermuten, daß ansonsten in ärmeren Gesellschaften die verlangte Sparrate niedriger sein wird als in reicheren. [Rawls 1999g, 253-258; Rawls 2001, 159 f.]

Vor dem Versuch einer Anwendung dieses Verständnisses intertemporaler (intergenerationaler) Gerechtigkeit auf unser Modell ist zunächst auf eine der Rawlsschen Theorie interne Problematik hinzuweisen.

Rawls entwickelt ja seine Theorie der Gerechtigkeit zunächst und hauptsächlich als eine ideale Theorie, d. h. es wird u.a. vorausgesetzt, daß die (materiellen) Bedingungen der Gerechtigkeit gegeben sind. Davon gehen auch die Parteien im Urzustand aus. Die Parteien wissen also zwar nicht, welcher Generation sie angehören, sie wissen aber sicher daß sie keiner Generation angehören, die in einer Gesellschaft lebt, die die materiellen Bedingungen einer gerechten Gesellschaft nicht erfüllt. [Rawls 1999g, 109-111, 215 f.; Rawls 2001, 84, 101, 197] Das hat zur Folge, daß Parteien, die ja nur die eigenen Interessen vertreten, in keiner Weise die Interessen von Generationen berücksichtigen werden, die in einer solchen Gesellschaft leben, auch wenn sie wissen, daß ihr Spargrundsatz auch für diese Generationen gelten wird.

Rawls selbst geht aber – inkonsequenterweise – davon aus, daß die Parteien es für möglich halten, in jeder beliebigen Generation zu enden, insbesondere auch in armen Gesellschaften, die die materiellen Bedingungen für eine gerechte Gesellschaft noch nicht besitzen und deshalb sparen müssen. [Rawls 1999g, 255 f.]

Rawls hat in der Frage des gerechten Sparprinzips also ideale und nichtideale Theorie¹⁸ in einer problematischen Weise gemischt. Wenn man die beiden Theorieteile wieder sauber trennt, bleibt für die ideale Theorie nur noch die Frage nach der Sparrate

¹⁷ Entgegen einer Bemerkung in Fußnote 39 in Rawls [2001, 160] hat Rawls das auch schon in seinem Hauptwerk so verlangt [Rawls 1973, 287 f.]. *Zusätzlich* hat er dort noch angenommen, daß die Parteien sich um das Wohl ihrer Nachkommen sorgen [Rawls 1973, 126, 288 f., 292]. Diese Annahme hat er später aufgegeben. [Rawls 2001, 160]

¹⁸ Genauer geht es hier um den Teil der nichtidealen Theorie, der Gerechtigkeit unter ungünstigen Bedingungen behandelt. S. o. S. 132.

in Gesellschaften übrig, die die materiellen Bedingungen für eine gerechte Gesellschaft erfüllen. Mit Rawls kann man für diesen Fall als eine plausible Lösung annehmen, daß die Parteien im Urzustand sich für eine von der Gerechtigkeit geforderte Nettosparrate von Null entscheiden. Ja man kann sogar weitergehen und die Voraussetzung der Gegenwart als Eintrittszeit aufgeben, was zur Anwendung des Differenzprinzips auch für die Fragen intergenerationaler Gerechtigkeit führen würde. Anders als Rawls argumentiert, wäre es dann – wenigstens aus Rawlsscher Sicht – kein Problem mehr, daß die Anwendung des Differenzprinzips dazu führt, „daß entweder überhaupt nicht oder nicht genug gespart wird, um die gesellschaftlichen Verhältnisse so weit zu verbessern, daß die gesamten gleichen Freiheiten für alle wirksam werden können“ [Rawls 1975, 321 f.]. Denn zum einen sind die Voraussetzungen dafür für alle in der idealen Theorie betrachteten Gesellschaften annahmegemäß erfüllt, und zum anderen führen die Überlegungen Rawls' für die *betrachteten Gesellschaften* zum gleichen Ergebnis.

Die Frage der Sparrate in einer Gesellschaft, die die materiellen Bedingungen der Gerechtigkeit noch nicht erfüllt, gehört also zur nichtidealen Theorie, die von Rawls nicht in der gleichen Weise systematisch entwickelt wurde wie die ideale Theorie. [Rawls 1999g, 216] Insbesondere hat er keine systematische Antwort auf die Frage, in welcher Weise genau das Modell der Urzustands für die nichtideale Theorie fruchtbar gemacht werden soll: Klar ist, daß Rawls die ideale Theorie für fundamental hält, daß die nicht-ideale Theorie die ideale Theorie voraussetzt und daß er auch die Probleme der nicht-idealen Theorie im Rahmen der Vertragstheorie behandeln möchte. [Rawls 1999g, 311, 336 f., 343] Doch nach welchem *Prinzip* er etwa einen Aspekt der nichtidealen Theorie – die Frage des Gehorsams gegenüber ungerechten Gesetzen – im Rahmen der Verfassungsversammlung behandelt [Rawls 1999g, 310-312], einen anderen – die eng mit dem ersteren zusammenhängende Frage des zivilen Ungehorsams – aber in einer Art zweitem Teil des Urzustands [Rawls 1999g, 336 f.], bleibt offen. Unklar ist auch, wie genau Rawls sich die Änderungen der Voraussetzungen für den Urzustand vorstellt und wie er sie rechtfertigt als Veränderungen in den Spezifikationen *eines* Urzustands.¹⁹

Wir können also festhalten, daß die Aussagen Rawls' zum gerechten Spargrundsatz zum einen äußert vage bleiben²⁰ und zum anderen nicht mit der gleichen Stringenz begründet sind wie seine zwei Prinzipien der Gerechtigkeit. Dies kommt uns hier insoweit entgegen, als die Sparraten der verschiedenen Ländergruppen im RICE-Modell aufgrund von Verhaltensannahmen bestimmt werden und nicht eigentlich Objekt der normativen Bewertung sind.²¹ Natürlich weichen diese utilitaristisch bestimmten Sparraten von den Ergebnissen der „intuitiven Überlegungen“ Rawls' zu den gerechten Sparraten ab: So wird durchgängig, auch nach dem Erreichen eines hohen Wohlstand-

¹⁹ Dessen, daß solche Veränderungen der Voraussetzungen notwendig sind, ist sich Rawls bewußt [Rawls 1999g, 308 f.]. Nur das *Wie* der Einführung der veränderten Annahmen ist nicht genau erklärt. Unproblematisch ist im übrigen die Behandlung von (bestimmten) Fragen der nichtidealen Theorie in der verfassungsgebenden Versammlung, und von Fragen des Völkerrechts in einem zweiten Urzustand, da diese von vornherein als vom (ersten) Urzustand unterschieden und mit neuen Voraussetzungen konzipiert wurden. Vgl. Rawls [1999g, 172-174, 331 f.] und Rawls [1999d, 30-35]

²⁰ Was Rawls selbst genau so sieht. [Rawls 1999g, 252 f., 255]

²¹ Wobei man allerdings annehmen darf, daß Nordhaus diese Sparraten auch für normativ angemessen hält. Siehe oben S. 200.

niveau, viel gespart. Jedenfalls sind – abgesehen von den letzten Werten, die nur den Modellabbruch dokumentieren – auch die Nettosparraten weit von Null entfernt. (Die USA würden etwa im Referenzmodell im Jahr 2275 bei einem Pro-Kopf-Konsum von mehr als 87000 Dollar im Jahr noch eine Nettosparrate von knapp unter 14 % besitzen.) Da allerdings die gerechte Sparrate in wohlgeordneten Gesellschaften²² nur eine Art untere Grenze darstellt, ist dies in sich aus einer Rawlsschen Sicht kein Gerechtigkeitsproblem. Wenn wir aber annehmen, daß anfänglich wenigstens die beiden ärmsten Ländergruppen (China und „Low income“ Länder) die materiellen Voraussetzungen für eine wohlgeordnete Gesellschaft in den ersten Perioden noch nicht besitzen²³, dann wird man außerdem sagen können, daß in den ersten Perioden zuviel gespart wird: Entgegen der Rawlsschen Vermutung, daß – solange die materiellen Bedingungen einer gerechten Gesellschaft noch nicht bestehen – von früheren und ärmeren Generationen geringere Sparraten verlangt werden als von späteren und reicheren, beginnen die Nettosparraten relativ hoch, um dann zunächst regelmäßig abzunehmen.²⁴ Doch selbst das ist, wenn wir die Sparraten des Modells konsequent nicht als normativ geforderte sondern als Resultat der Entscheidungen ökonomischer Agenten verstehen, nicht eindeutig. Auch auf dem Weg zu einer gerechten Gesellschaft kann aus anderen Gründen als denen der Gerechtigkeit gespart werden.²⁵

7.3.3 Das Differenzprinzip

Nachdem die Rawlssche Ausarbeitung des Spargrundsatzes für die Beurteilung der Klimastrategien unergiebig ist, jedenfalls wenn wir die Verhaltensannahmen des Modells nicht ändern, ist zu fragen, inwieweit die zwei Prinzipien zu einer solchen Beurteilung beitragen können. Da weder Fragen der gewährten Freiheiten noch der Zugangsmöglichkeiten zu gesellschaftlichen Positionen in dem Modell erfaßt werden, können wir uns dabei auf das Differenzprinzip beschränken.

Rawls selbst würde das Differenzprinzip in unserem Kontext als nicht relevant betrachten: Das Differenzprinzip ist nur auf Verteilungsfragen innerhalb einer Gesellschaft anzuwenden, nicht aber auf Verteilungsfragen zwischen verschiedenen Gesellschaften.[Rawls 1999d, 115-120] Da das Modell die interne Verteilung in den verschiedenen Ländergruppen nicht erfaßt, ist das Differenzprinzip also nicht anwendbar.

Rawls kommt nicht zu einem internationalen Differenzprinzip wegen der spezifischen Form, in der er den Urzustand konstruiert, in dem über die Fragen des Völkerrechts nachgedacht wird. Es ist klar, daß andere Formen, diesen Urzustand zu konstruieren, zum Differenzprinzip²⁶ führen würden. Insbesondere wäre das der Fall, wenn man den ersten Urzustand statt einer Versammlung der Vertreter von Menschen eines

²² Zum Begriff s. u. S. 218 f.

²³ Dazu weiter unten (Abschnitt 7.3.5) mehr.

²⁴ Folgendes sind die Nettosparraten der Ländergruppe „Low income“ in den ersten zehn Perioden im Referenzfall (Prozent): 25,20 – 22,23 – 20,29 – 19,00 – 18,09 – 17,42 – 16,91 – 16,50 – 16,16 – 15,88.

²⁵ Rawls jedenfalls schränkt die entsprechende Bemerkung [Rawls 1999g, 255] nicht auf wohlgeordnete Gesellschaften ein.

²⁶ Oder – im Sinne meiner Rekonstruktion – jedenfalls zu einem dem Differenzprinzip ähnlichen Prinzip.

Landes als Versammlung aller Menschen der Welt versteht, und dort über ein weltweites Verteilungsprinzip entscheiden läßt. Als Grund für einen solchen erweiterten Urzustand könnte man anführen, daß aufgrund der vielen wirtschaftlichen, politischen und kulturellen Verflechtungen die Welt und nicht mehr nur einzelne Nationen als ein (idealerweise faires) System sozialer Kooperation zu verstehen ist.²⁷ [Vgl. Rawls 1999g, 19, 47, 73 f., 109; Rawls 2001, 5-7] Auch die Bedingung, daß wir von einem „geschlossenen Systemen“, von einer „sozialen Welt“ sprechen, die wir im Gegensatz etwa zu freien Assoziationen mit der Geburt betreten und nur durch den Tod verlassen [Rawls 1999g, 229; Rawls 1993b, 12, 40 f.; Rawls 2001, 55] ist für die Welt wörtlich, für Nationalstaaten²⁸ dagegen nur annähernd erfüllt.²⁹ Nicht nur die Grundstruktur der Gesellschaft, in der wir leben, sondern auch die Organisation der weltweiten Kooperation und internationale Konflikte „beeinflussen“ die „Lebenschancen“ der Menschen. Auch ihre Wirkungen sind „tiefgreifend und von Anfang an vorhanden“. [Vgl. Rawls 1975, 23]

Im übrigen hält Rawls die Konsequenzen der Anwendung des Differenzprinzips auf internationale Verteilungsfragen für nicht akzeptabel, letztlich, da sie die politische Autonomie freier und gleicher Völker verletzen würde. Er illustriert das mit zwei Beispielen, in denen jeweils freie Entscheidungen – im ersten Beispiel über die reale Sparrate, im zweiten über die Rolle von Frauen³⁰ und damit über das Bevölkerungswachstum – bei gleicher Ausgangslage längerfristig zu unterschiedlichem Wohlstand führen. Eine Position, die unter diesen Bedingungen von der jeweils reicheren Gesellschaft verlangen würde, die jeweils ärmere zu unterstützen, scheint Rawls inakzeptabel. [Rawls 1999d, 117 f.]

Wenn wir trotzdem versuchen, das Differenzprinzip auf das RICE-Modell anzuwenden, stoßen wir sofort auf die schon erwähnte Schwierigkeit, daß das Prinzip die Verteilung zu bestimmten Zeitpunkten oder in geeignet zusammengefaßten Perioden beurteilt und nicht unmittelbar dazu geeignet ist, Zeitpfade zu vergleichen.³¹

Die Schwierigkeit wird deutlich sichtbar im Blick auf Abbildung 7.1. Sie zeigt die Rangfolge aller Strategien, die zumindest in einer Periode entweder den minimalen Pro-Kopf-Konsum unter allen Ländergruppen (faktisch immer den der Ländergruppe „Low Income“) maximieren oder zumindest den zweitgrößten minimalen Pro-Kopf-Konsum ausweisen, und des Referenzfalls.³²

Selbst wenn wir die Strategie des Geoengineering mit in Betracht ziehen, kann das Differenzprinzip – nur ergänzt durch ein einfaches und offensichtliches Dominanzprin-

²⁷ Ähnlich argumentieren auch Beitz [2002] und Pogge [1989, 246-259].

²⁸ Rawls spricht konsequent von „Gesellschaften“ oder „Völkern“, sachlich sind damit aber Nationen gemeint. Vgl. Rawls [1999d, 23-30]. Dort auch eine Begründung dafür, weshalb Rawls den Begriff des „Staates“ für die Bezeichnung der gemeinten Einheiten vermeidet.

²⁹ Rawls selbst nennt diese Annahme bezogen auf Nationalstaaten „a considerable abstraction“. [Rawls 1993b, 12]

³⁰ Die Forderung gleicher Freiheiten für Frauen wird dabei als erfüllt vorausgesetzt.

³¹ Forderungen der intergenerationalen Gerechtigkeit werden für Rawls durch die gerechte Sparrate erfaßt. Das Differenzprinzip wird dann auf die nach Erfüllung der Forderungen der intergenerationalen Gerechtigkeit noch in der betrachteten Periode möglichen Verteilungen der Grundgüter angewandt. [Vgl. Rawls 1975, 319 f., 326 f.]

³² Ein zahlenmäßig höherer Rang entspricht dabei einem höheren Pro-Kopf-Konsum.

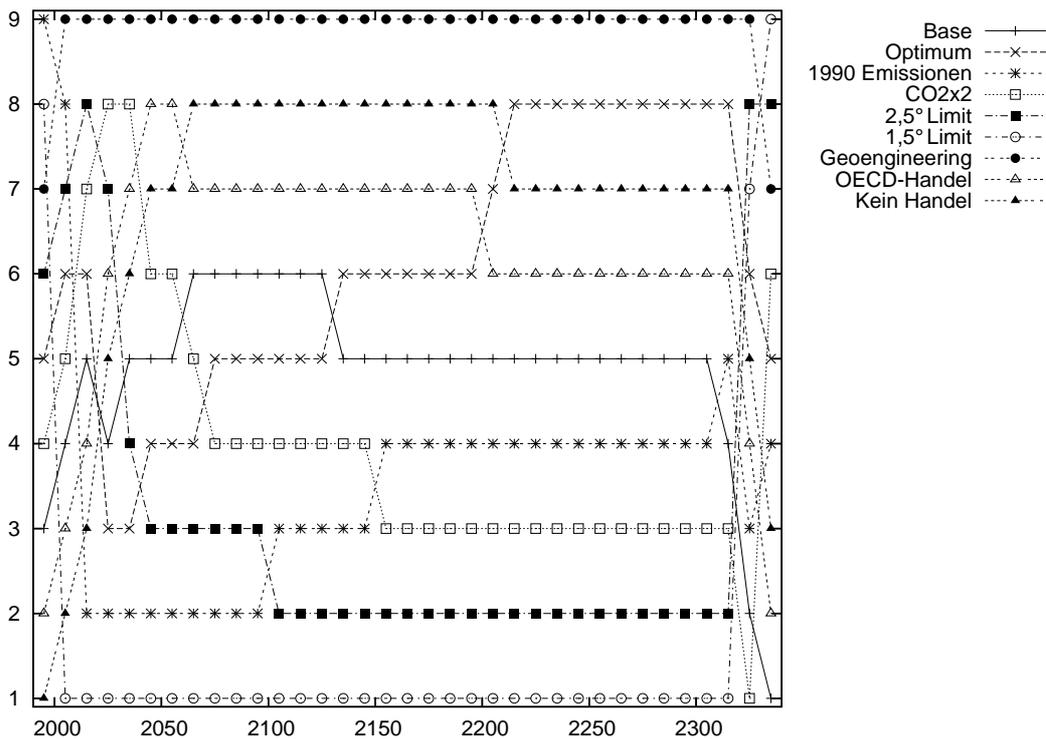


Abbildung 7.1: Rang der „besten“ Strategien (Pro-Kopf-Konsum).

zip – keine beste Strategie bestimmen. Wenn wir von Geoengineering absehen, bleiben sieben Strategien, die durch die Kombination von Differenz- und Dominanzprinzip deshalb nicht geordnet werden können.

Relativ leicht ist es noch, die Strategie der Beschränkung auf einen Temperaturanstieg von 1,5°C auszuschließen. Offensichtlich maximiert sie nur aufgrund des Modellabbruchs den minimalen Pro-Kopf-Konsum in der letzten Periode. Ähnlich leicht kann man es sich aber nicht mit den Strategien machen, die in einer oder zwei der frühen Perioden den minimalen Pro-Kopf-Konsum maximieren. Anders als die Endphase der Strategien ist die Anfangsphase nicht einfach ein Artefakt des Modells sondern (möglicherweise) auch ein Aspekt der modellierten Realität. Vielleicht kann man diese Strategien *intuitiv* mit dem Argument ausschließen, daß ihre kurzfristigen und anfänglichen „Erfolge“ wenig Gewicht haben gegenüber den längerfristigen aber späteren positiven Folgen der übrigen beiden Strategien. Und in der Tat hält Rawls von einem bestimmten Punkt an die Berufung auf die Intuition für unvermeidbar. [Rawls 1975, 59, 354 f.] Auch wenn man der Meinung ist, daß hier ein solcher Punkt erreicht ist, bleibt das Problem des Vergleichs der sogenannten optimalen Strategie und „Kyoto ohne Handel“. Hier liegt wohl keine einfache, plausible Intuition vor, die zu einer eindeutigen Entscheidungen zwischen den beiden führen könnte. Es bleibt also die Frage, ob aus rawlsianischer Sicht nicht doch eine systematischere und prinzipiellere Antwort

auf das vorliegende Entscheidungsproblem gefunden werden kann.

Ich habe bislang für die Anwendung des Differenzprinzips die Konsumhöhe zugrundegelegt. Das entspricht der üblichen Vorgehensweise in der normativen Ökonomik, die – wie auch Nordhaus – den Konsum als die wohlfahrtstheoretisch relevante Größe behandelt. Für Rawls ist allerdings das relevante „ökonomische“ Grundgut das gesamte Einkommen und nicht nur der konsumierte Teil.³³ Abbildung 7.2 zeigt die Rangordnung der sechs Strategien die zumindest einmal den minimalen Pro-Kopf-Output (abzüglich der Klimaschäden) maximieren oder den zweitgrößten minimalen Pro-Kopf-Output erzielen.

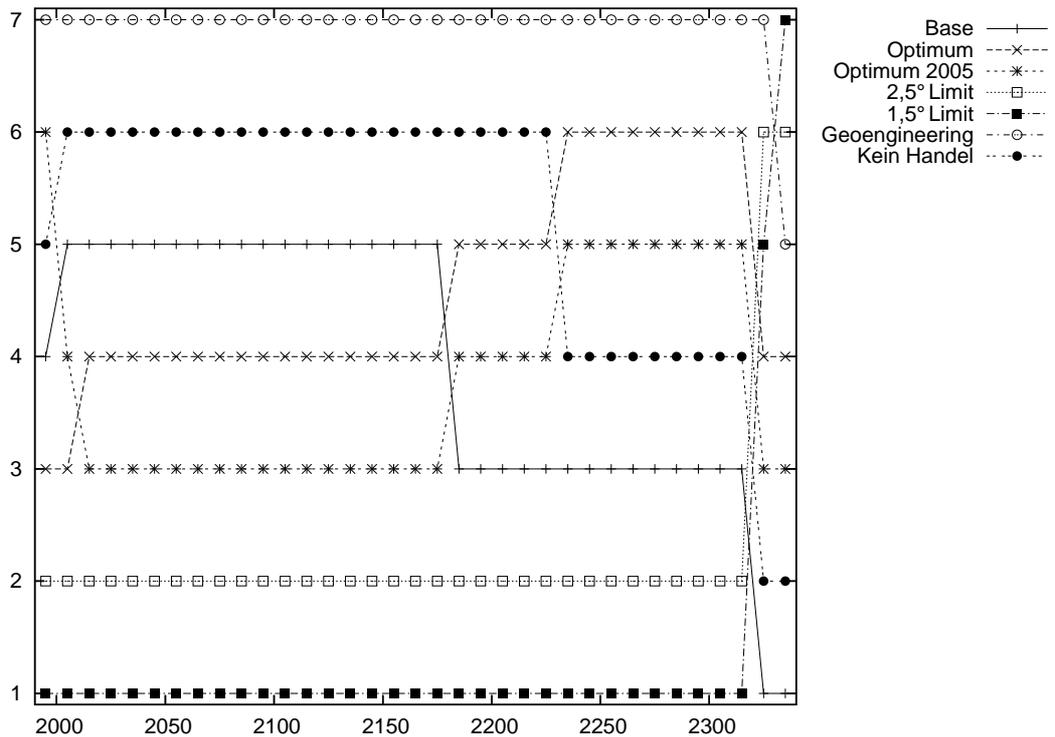


Abbildung 7.2: Rang der „besten“ Strategien (Pro-Kopf-Produktion).

Wenn wir wieder von Geoengineering absehen und die Ergebnisse der letzten Perioden als Artefakte des Modellabbruchs ausschließen, bleibt auch hier in jedem Fall immer noch die Frage des systematischen Vergleichs zwischen der „optimalen Strategie“ und des Kyotoprotokolls ohne Zertifikathandel, auch wenn sich „intuitiv“ die Gewichtung wohl in Richtung des Kyotoprotokolls verschoben hat.

³³ Vgl. Hinweise zu Rawls' weitem Einkommensbegriff Rawls [2001, 172].

7.3.4 Ein rawlsianischer Utilitarismus

Eine mögliche prinzipielle Antwort auf die Frage nach der Wahl zwischen den verschiedenen Strategien wäre der Versuch, Rawls mit utilitaristischen Ideen zu verknüpfen. Dabei würde man von Rawls das Prinzip übernehmen, daß die relevante (hier globale) gesellschaftliche Position, von der aus Fragen zu beurteilen sind, für die soziale und ökonomische Ungleichheiten wichtig sind, die der jeweils am schlechtesten Gestellten ist [Rawls 1999g, 81-86], in unserem Fall in jeder Periode die Position der Bewohner der „Low Income“ Ländergruppe.

Für die Beurteilung der unterschiedlichen Strategien würde man dann aber nicht vom Differenzprinzip angewandt auf das hier einzige Gut ausgehen, sondern dafür die im Modell implizit enthaltene (kollektive) Bewertung durch die Betroffenen, d. h. die regionale Wohlfahrtsfunktion benutzen. Da die betroffene Region im gegebenen Fall nicht wechselt, ist dies sehr einfach möglich und verlangt keinen interregionalen Nutzenvergleich.³⁴ Ein Blick auf z. B. Tabelle 7.5 zeigt, für was sich die Ländergruppe „Low Income“ entscheiden würde: das Kyoto-Protokoll ohne Handel.³⁵

Wenigstens im gegebenen Fall hat diese Lösung ihren Charme: Sie führt zu einem eindeutigen Ergebnis, und die in der regionalen Wohlfahrtsfunktion enthaltene Bewertung zu übernehmen, macht den angenehmen Eindruck, daß auf Paternalismus verzichtet wird.³⁶

Ein wichtiges Problem dieser Lösung ist jedoch, daß sie sich nicht direkt auf andere Fälle übertragen läßt, in denen die am schlechtesten gestellte Gruppe wechselt. Eine teilutilitaristische Lösung, die die Wohlfahrtssumme der verschiedenen benachteiligten Gruppen als Kriterium verwenden würde, wäre jedenfalls auf einen (interregionalen) Nutzenvergleich angewiesen, dessen Schwierigkeiten einer der Gründe sind, die den Utilitarismus als problematisch erscheinen lassen.³⁷

Für Rawls selbst wäre diese Lösung sicher nicht akzeptabel, unter anderem, da diese Wohlfahrtssummen unter Voraussetzung einer reinen Zeitpräferenz bestimmt werden, die Rawls allenfalls als Korrektiv unzureichender Kriterien akzeptiert. Ansonsten lehnt er sie ab: Für Individuen ist sie irrational und aus dem Urzustand, der gegenüber verschiedenen Zeitpunkt symmetrisch ist, ist sie nicht herleitbar. [Rawls 1999g, 259-262] Im übrigen ist aus der Sicht des Urzustandes die Summierung der Wohlfahrt über verschiedene Generationen denselben Problemen ausgesetzt, wie allgemein der Utilitarismus: Es ist nicht ausgeschlossen, daß von bestimmten Generationen Opfer verlangt werden, die nicht akzeptabel sind [vgl. z. B. Rawls 2001, 101-104], es gibt keine objektive Basis für die Annahme der Gleichwahrscheinlichkeit für die verschiedenen Generationen [vgl. Rawls 1999g, 145-150] und es wird für die Parteien – entgegen

³⁴ Nachdem die am schlechtesten Gestellten mit einem Rawlsschen Kriterium bestimmt wurden.

³⁵ Wie im oben betrachteten rein utilitaristischen Fall mit egalitärer Nutzenfunktion wäre auch das Kyoto-Protokoll in seiner tatsächlichen Fassung der „optimalen Strategie“ – wenn auch knapp – überlegen.

³⁶ Da es sich bei dieser Wohlfahrtsfunktion nicht um eine individuelle Nutzenfunktion handelt, sondern um die Modellierung des Verhaltens eines – sehr disparaten – Kollektivs, geht es dabei allerdings wirklich nur um einen *Eindruck*.

³⁷ Auch wenn Rawls seine Ablehnung des Utilitarismus nicht hauptsächlich darauf aufbauen möchte. [Rawls 1999g, 282-285]

der Rawlsschen Spezifikation des Urzustands – vorausgesetzt, daß „sie keine Personen mit bestimmten letzten Interessen oder einer bestimmten Vorstellung von ihrem Wohl sind, das sie fördern möchten“. [Rawls 1975, 201; für das ganze Argument Rawls 1999g, 150-153] Die verschiedenen Generationen werden in einer Weise zusammengefaßt, als handle es sich um das Entscheidungsproblem eines langlebigen Individuums: „Der Utilitarismus nimmt die Verschiedenheit der einzelnen Menschen nicht ernst“. [Rawls 1975, 45, 214 f.]

7.3.5 Das Völkerrecht

Wie schon ausgeführt, hält Rawls das Differenzprinzip nicht für geeignet, um Fragen der internationalen Verteilungsgerechtigkeit zu entscheiden.³⁸ Statt dessen argumentiert er, daß die Parteien des zweiten Urzustands eine Beistandspflicht gegenüber „belasteten Gesellschaften“ beschließen würden, d. h. Gesellschaften, denen „politische und kulturelle Traditionen, das Humankapital, das Know-How, und oft auch die nötigen materiellen und technologischen Ressourcen“ fehlen, um wohlgeordnet zu sein. Das langfristige Ziel dieser Beistandspflicht ist es, den belasteten Gesellschaft zu helfen, wohlgeordnet zu werden. [Rawls 2002, 131 f.]

Abstrakt und allgemein handelt es sich bei einer wohlgeordneten Gesellschaft um eine Gesellschaft, „die von einer öffentlichen Gerechtigkeitskonzeption wirksam reguliert wird“. [Rawls 2003, 29]

Konkreter sind wohlgeordnet zum einem „vernünftige liberale“ und zum anderen „achtbare“ Gesellschaften (bzw. Völker). [Rawls 1999d, 4]

Vernünftige liberale Völker besitzen eine „annehmbare gerechte konstitutionelle demokratische Regierung, die ihren grundlegenden Interessen dient“, ihre Bürger sind durch gemeinsame Zuneigungen geeint, und sie besitzen einen moralischen Charakter, d. h. sie sind rational und vernünftig (d. h. ihr rationales Handeln wird durch einen Sinn für das Vernünftige eingeschränkt) und sie bieten anderen vernünftigen Völkern faire Bedingungen der Kooperation an, die sie dann auch beachten. [Rawls 2002, 26-28]

Intern folgt eine annehmbare gerechte demokratische Gesellschaft einer Gerechtigkeitskonzeption, die eine Verallgemeinerung der Rawlsschen zwei Prinzipien darstellt und aus drei Grundsätzen besteht:

the first enumerates basic rights and liberties of the kind familiar from a constitutional regime;

the second assigns these rights, liberties and opportunities a special priority, especially with respect to the claims of the general good and perfectionism values; and

the third assures for all citizens the requisite primary goods to enable them to make intelligent and effective use of their freedoms. [Rawls 1999d,

³⁸Rawls folgend stelle ich seinen Ansatz zunächst als Alternative zu einem internationalen Differenzprinzip dar. Aber man kann seinen Vorschlag auch als die systematische Antwort auf die Frage verstehen, die ein internationales Differenzprinzip nicht beantworten konnte. Siehe dazu unten Anm. 71.

14]

Außerdem muß es Institutionen geben, die exzessive Ungleichheiten verhindern, wie etwa folgende:

(a) A certain fair equality of opportunity, especially in education and training. (Otherwise all parts of society cannot take part in the debates of public reason or contribute to social and economic policies.)

(b) A decent distribution of income and wealth meeting the third condition of liberalism: all citizens must be assured the all-purpose means necessary for them to take intelligent and effective advantage of their basic freedoms. (In the absence of this condition, those with wealth and income tend to dominate those with less and increasingly to control political power in their own favor.)

(c) Society as employer of last resort through general or local government, or other social and economic policies. (The lack of a sense of long-term security and of the opportunity für meaningful work and occupation is destructive not only of citizens' self-respect, but of their sense that they are members of society and not simply caught in it.)

(d) Basic health care assured for all citizens.

(e) Public financing of elections and ways of assuring the availability of public information on matters of policy. (A statement of the need for these arrangements merely hints at what is needed both to ensure that representatives and other officials are sufficiently independent of particular social and economic interests and to provide the knowledge and information upon which policies can be formed and intelligently assessed by citizens.) [Rawls 1999d, 50]

Achtbare Gesellschaften³⁹ erfüllen zwar die Anforderungen an eine liberale Gesellschaft nicht vollständig [Rawls 1999d, vgl. 63, 67], in jedem Fall aber respektieren sie „die politische und soziale Ordnung anderer Gesellschaften“ und akzeptieren dieselben Grundsätze des Rechts der Völker wie liberale Gesellschaften. [Rawls 2002, 79, 84] In achtbaren Gesellschaften werden die Menschenrechte für alle Mitglieder geachtet, darunter:

the right to life (to the means of subsistence and security); to liberty (to freedom from slavery, serfdom, and forced occupation, and to a sufficient measure of liberty of conscience to ensure freedom of religion and thought); to property (personal property); and to formal equality as expressed by the rules of natural justice (that is, that similar cases are treated similarly). [Rawls 1999d, 65]

³⁹ Genauer: achtbare hierarchische Gesellschaften. Rawls geht davon aus, daß es noch andere Formen achtbarer Gesellschaften gibt. Da er aber weder Hinweise darauf gibt, wie sie aussehen könnten, noch darauf, welche Kriterien sie erfüllen, um als achtbar anerkannt zu werden, identifiziere ich hier ein-fachheitshalber achtbare mit achtbaren hierarchischen Gesellschaften. Vgl. Rawls [1999d, 63].

Außerdem ordnet eine achtbare Gesellschaften allen ihren Mitgliedern gewisse Rechte und Verpflichtungen zu, sein Rechtssystem folgt einer Gemeinwohlvorstellung, welche grundlegende Interessen aller ihrer Mitglieder berücksichtigt, und Richter und Beamte sind ehrlicher- und vernünftigerweise davon überzeugt, daß dies der Fall ist. [Rawls 1999d, 65-67] Um einige der bisher erwähnten Bedingungen zu erfüllen, ist es zuletzt notwendig, daß sein Rechtssystem eine „achtbare Konsultationshierarchie“ einschließt, d. h. ein etabliertes Konsultationsverfahren, in dem alle Gruppen der Gesellschaft (die gemeinsam alle Mitglieder erfassen müssen) in einer maßgeblichen Weise repräsentiert sind. [Rawls 1999d, 71 f., 77 f.]

Will man die Beistandspflicht praktisch umsetzen, muß (u.a.) geklärt werden, welches Niveau gesellschaftlichen Reichtums für die Verwirklichung einer wohlgeordneten oder achtbaren Gesellschaft notwendig ist.⁴⁰ Man kann vermuten, daß dieses Niveau für nur achtbare Gesellschaften niedriger ist, da sie ja ihren Bürgern deutlich weniger Rechte garantieren müssen. Rawls erklärt aber nicht eindeutig, wann genau die (materielle) Beistandspflicht beendet ist: wenn alle Völker die Bedingungen für eine liberale Gesellschaft, wenn alle Völker die Bedingungen für eine achtbare Gesellschaft, oder wenn alle Völker die Bedingungen für jeweils den wohlgeordneten Gesellschaftstyp erfüllt haben, den sie anstreben. Vermutlich denkt er an Letzteres. Dies würde jedenfalls erklären, weshalb er im Zusammenhang des Endes der Beistandspflicht immer beide Gesellschaftstypen erwähnt.⁴¹

Trotzdem werde ich für die folgende Diskussion voraussetzen, daß die Beistandspflicht verlangt, daß für alle Völker die materiellen Bedingungen einer wohlgeordneten liberalen (d. h. gerechten) Gesellschaft gesichert werden. Das dient zum einen der Vereinfachung, zum anderen ist zu fragen, ob für liberale Völker nicht eine Pflicht besteht, allen Völkern die Verwirklichung einer liberalen Demokratie zu ermöglichen.⁴²

Rawls selbst macht auch keine genauen Angaben darüber, was er als das für die Verwirklichung einer gerechten Gesellschaft notwendige Niveau gesellschaftlichen Reichtums erachtet. Deutlich macht er nur, daß dieses Niveau für verschiedene Gesellschaften verschieden sein kann (abhängig z. B. von ihrer Geschichte) und daß es nicht sehr hoch sein muß. Nicht nur reiche Gesellschaften können gerecht sein. [Rawls 1999d, 106-108]

⁴⁰ Ich werde mich im folgenden auf eine Diskussion einer materiellen Beistandspflicht beschränken. Rawls betont, daß „die bloße Verteilung von Geldern nicht ausreicht, um grundlegende politische und soziale Ungerechtigkeiten zu berichtigen“, erwähnt aber auch, daß „Geld dafür oft unverzichtbar ist“. [Rawls 2002, 134]

⁴¹ Nur ein Beispiel: „In the society of the Law of Peoples the duty of assistance holds until all societies have achieved just liberal or decent basic institutions.“ [Rawls 1999d, 118] Im Unterschied zu meiner Darstellung im Text spricht Rawls nicht von der Erfüllung der Bedingungen für eine wohlgeordnete Gesellschaft, sondern von ihrer Verwirklichung. Das erklärt sich damit, daß ich nur einen (den materiellen) Teil der Bedingungen für eine wohlgeordnete Gesellschaft betrachte, Rawls dagegen alle. Wenn alle Bedingungen einer wohlgeordneten Gesellschaft erfüllt sind, und wir „Schurkenstaaten“, die ein anderes Ziel anstreben, gesondert betrachten, dann ist mit den Bedingungen auch die Verwirklichung gegeben. Die im Text kurz behandelte Problem ist zu unterscheiden von der Frage, ob liberale Gesellschaften Anreize setzen sollten, um achtbare Gesellschaften in Richtung von mehr Liberalität zu verändern. Eine Frage, die Rawls verneint. [Rawls 1999d, 84 f.]

⁴² Eine Frage, die ich hier allerdings nicht vertiefe.

Angewendet auf das RICE-Modell möchte ich deshalb fragen, welches Maß an Reichtum (Bruttosozialprodukt pro Kopf) plausiblerweise notwendig ist, um die von Rawls angeführten Bedingungen einer gerechten liberalen Gesellschaft erfüllen zu können.⁴³

Zur Operationalisierung schlage ich vor, auf die von Nordhaus u. Boyer [2000c] für die Kalibrierung ihres Modells erfaßten Länder zu schauen, und zu fragen, welches die Länder sind, die bei niedrigstem Bruttosozialprodukt pro Kopf⁴⁴ (um) 1995 dem Ideal einer gerechten liberalen Gesellschaft, wie es von Rawls beschrieben wurde, hinreichend nahe kommen.⁴⁵ Dem liegt zum einen die Idee zugrunde, daß, wenn es bislang keinem Land unter einem bestimmten Wohlstandsniveau gelungen ist, dem Ideal wenigstens nahe zu kommen, es nahe liegt, zu vermuten, daß es wenigstens *praktisch* auch nicht möglich ist.⁴⁶ Zum anderen die umgekehrte, wenig umstrittene Idee, daß Sein Möglichkeit impliziert.⁴⁷

Genauer werde ich für jede der Bedingungen Rawls' ein praktisches Kriterium vorschlagen, bei dessen Erfüllung man (für praktische Zwecke) die Bedingung als hinreichend erfüllt betrachten kann.⁴⁸

1. Die ersten beiden Grundsätze einer gerechten liberalen Gesellschaft betrachte ich als erfüllt, wenn das betrachtete Land von „Freedom House“ als frei eingestuft wird.⁴⁹
2. Den dritten Grundsatz betrachte ich durch die folgenden fünf Kriterien konkretisiert.
3. Ein „gewisses Maß“ an fairer Chancengleichheit besonders in der Bildung (a) betrachte ich als gegeben, wenn die geschätzte oder berichtete Alphabetisierungsrate bei Erwachsenen 99 % oder mehr beträgt.⁵⁰

⁴³ Dem hochaggregierten Charakter des Modells entsprechend werde ich hier nur *eine* solche Grenze vorschlagen und nicht zwischen den Erfordernissen verschiedener Völker differenzieren. Im übrigen möchte ich betonen, daß die folgenden Überlegungen nur einige Hinweise in Richtung der Bestimmung eines solchen Maßes geben wollen und als sehr vorläufig gedacht sind.

⁴⁴ Für die Berechnung dieses Wertes benutze ich – soweit sie vorliegen – die Angaben von Nordhaus u. Boyer [2000c, 28-39]. Ansonsten verwende ich die Zahlen von UNSP [2004].

⁴⁵ Da es sich um ein Ideal handelt, können wir nicht erwarten, daß irgendeine existierende Gesellschaft alle Bedingungen vollständig erfüllt. Vgl. Rawls [1999d, 75].

⁴⁶ Das setzt voraus, daß eine hinreichende Anzahl von Ländern versucht, dem Ideal einer gerechten liberalen Gesellschaft nahe zu kommen.

⁴⁷ Von Eigenschaften der Länder, die unterschiedliche Wohlstandsniveau verlangen würden, sehen wir ja ab.

⁴⁸ Angesichts des vorläufigen Charakters der Übung wird die leichte Erreichbarkeit von Daten bei der Zuordnung eine große Rolle spielen.

⁴⁹ Datenbasis: Tabelle mit allen Ländereinstufungen von Freedom House 1972-2003 [Freedom House 2004].

Genau genommen gehen die Kriterien von Freedom House allerdings in einem Punkt zu weit, da sie eine gerechte Gesellschaft auf eine „property owning democracy“ festlegen. Kriterium D 3 der „Civil Liberties Check List“ (in der Fassung von 2001) fragt: „Are property rights secure? Do citizens have the right to establish private businesses? Is private business activity unduly influenced by government officials, the security forces, or organized crime?“ [Freedom House 2001]

⁵⁰ Für entwickelte Industrienationen, für die keine andere Daten vorliegen, schätzt UNDP eine Alphabe-

4. Eine „akzeptable Verteilung“ (b) sei gegeben, wenn weniger als 0,5 % der Bevölkerung in absoluter Armut leben, d. h. von einem Einkommen von unter 1 Dollar (KKP) leben müssen.⁵¹
5. Die Gesellschaft werde ihrer Rolle als „Arbeitgeber letzter Instanz“ (c) annähernd gerecht, wenn die Arbeitslosenquote unter 5 % liegt.⁵²
6. Eine medizinische Grundversorgung (d) für alle sei gesichert, wenn (statistisch) 100 % der Bevölkerung Zugang zu Gesundheitsdiensten hat.⁵³
7. Eine öffentliche Finanzierung von Wahlen (e) wird als gegeben angenommen, wenn Parteien direkt (sei es ohne Verwendungsbeschränkungen, sei es für Wahlkampfaufwendungen, für allgemeine Administration oder andere Zwecke) mit staatlichen Finanzmitteln unterstützt werden.⁵⁴ Die allgemeine Zugänglichkeit von politischer Information wird als gegeben angenommen, wenn (Pressefreiheit vorausgesetzt) 100 oder mehr Tageszeitungen pro 1000 Einwohner zirkulieren.⁵⁵

tisierungsquote von 99 %, die offensichtlich in Rechnung stellen soll, daß eine kleine Zahl von erwachsenen Analphabeten praktisch nicht zu vermeiden ist. Die höchsten *berichteten* Alphabetisierungsquoten betragen ebenfalls 99 % (hauptsächlich frühere Ostblockstaaten bzw. Mitglieder der Sowjetunion). Datenbasis: Bericht über die menschliche Entwicklung 1998 [Deutsche Gesellschaft für die Vereinten Nationen e.V. 1998, 152-154].

⁵¹ Das könnte man als eine zu lockere Forderung ansehen. Man beachte aber, daß es zum einen nur um eine Annäherung an ein Ideal geht, und daß es zum anderen wieder praktisch unvermeidbar sein könnte, daß eine kleine Zahl von Bürgern ein sehr geringes Einkommen angeben. Zuletzt muß ein gewisser Raum bleiben für Ungenauigkeiten in den Schätzmethoden. Aufgrund der unvermeidbaren Ungenauigkeiten der Schätzungen weist etwa die Weltbank in ihren Publikationen Anteile der Bevölkerung in absoluter Armut unter 2 % nur mit der Angabe „<2“ aus [vgl. Weltbank 2004, 54 ff.]. Datenbasis: Global Poverty Monitoring Database [Chen u. Ravallion 2004].

⁵² Das Rawlssche Kriterium (d) ist sicher das problematischste in einem doppelten Sinn: Zum einen dürfte es umstritten sein, da die kapitalistischen Marktwirtschaften zwar Vollbeschäftigung als ein Ziel unter anderen anstreben, aber ihm kein solches Gewicht zumessen, daß sie sich zu seiner tatsächlichen Verwirklichung verpflichtet fühlten. Auch ideologisch wird eine Rolle des Staates als „Arbeitgeber letzter Instanz“ zurückgewiesen: Der Staat hat – in dieser Auffassung – *nicht* die Aufgabe, Vollbeschäftigung – wenn nötig durch direkte Intervention auf dem Arbeitsmarkt – zu sichern. Zum andern scheitern fast alle Nationen der Welt an diesem Kriterium, wenn man es ähnlich streng anlegen würde wie die anderen Kriterien. Datenbasis: Yearbook of Labor Statistics 2003 [ILO 2003, 485-505].

⁵³ Hier könnte man einwenden, daß die Forderung zu streng ist, da sein keinen Raum für statistische Ungenauigkeiten läßt. Im Unterschied zu den Kriterien Einkommen und Alphabetisierungsquote ist hier aber nur das Angebot eines Zugangs im Blick, nicht seine Inanspruchnahme und sein Ergebnis. Außerdem nimmt die Konkretisierung der Kriterien – durchaus im Einklang mit der gewünschten Praktikabilität der Vorgehensweise – natürlich Rücksicht auf die Datenlage in tatsächlich vorliegenden Statistiken. Datenbasis: Social Watch Report 2003 [Social Watch 2003, 62] ergänzt durch Angaben aus den „Epidemiological Fact Sheets on HIV/AIDS and Sexually Transmitted Infections, 2004 Update“ [UNAIDS, UNICEF, WHO 2004].

⁵⁴ Wieder handelt es sich um ein eher umstrittenes Kriterium, da keine Einigkeit darüber besteht, daß der Staat Wahlkämpfe finanzieren sollte. Ansonsten ist das Kriterium eher mild, da es nicht die Höhe der Unterstützung berücksichtigt. Datenbasis: IDEA Handbook on Funding of Political Parties and Election Campaigns [IDEA 2004], World Cultural Report 1998 [UNESCO 1998].

⁵⁵ Hier eine sinnvolle Grenze zu finden, ist noch schwieriger als bei den anderen Kriterien, da es keine „natürliche“ obere Schranke gibt. Die gewählte Zahl von 100 ist jedenfalls eine eher schwache Anforderung.

Wie sich herausstellt, ist das Land mit dem niedrigsten Bruttosozialprodukt pro Kopf, das alle diese Bedingungen erfüllt, die Tschechische Republik, mit – nach den Zahlen von Nordhaus – einem Bruttosozialprodukt von 3581,80 Dollar pro Kopf, einer Arbeitslosenrate von 4 %, einer Armutsquote von 0.12 % (1996; 0 % bei alternativer Schätzmethode) und einer Verbreitung von 296 Tageszeitungen auf 1000 Bewohner.⁵⁶

Das Ergebnis ist – wie zu erwarten – sehr sensibel gegenüber den gewählten Indikatoren.⁵⁷ Wenn wir etwa das umstrittene Kriterium der Vollbeschäftigung weglassen,⁵⁸ gebührt Polen der Preis: bei einem Bruttosozialprodukt von 1916,60 Dollar pro Kopf, einer Armutsquote von 0,09 % (1996) und einer Verbreitung von 140 Tageszeitungen auf 1000 Bewohner.

Der Indikator für die medizinische Grundversorgung scheint aus mehreren Gründen problematisch: Zum einen fehlen für eine ganze Reihe von Ländern statistische Angaben über den Zugang zu Gesundheitsdiensten, zum anderen ist er möglicherweise zu optimistisch. Nicht nur, daß in einer Reihe von Ländern, die einen hundertprozentigen Zugang zu Gesundheitsdiensten ausweisen, faktisch Zugangsschranken in der Form von Bestechungszahlungen oder sonstigen notwendigen direkten Barauslagen vorliegen,⁵⁹ kann man von einer gesicherten medizinischen Grundversorgung wohl dann nicht sprechen, wenn ein signifikanter Anteil der Bevölkerung keinen Zugang zu essentiellen Arzneimitteln haben. Wenn wir den Zugang zu Gesundheitsdiensten durch den Zugang zu essentiellen Arzneimitteln als Indikator ersetzen⁶⁰ und aufgrund der geringen Genauigkeit der Daten nur verlangen, daß über 95 % der Bevölkerung Zugang zu essentiellen Arzneimitteln haben,⁶¹ dann hat das weitreichende Konsequenzen: Solange wir die übrigen Kriterien unverändert beibehalten, dann erfüllen nur noch industrialisierte Länder mit hohem Einkommen alle Kriterien. Dabei müssen wir allerdings einfach *annehmen*, daß in diesen Länder der Anteil der Bevölkerung in absoluter Armut kleiner als 0,5 % ist, da für diese Ländergruppe in der Regel keine entsprechenden Statistiken vorliegen. Unter letzterer Bedingung ist das Land mit dem niedrigsten Bruttosozialprodukt pro Kopf, das alle Kriterien erfüllt Island, mit einem Bruttosozialprodukt pro Kopf von 22222,20 Dollar.⁶²

derung, wie ein Blick auf die Datenbasis zeigt. Datenbasis: IDEA Handbook on Funding of Political Parties and Election Campaigns [IDEA 2004], World Cultural Report 1998 [UNESCO 1998].

⁵⁶ Die übrigen Angaben ergeben sich direkt aus den Kriterien und den Erläuterungen dazu. Die Quelle für den Zugang zu Gesundheitsdiensten (100 %) ist [UNAIDS, UNICEF, WHO 2004].

⁵⁷ Aber s. u. die Anmerkung 71!

⁵⁸ Auch die Tschechische Republik hat seit 1998 Arbeitslosigkeitstraten von über 5 %. Die höheren Raten sind allerdings möglicherweise z.T. mit einem Wechsel der Definition von Arbeitslosigkeit zu erklären. Vgl. ILO [2003, 496 Anm. 9].

⁵⁹ Wie z. B. für Polen [Shahriari u. a. 2001]. Weitere Beispiele siehe Belli [2001, 18-22]. In der Tschechischen Republik dagegen scheinen solche Bestechungszahlungen keine bedeutende Rolle zu spielen. [Belli u. a. 2003]

⁶⁰ Datenbasis: Bericht über die menschliche Entwicklung 2001 [Deutsche Gesellschaft für die Vereinten Nationen e.V. 2001, 190-193].

⁶¹ Vgl. Deutsche Gesellschaft für die Vereinten Nationen e.V. [2001, 193].

⁶² Vernünftigerweise kann man annehmen, daß auch über 95 % der Bewohner Monacos Zugang zu wesentlichen Arzneimitteln haben. Das geringste Bruttosozialprodukt eines Landes, das alle Bedingungen erfüllt, wäre dann 22051,65 Dollar und damit in der gleichen Größenordnung.

Wenn wir auf die beiden „umstrittenen“ Kriterien (Arbeitslosigkeit und Wahlfinanzierung) verzichten⁶³ und weiter das Kriterium des Zugangs zu wesentlichen Arzneimitteln beibehalten, dann ist Ungarn das Land mit dem niedrigsten Bruttosozialprodukt pro Kopf, das alle Bedingungen erfüllt, mit einem Bruttosozialprodukt pro Kopf von 2639,30 Dollar, gefolgt von Estland (2702,70 Dollar) und der Slowakei (3538,20 Dollar).

Auf der Basis der vorangegangenen Diskussion erscheint es als vernünftig, überschlagsmäßig anzunehmen, daß das Bruttosozialprodukt, von dem an ein Land eine liberale Gesellschaft verwirklichen kann, irgendwo in einem Korridor von 1900,- zu 3600,- Dollar liegt, bzw. um uns auf eine Zahl festzulegen⁶⁴, bei 2750,- Dollar.⁶⁵

Abgesehen von der allgemeinen und vagen Aussage, daß gerechte, liberale Gesellschaften keineswegs reich sein und kein besonders hohes Durchschnittseinkommen aufweisen müssen [Rawls 1999g, 257], macht Rawls – wie schon erwähnt – keine genauen Angaben darüber, bei welchem Wohlstandsniveau die Beistandspflicht belasteten Gesellschaften gegenüber endet. Aus dem einleitenden Abschnitt zu seiner Diskussion belasteter Völker [Rawls 1999d, 105 f.] kann man aber vielleicht doch entnehmen, daß er jedenfalls das Deutschland der dreißiger Jahre des letzten Jahrhunderts und wohl auch Spanien, Frankreich und Österreich der frühen Neuzeit für keine belasteten Gesellschaften hält. Auf der Basis von Maddisons [Maddison 2004] historischen Schätzungen für das Bruttosozialprodukt pro Kopf umgerechnet auf eine Wechselkursbasis würde sich damit eine obere Schranke von 4532,20 Dollar (Deutschland 1933) bzw. eine Spanne von 536,70 (Spanien 1500) bis 1252 Dollar (Frankreich 1820) ergeben. Während die erst obere Schranke mit unserer Schätzung gut vereinbar ist, scheint die aus der frühen Neuzeit übernommene Spanne unverhältnismäßig niedrig: Länder mit einem Bruttosozialprodukt pro Kopf von unter 1252 sind (soweit entsprechende Daten vorliegen, was für einige Kleinstaaten nicht der Fall ist) in aller Regel von der Erfüllung der von Rawls inspirierten Kriterien *weit* entfernt. Am nächsten kommt den Kriterien das als frei eingestufte Chile mit einer Alphabetisierungsquote von 95,2 %, einem Zugang zu Gesundheitsdiensten von 97 %, einem Zugang zu essentiellen Arzneimitteln von 88 %, einer Arbeitslosenquote von 4,9 %, einer Armutsquote von 0 % und einer Zirkulation von 99 Tageszeitungen pro 1000 Einwohner. Eine Wahlfinanzierung wird nicht gewährleistet. Entsprechend der Schätzungen von Nordhaus u. Boyer [2000c, 33] hatte Chile 1995 ein Bruttosozialprodukt pro Kopf von 1124,40 Dollar. Diese Schätzung ist aber offensichtlich viel zu niedrig: Nach den Zahlen von UNSP [2004] hatte Chile in 1990 US-Dollar ein Bruttosozialprodukt pro Kopf von 3576,20 Dollar, d. h. mehr als das Dreifache des von Nordhaus verwendeten Wertes! Auch im Vergleich mit anderen von Nordhaus verwendeten Werten ist die Schätzung für Chile unglauwbüdig: Nur zum Beispiel hatte Chile demnach 1995 ein kleineres Pro-Kopf-Einkommen als Kongo und Papua-Neuguinea, deren Bruttosozialprodukt pro Kopf (in Dollar) nach Weltban-

⁶³ Da nicht sicher ist, ob die Erfüllung dieser Kriterien effektiv angestrebt wird.

⁶⁴ Aber s. u. die Anmerkung 71!

⁶⁵ Es entspricht wohl auch dem Selbstbild der genannten Nationen des früheren Ostblocks, daß sie die Vorbedingungen für die Verwirklichung einer liberalen (freiheitlichen und sozialen) Gesellschaft besitzen.

kangaben [Weltbank 1997, 246 f.] 1995 nur 16 % bzw. 28 % des chilenischen betragen.⁶⁶ Nordhaus u. Boyer [2000b, 3-40] haben nach eigenen Angaben an den Werten der Weltbank „multiple adjustments“ vorgenommen.⁶⁷ Welcher Natur diese Anpassungen waren, und weshalb für Chile ein so unverhältnismäßig niedriges Bruttosozialprodukt ausgewiesen wird, bleibt unerklärt. Für die Beurteilung der Angemessenheit der von mir vorgeschlagenen Grenze für eine Beistandspflicht erscheint der chilenische Wert in jedem Fall ungeeignet.

Wenn es die Pflicht entwickelter liberaler Gesellschaften ist, „belasteten Gesellschaften“ zu helfen, die Bedingungen für eine wohlgeordneten liberale Gesellschaften zu schaffen, und es ebenso die Pflicht früherer Generationen der belasteten Gesellschaften ist, zu diesem Ziel beizutragen, liegt es nahe, anzunehmen, daß anzustreben ist, dieses Ziel möglichst früh zu erreichen.⁶⁸ Diese Zielvorstellung ist allerdings abzuwägen gegen die berechtigten Interessen der früheren Generationen, von denen – so Rawls – eben nicht jedes Opfer zum Wohl der künftigen Generationen abzuverlangen ist. [Rawls 1975, 321, 331 f.]

Wie sich allerdings herausstellt, unterscheiden sich die verschiedenen Strategien nur unwesentlich, was den Zeitpunkt der Erreichung (z. B.⁶⁹) des Ziels eines Bruttosozialprodukts pro Kopf von 2750 Dollar angeht: Zehn der betrachteten Strategien erreichen dieses Niveau 2095, drei 2105 (1990-Emissionen, Doppelte CO₂-Konzentration, Temperaturlimit: 2,5 °C) und eine 2115 (Temperaturlimit: 1,5 °C). Wenn man den Zeitpunkt der genauen Zielerreichung interpoliert⁷⁰, dann unterscheiden sich die einzelnen Strategien zum Teil nur um Tage oder wenige Monate: Wenn wir von Geoengineering absehen, dann wird das Niveau als erstes mit der Strategie Kyoto ohne Zertifikatshandel erreicht, – nur zwei Tage später dann von der Strategie Kyoto mit Handel nur zwischen OECD-Ländern. Von den Strategien, die bis 2095 das angezielte Niveau erreichen, ist die letzte die um zehn Jahre verzögerte „Optimalstrategie“ (fünf Monate nach der ersten), die eigentliche „Optimalstrategie“ kommt drei Tage früher. Immerhin über zwei Jahre abgeschlagen sind dann die Strategien der Beschränkung der Emissionen auf das Niveau von 1990 (zwei Jahre und elf Monate) und das Temperaturlimit von 2,5 °C (zwei Jahre und neun Monate). Signifikant später wird das Ziel erreicht, wenn der Temperaturanstieg auf 1,5 °C beschränkt wird (mehr als 20 Jahre).

⁶⁶ Neben Kongo und Papua-Neuguinea hatten entsprechend der von Nordhaus verwendeten Daten noch Peru, Guyana Paraguay und El Salvador ein Bruttosozialprodukt über dem Chiles und unter 1252 Dollar. Keines dieser Länder erfüllte 1995 die operationalisierten Kriterien für eine gerechte Gesellschaft.

⁶⁷ Die Tabellen in Nordhaus u. Boyer [2000b] und Nordhaus u. Boyer [2000c] sind identisch.

⁶⁸ Wenn es nicht darauf ankäme, die Bedingungen einer wohlgeordneten liberalen Gesellschaft – unter Beachtung möglicher anderer konkurrierender ethischer Ziele – möglichst früh zu erreichen, würden in einer Welt, die durch Modelle vom Typ des RICE-Modells für Fragen der intergenerationellen und internationalen Gerechtigkeit angemessen abgebildet wird, sowohl die Pflicht zur Akkumulation wie die zu einer auch materiellen Beistandspflicht zu hohlen Worten verkommen, da in jedem Fall ausreichend akkumuliert wird, um ein entsprechendes – angemessen bestimmtes – Wohlstandsniveau *irgendwann* zu erreichen.

⁶⁹ Entsprechendes gilt natürlich auch für alternative Niveaus eines für die Einrichtung einer wohlgeordneten liberalen Gesellschaft notwendigen Pro-Kopf-Einkommens.

⁷⁰ Unter Voraussetzung eines exponentiellen Wachstums innerhalb der Zehnjahresperioden.

Wenn wir uns allein auf das Kriterium des frühestmöglichen Erreichens der Bedingungen einer wohlgeordneten liberalen Gesellschaft stützen würden, wäre die Entscheidung klar: Von den zur Wahl stehenden Alternativen (zu denen das wunderbare Verschwinden der Umweltprobleme nicht zu rechnen ist), wäre Kyoto ohne Zertifikatshandel die beste, dicht gefolgt von Kyoto mit Handel nur zwischen den OECD-Ländern.⁷¹

Während aus einer Rawlsschen Perspektive – wenigstens in der hier diskutierten Interpretation – die Tatsache, daß es Strategien gibt, die nach dem Erreichen des Ziels der Akkumulation ein höheres Bruttosozialprodukt pro Kopf gewährleisten, ethisch irrelevant ist, sind die Belastungen, die früheren Generationen auferlegt werden, aber relevant. Deshalb kann der Zeitpunkt der Zielerreichung nicht das einzige Kriterium sein, und das Niveau des Bruttosozialprodukt in früheren Generationen muß in irgendeiner Form berücksichtigt werden.

Wieder stehen wir also vor der Aufgabe verschiedene Ziele gegeneinander abzuwägen, und es stellt sich die Frage, ob wir hiermit einen Punkt erreicht haben, wo intuitives Abwägen im Ende unausweichlich ist.

Allerdings ist die einzige Strategie, die einer früheren Generation der am meisten belasteten Region (der „low income“ Ländergruppe) ein höheres Bruttosozialprodukt pro Kopf gewährt als „Kyoto ohne Handel“, die Verschiebung der „optimalen“ Strategie um 10 Jahre, die in der ersten Periode für die ärmsten Länder ein *minimal* höheres Bruttosozialprodukt pro Kopf gewährleistet. Der Unterschied ist außerdem so klein (weniger als 0,003 Cent), daß das Abwägen leicht fällt, da in allen anderen Perioden vor Erreichen des Zielpunkts das Bruttosozialprodukt pro Kopf der Strategie „Kyoto ohne Handel“ *vergleichsweise* deutlich über dem der verschobenen Optimalstrategie liegt (zwischen 0,88 Cents und etwas unter 13 Dollar).⁷²

7.3.6 Umverteilung

Die Unterschiede zwischen den verschiedenen Strategien, die in der hier vorgeschlagenen Interpretation der Rawlsschen Theorie der Gerechtigkeit und des Rechtes der Völker den Ausschlag für und gegen einzelne Strategien geben, sind vielleicht überraschend gering: Einige Monate in der Erreichung des Zielpunkts der Akkumulation bzw. einige wenige Dollar in der Höhe des Bruttosozialprodukts pro Kopf.

⁷¹ Dieses Ergebnis ist robust gegenüber Variationen des gewählten Zielpunkts der Beistandspflicht. Aus Abbildung 7.2 läßt sich entnehmen, daß von der zweiten bis zur vierundzwanzigsten Periode durchgängig Kyoto ohne Zertifikatshandel die Produktion pro Kopf der ärmsten Region maximiert. Daraus ergibt sich ein Korridor für den Zielwert der Beistandspflicht von 605 \$ bis 9 873 \$.

Es zeigt sich hier übrigens eine überraschende Nähe zwischen dem Ansatz eines internationalen Differenzprinzips und dem Kriterium der Rawlsschen Beistandspflicht, das man durchaus als die systematische Antwort auf das Problem der Abwägung zwischen Strategien, die zu verschiedenen Zeitpunkten gemäß dem Differenzprinzip als beste ausgezeichnet sind.

⁷² Da in beiden Strategien in der ersten Periode die ärmsten Länder in keiner Weise zum Klimaschutz beitragen, ist einzige Faktor, der den Unterschied des Bruttosozialprodukts in der ersten Periode begründet, der Knappheitspreis („Hotelling rent“) für Energiequellen. Der leicht höhere Grades des Klimaschutzes im Fall der verschobenen Optimalstrategie führt zu einem niedrigeren Knappheitspreis in der ersten Periode.

Ein Grund dafür ist, daß die Schäden des Treibhauseffekts im RICE-Modell eher klein sind. In den Worten von Nordhaus u. Boyer [2000c, 142]: „... the impact of warming upon the global economy is relatively small, amounting to around 2 percent of global output for a 2.5 °C average warming.“⁷³ Angesichts der doch beachtlichen Wachstumsraten⁷⁴ übersetzen sich die geringen Unterschiede in der jährlichen Produktion dann in auch in nur wenige Jahre (unter vier), die durch Umweltschutz für eine frühere Zielerreichung maximal gewonnen werden können.

Ein weiterer Grund liegt in der Natur der von Nordhaus untersuchten Strategien, die entweder den Temperaturanstieg im Vergleich mit dem Referenzfall nur ganz unwesentlich begrenzen oder die ärmsten Länder an den Kosten des Umweltschutzes gleichmäßig beteiligen (oder beides). Nur die vier direkten Varianten des Kyoto-Protokolls stellen die ärmsten Länder von den Kosten des Umweltschutzes frei⁷⁵, aber der Unterschied im Temperaturanstieg bis 2105 beträgt maximal 0,06 °C (für Kyoto ohne Zertifikatshandel⁷⁶). Die stärkste Einschränkung des Temperaturanstiegs gegenüber dem Referenzfall ergibt sich (wie zu erwarten) bei der Strategie der Beschränkung des gesamten Temperaturanstiegs auf 1,5 °C (1,06 °C), aber hier werden die ärmsten Länder erheblich durch die Kosten belastet, die dafür anfallen.⁷⁷ Da allerdings die Schäden insgesamt so klein sind und die ärmsten Länder selbst mit der Zeit erheblich zur Emission von CO₂ beitragen, wären auch bei vollkommener Freistellung der ärmsten Länder und aggressiver Umweltpolitik im Rest der Welt der Zeitpunkt der Zielerreichung nur wenig nach vorne verschoben.

Ein letzter Grund ist, daß die betrachteten Strategien alle keine wirkliche Umverteilungskomponente enthalten. Wenn das RICE-Modell die Schäden des Treibhauseffekts korrekt modelliert, dann sind auch nach unserer Interpretation der Rawlsschen Kriterien die damit verbundenen Umweltschäden kein großes Problem. Das große Problem in dieser Perspektive ist, daß die ärmsten Länder mit China solange (ein rundes Jahrhundert) warten müssen, bis sie endlich die materiellen Bedingungen für eine wohlgeordnete liberale Gesellschaft erreicht haben. Diesem Umstand wäre durch großzügige Transfers abzuhelpen, die – soweit sie möglich sind, ohne die Bedingungen einer wohlgeordneten liberalen Gesellschaft in den Geberländern zu zerstören – aufgrund der Beistandspflicht dann auch geboten wären, ohne daß eine Abwägung gegenüber anderen ethischen Forderungen notwendig wäre: Es gibt ja – Rawls folgend – keinen Anspruch und kein Gebot eines Wachstums über das Erreichen der materiellen Bedingungen einer wohlgeordneten liberalen Gesellschaft hinaus.

Diese Überlegungen legen es nahe, im Vergleich Strategien zu betrachten, die auf das

⁷³ Vgl. auch Nordhaus u. Boyer [2000c, 96].

⁷⁴ Von z. B. 40 % bis 14 % in zehn Jahren für die Low-Income Länder bis 2095.

⁷⁵ Das ist der Grund dafür, daß sie relativ gut abschneiden, wenn das Wohl dieser Länder entweder gleich gewichtet wird wie das Wohl der reichen Länder, oder wenn es gar zum eigentlich entscheidenden Kriterium gemacht wird.

⁷⁶ Deshalb schneidet diese Strategie unter den vier Kyoto-Varianten am besten ab, wenn das Wohl dieser Länder gleich gewichtet oder zum eigentlich entscheidenden Kriterium gemacht wird.

⁷⁷ Die „optimalen“ Strategien stellen die ärmsten nicht von den Kosten des Umweltschutzes frei und die Einschränkung des Temperaturanstiegs ist ähnlich unbedeutend wie im Fall der Kyoto-Varianten.

Wohl armer Gesellschaften hin konstruiert sind, indem sie ein großzügiges Element der Umverteilung enthalten. Natürlich gibt es unendlich viele solcher Strategien. Da das RICE-Modell keine Kosten der Transfers modelliert⁷⁸, ergeben sich aus ihm keine Grenzen für mögliche Transfers diesseits des ethisch Angezielten und theoretisch Möglichen. Damit paßt das Modell gut in den Rahmen einer Theorie der vollständigen Konformität. In der realen, von Ungerechtigkeiten geprägten Welt⁷⁹ kann man aber nicht davon ausgehen, daß Transfers von den Gebergesellschaften einfachhin in dem Maß gegeben werden, wie dies den Forderungen der Gerechtigkeit entspricht, noch auch davon, daß in den Empfängergesellschaften Transfers in der ethisch geforderten Weise eingesetzt werden. Angesichts der realen Schwierigkeiten einer weltweiten signifikanten Umverteilung und da der Transferprozeß und seine möglichen Kosten nicht Teil des Modells sind, werde ich zunächst (allerdings sehr weite) Grenzen *ad hoc* in die im folgenden diskutierte Alternativstrategie einbauen. Später werde ich dann jedoch noch einmal auf die Perspektive einer idealen Theorie zurückkommen.⁸⁰

Ich werde nur eine solche Alternativstrategie diskutieren. Die betrachtete Alternativstrategie verbindet, was Ökonomen „Optimalität“ nennen, mit einem zwar *erheblichen* Maß an materiellen Transfers, das aber von den Geberländern insofern keine *Einschränkungen* verlangt, als nur aus Einkommenszuwächsen umverteilt wird. Zwar gibt es keinen ethisch fundierten Anspruch auf die Nutzung von Vorteilen, die das Resultat ungerechter Strukturen sind (und Rawls betrachtet unsere Welt heute als von extremer Ungerechtigkeit gekennzeichnet [Rawls 1999d, 117]), aber die Strategie berücksichtigt die menschliche Tendenz, Besitzstände als ein moralisches Recht einzufordern,⁸¹ und berücksichtigt so in der politischen Realität ohne Zweifel vorhandene Widerstände gegen Formen der internationalen Umverteilung, die mit einer wahrnehmbaren tatsächlichen Einschränkung des Konsums in reichen Ländern verbunden wären.

Genauer stellen alle Regionen beginnend mit der Periode, wo sie (im Modell) das als Zielpunkt der Akkumulation bestimmte Bruttosozialprodukt pro Kopf überschritten haben (Periode A), einen für jede Periode festgelegten Prozentsatz *des* Teils des Outputs (abzüglich der Umweltschäden) zur Verfügung, der über das Niveau hinausgeht, das nötig wäre, um das Bruttosozialprodukt pro Kopf auf dem Niveau der Periode A zu halten. Die auf diese Weise aufgebrauchten Mittel werden dann auf China und die ärm-

⁷⁸ Das ist in keiner Weise als eine Kritik des Modells gedacht. Zum einen war die halbwegs realistische Darstellung eines Umverteilungsprozesses nicht Teil der Fragestellung des Modells. Zum anderen ist es fraglich, ob es überhaupt sinnvoll ist, im Rahmen einer normativen Diskussion Grenzen der Umsetzbarkeit ethischer Forderungen in der Form eines Modells zu *fixieren*, da dies allzuleicht als Vorwand gegen die Ansprüche der Gerechtigkeit mißbraucht werden kann. Vielleicht wäre es eher die Aufgabe ethisch motivierter Wissenschaftler und Politiker, ausgehend von den ethischen Forderungen im realen politischen Prozeß die Grenzen der Umsetzbarkeit zu testen und zu einer Ausweitung beizutragen. Die Auswahl möglicher Strategien ist dann eine Frage der auch politischen Klugheit und nicht einfach der Wissenschaft. Ein solches Vorgehen würde außerdem dem Vorrang korrespondieren, den Rawls der idealen Theorie einräumt, wenn es von einer auf der idealen Theorie aufbauenden ethischen Reflexion dieser Strategien begleitet wäre.

⁷⁹ So jedenfalls [Rawls 1999d, 117].

⁸⁰ S. u. 245

⁸¹ Empirische Evidenz für diese Tendenz (den „entitlement effect“) findet sich u.a. bei Kahnemann u. a. [1986] und Gaechter u. Riedl [2002]. Dort auch weitere Literaturverweise.

sten Länder so umverteilt, daß deren Bruttosozialprodukt pro Kopf gleich ist. Nachdem China ein Bruttosozialprodukt pro Kopf (ohne Transfers) erreicht hat, das den Zielwert überschreitet, gehen alle Transfers an die ärmsten Länder, bis diese ebenfalls selbstständig den Zielwert des Bruttosozialprodukts pro Kopf erreicht haben.

Auf diese Weise beginnen die USA, OECD Europa, andere Länder mit hohem Einkommen und Länder mittleren Einkommens 2005 mit Transfers von 90 % (sic!) des oben definierten Zuwachses. 2015 werden weiter 90 % des Zuwachses umverteilt, wobei Osteuropa und Russland mit Transfers beginnen. Ab 2025 zahlen auch die Länder mit niedrigerem mittleren Einkommen Transfers und die Prozentsätze beginnen zu sinken (zunächst auf 83 %, dann schrittweise bis auf 4 % 2085, der letzten Periode, in der Transfers stattfinden).⁸²

Die Transfers sind, wie schon erwähnt, erheblich: Für die Regionen hohen und mittleren Einkommens erreichen sie relativ zum Bruttosozialprodukt ihren Höhepunkt 2025 mit um die 25 % des Outputs für die Regionen hohen Einkommens, knapp 36 % für die Länder mittleren Einkommens und um 30 % für Osteuropa und Russland.⁸³ Die Länder mit niedrigem mittleren Einkommen zahlen den höchsten Teil des Outputs 2035 (knapp 18 %). Auch in der letzten Periode, in der Transfers geleistet werden, betragen diese noch zwischen 2 % und 3 % des Outputs der Geberländer.

Um einen Eindruck für die Größenordnung zu gewinnen, kann man diese Transfers mit der tatsächlich geleisteten Entwicklungshilfe der Mitgliedsländer des DAC⁸⁴ vergleichen, die von den 80er Jahren bis heute insgesamt jeweils deutlich unter 0,4 % des Bruttosozialprodukts der Geberländer lag, und seit Mitte der 90er Jahre sogar deutlich unter 0,3 %. [Vgl. OECD 2004, 30] Auch die vergleichsweise großzügigen Vorschläge der „Global Marshall Plan Initiative“ würden nur einen kleinen Anteil der hier betrachteten Transfers decken. Sie zielen auf eine – allerdings als unzureichend betrachtete – (ungefähre) Erfüllung der 0,7 % Zusage bis 2015. Nach diesen Vorschlägen würden im Zeitraum von 2005 bis 2014 insgesamt 1,394 Billionen Dollar Entwicklungshilfe geleistet.⁸⁵ Nach unserem Schema würden im selben Zeitraum 48,9 Billionen gezahlt, also fast das 38-fache. [Vgl. Rademacher 2004, 143-148]

Sowohl nach dem egalitär-utilitaristischen Kriterium, wie der hier vorgelegten Konkretisierung der Rawlsschen Theorie der Gerechtigkeit ist diese Umverteilungsstrategie deutlich die beste von allen betrachteten Strategien: Einschließlich der Transfers haben alle Regionen die materiellen Bedingungen einer wohlgeordneten liberalen Ge-

⁸² Die Prozentsätze sind zunächst willkürlich gewählt und wurden dann so nach unten angepaßt, daß nach Überschreiten des Zielwerts das gesamte Sozialprodukt (inklusive der Transfers) pro Kopf nur sehr leicht wächst, und es zu einer relativ „sanften Landung“ in der Periode kommt, in der die Transfers eingestellt werden.

⁸³ Dieser Prozentsatz für Länder mittleren Einkommens könnte unfair erscheinen. Es wäre möglich, ein alternatives Beitragsschema zu finden, bei dem diese Länder entlastet und die Länder mit hohem Einkommen stärker belastet werden. Für den Zweck des Beispiels im Text sind solche Detailfragen aber unerheblich. Immerhin ist zu bemerken, daß die Rawlsschen Kriterien gegenüber solchen Umverteilungen unter Gesellschaften, die die Bedingungen einer wohlgeordneten Gesellschaft schon erfüllt haben, unsensibel sind. Von einem egalitären Utilitarismus wären sie als positiv zu beurteilen.

⁸⁴ Ausschuß für Entwicklungshilfe der OECD

⁸⁵ Wobei bis 2007 nur schon faktisch zugesagte Leistungen vorausgesetzt werden.

sellschaft schon 2025 erreicht, d. h. gute 70 Jahre vor dem Zeitpunkt, der bei den anderen Strategien zu erwarten war. Beeindruckend ist auch der Wohlfahrtsgewinn, wenn die Nutzen der Bewohner der verschiedenen Regionen gleich gewichtet werden: Der Nutzengewinn über den Referenzfall ist mehr als 2600mal so groß wie im Fall der sogenannten optimalen Strategie ohne Umverteilung. Selbst im Vergleich mit dem Geoengineering (das auf ein Wegzaubern des Treibhauseffekts hinausläuft) wäre der Nutzengewinn noch fast 30mal größer.

Nordhaus u. Boyer [2000c, 123] geben einen Hinweis zur Interpretation ihrer „optimalen“ Strategie:

It will be useful to provide a word of interpretation of the optimal case. This is not presented in the belief that an environmental pope will suddenly appear to provide infallible canons of policy that will be scrupulously followed by all. Rather, the optimal policy is provided as a benchmark for policies to determine how efficient or inefficient alternative approaches may be.

Ein ähnlicher Hinweis ist auch hier angebracht: Es wäre naiv zu glauben, daß eine Strategie, die sowohl aus der Sicht eines egalitären Utilitarismus wie aus der Sicht einer (bestimmten) Interpretation der Rawlsschen Theorie der Gerechtigkeit als in hohem Maße erwünscht erscheint, deshalb auch realisiert werden wird. Aber auch eine solche Strategie kann als „benchmark for policies“ dienen. Sie zeigt dann, welche Elemente politischer Strategien vielversprechend sind und welche nicht. Genauer zeigt sie, wie wenig von dem Bemühen um Effizienz und wie viel von dem Bemühen um effektive Umverteilung zu erwarten ist.

Die betrachtete Strategie, die großzügig zugunsten der Armen umverteilt, ist aus der Sicht der hier betrachteten egalitären Theorien in hohem Maße ethisch erwünscht, aber ihrer politische Umsetzung ist nicht zu erwarten. Die relevanten Grenzen der Umverteilung sind also durch die faktische politische Umsetzbarkeit gesetzt und nicht durch die den Theorien inhärenten Grenzen der Wünschbarkeit von Umverteilung. Die betrachtete Strategie gibt zwar eine klare Richtung vor. Für Vorschläge, die auf eine konkrete politische Umsetzung zielen, wird man sich aber mit bescheideneren Ergebnissen zufrieden geben müssen. Auch die genaue Ausgestaltung politisch viabler Strategien wird nicht einfach durch die ethischen Vorgaben, sondern auch durch das politisch Mögliche bestimmt sein.⁸⁶

7.3.7 In the long run ...

Die Schäden, die im RICE-Modell durch den Treibhauseffekt bewirkt werden, sind sehr moderat. Und es fällt schwer, sie mit den alarmierenden Szenarien in Verbindung zu bringen, die von Vertretern einer strikten Politik der Bekämpfung der Emission von Treibhausgasen in die politische Diskussion eingebracht werden.

⁸⁶ Vgl. oben Anmerkung 78.

Einen weiteren Beleg für die Harmlosigkeit des Treibhauseffekts im RICE-Modell liefert ein Blick auf mögliche Zeitpfade in einer mittleren langfristigen Sicht von mehreren tausend Perioden. Wenn wir etwa annehmen, daß bis in das Jahr 2285 dem Referenzfall gefolgt wird, d. h. keine aktive gemeinsame Politik der Emissionsbeschränkung unternommen wird, und von da an nur noch der Kapitalstock und der erreichte Pro-Kopf-Konsum erhalten wird⁸⁷, dann führt diese „steady state“ Politik zunächst über mehrere Jahrhunderte zu sinkenden Durchschnittstemperaturen (von 4,539 °C über dem vorindustriellen Niveau im Jahr 2285 bis leicht über 1,568 °C im Jahr 8455) und in der Folge dann zu *sehr langsam* steigenden Durchschnittstemperaturen. So läge etwa in Periode 25 000 (Jahr 251985) die Temperatur 1,583 °C über dem vorindustriellen Niveau, in Periode 50 000 (Jahr 501985) 1,598 °C und in Periode 100 000 (Jahr 1001985) 1,628 °C. Auch die jährlichen Klimaschäden würden sich engen Grenzen halten: Die ärmsten Länder etwa würden in dieser letzten Periode weniger als 2,29 % der Bruttoproduktion verlieren.

Nordhaus u. Boyer [2000c, 56] warnen nun jedoch zur Vorsicht, wenn man ihr Modell für langfristige Voraussagen verwenden möchte. Die Gründe⁸⁸, die sie dafür anführen, berühren aber meine Aussage über die harmlosen Folgen des Treibhauseffekt im RICE-Modell in dem, was ich eine „mittlere lange Frist“ genannt habe, nicht. Zum einen verweisen sie auf das Fehlen einer „backstop technology“, was zu unvernünftigen Ergebnissen führen könne, wenn die Vorräte an fossilen Energieträgern ausgehen. Dieses Argument bedeutet aber einfach, daß die Autoren den Treibhauseffekt für noch harmloser halten, als er in ihrem Modell erscheint: Das Modell überschätzt den Treibhauseffekt, weil es den Umstieg auf Energieformen, die nicht mit Emissionen von Treibhausgasen verbunden sind, vernachlässigt. Für eine frühere Version des RICE-Modells hat Nordhaus [1999, 22, 41] einen Preis der Backstop-Technologie von 300 Dollar pro Tonne Kohlenstoffenergie angenommen. Damit würde in meiner Projektion ungefähr in der Periode 25070 (Jahr 252685) mit dem Umstieg auf die Backstop-Technologie begonnen. Von da an würde meine Projektion also die Schäden des Treibhauseffekts überschätzen. In Nordhaus u. Boyer [2000c, 55] sprechen die Autoren allerdings von einer Obergrenze für den Preis für Kohlenstoffenergie von 500 Dollar pro Tonne. Wenn damit wieder der weltweit gemeinsame Preis gemeint ist, dann wären die Bedingungen für den Umstieg auf die Backstop-Technologie selbst im Jahr 1001985 – bei einem Knappheitspreis für fossile Energieträger um 351,54 Dollar pro Tonne Kohlenstoffenergie – noch nicht gegeben.⁸⁹ Der zweite Grund ist die Gefahr, daß sich eine langfristige Projektion aus dem Bereich entfernen könnte, über den das Modell kalibriert wurde. Aber auch davon kann bei einem insgesamt Temperaturanstieg von etwa nur 1,628 °C nicht die Rede sein.⁹⁰

Der Blick auf die längerfristigen Eigenschaften des RICE-Modells ist in einer rawlsia-

⁸⁷ Dieser beträgt für die Region der ärmsten Länder dann 10 512 \$ jährlich.

⁸⁸ Die Autoren kombinieren das, was ich hier als zwei Gründe ausführe, in ihrer Formulierung zu *einem* Argument.

⁸⁹ Der kumulierte Verbrauch würde 4403,59 GtC bis 501985 und 4584,25 GtC bis 1001985 betragen, deutlich unter der von den Autoren in der Angebotsfunktion implizierten Knappheitsgrenze von 6000 GtC.

⁹⁰ Die Schadensfunktionen sind auf einen Bereich zwischen 0 °C und 6 °C hin kalibriert. [Nordhaus u. Boyer 2000c, 94 f.]

nischen Perspektive wichtig, weil für Rawls' Gerechtigkeitsauffassung die Idee „von der Gesellschaft als einem fairen und langfristig von einer Generation zur nächsten fortwirkenden System der sozialen Kooperation“ [Rawls 2003, 25] grundlegend ist. Deshalb wäre ein Zeitpfad, der nur kurz- oder mittelfristig die materiellen Bedingungen einer wohlgeordneten Gesellschaft sichert, dann aber diese Bedingungen wieder zerstört, mit seiner Theorie der Gerechtigkeit nicht vereinbar.⁹¹

Da die Rawlssche Theorie der Gerechtigkeit kein ständiges Wachstum verlangt, zeigen die hier vorgelegten Projektionen für eine mittlere langfristige Periode, daß auch bei einem sehr großzügig angesetzten notwendigen Niveau der materiellen Bedingungen einer wohlgeordneten Gesellschaft diese Bedingungen weit in eine Zukunft hin gesichert werden können, die jenseits eines Zeithorizonts liegt, für den wir sinnvollerweise planen können. Jedenfalls ist das so, wenn wir meinen, daß das RICE-Modell die relevanten Faktoren, insbesondere auch die vom Treibhauseffekt ausgehenden Bedrohungen, korrekt modelliert hat.

7.3.8 Unsicherheit, Katastrophen und Tote

In dem harmlosen Bild, das das RICE-Modell vom Treibhauseffekt zeichnet, kommen scheinbar einige Elemente nicht vor, die ansonsten in der Auseinandersetzung doch eine große Rolle spielen: Die verschiedenen Zeitpfade, die wir betrachtet haben, sind – gegeben die gewählte Strategie – determiniert und sicher, *Unsicherheit* spielt – so scheint es – keine Rolle; bei im betrachteten Zeitrahmen stetig wachsendem oder auf einem großzügigen Niveau fixierten Pro-Kopf-Konsum fällt es schwer, von *Katastrophen* zu reden, nur weil ohne Treibhauseffekt der Konsum noch höher liegen könnte; wir begegnen im Modell nur monetären und Nutzengrößen, Ereignisse und Vorgänge wie *Flucht vor Umweltkatastrophen, Krankheiten, Hungern, Sterben* kommen in dem Modell nicht vor.

Dieser Eindruck ist korrekt, wenn wir auf das Modell schauen, mit dem die Autoren ihre Projektionen produzieren, mit deren Hilfe sie dann die verschiedenen Strategien bewerten. Metaphorisch gesprochen „hinter“ oder „vor“ dem Modell liegt allerdings der Versuch, alle diese Elemente in einer (dem aktuellen „Forschungsstand“ entsprechenden⁹²) sinnvollen Weise einzufangen. Insbesondere versuchen die Autoren, in den im Modell enthaltenen Schadensfunktionen nicht nur marktbewertete Schäden zu erfassen (unter den Kategorien „Landwirtschaft“, „Erhöhung des Meeresspiegels“, „Andere Marktsektoren“ [Nordhaus u. Boyer 2000c, 74-78]) sondern auch nichtmarktliche Aspekte angemessen zu berücksichtigen. [Vgl. Nordhaus u. Boyer 2000c, 71] Neben nicht marktbewerteten Einflüssen auf das Freizeitverhalten [Nordhaus u. Boyer 2000c, 82-85], die wir für unsere Zwecke außer acht lassen können, geht es dabei um Gesundheit, menschliche Ansiedlungen und Ökosysteme und das Risiko katastrophaler Ent-

⁹¹ Das ist auch der bestimmende Gesichtspunkt bei der Festlegung der gerechten Sparquote für eine wohlgeordnete liberale Gesellschaft. Vgl. Rawls [2001, 159] u. Rawls [1999g, 255].

⁹² Hinweise auf aktuelle Studien, den derzeitigen Forschungsstand, gegenwärtige Tendenzen unter Forschern u.ä. finden sich im ganzen Kapitel 4 von Nordhaus u. Boyer [2000c], das der Herleitung der Schadensfunktionen gewidmet ist.

wicklungen. [Nordhaus u. Boyer 2000c, 78-82, 85-89] Alle Schadenskategorien werden (im Prinzip⁹³) über die (angenommene) Zahlungsbereitschaft zu ihrer Verhinderung monetär bewertet, für jeweils die einzelnen Regionen aufsummiert und schließlich an quadratische Schadensfunktionen angepaßt, die dann in das Modell eingehen. [Nordhaus u. Boyer 2000c, 71-74, 94 f.]

Man kann den Autoren nicht vorwerfen, sie würden ihre Leser über die *praktischen* Unsicherheiten und Schwierigkeiten ihres Versuches, die Auswirkungen der verschiedenen Kategorien des Treibhauseffekts zu quantifizieren, im Unklaren lassen. Es fehlt nicht an Hinweisen auf die vielen Bereiche (regional und kategorial), in denen halbwegs verlässliche Schätzungen fehlen, die Ungenauigkeit bestehender Schätzungen und die vielen ad hoc Annahmen und bewertenden Entscheidungen, mit denen sie selbst fehlende fundiertere Schätzungen ersetzen. [Z. B. Nordhaus u. Boyer 2000c, 69-71, 75, 77, 78, 80-82, 85-89] Ja, abschließend betonen die Autoren den „hoch spekulativen“ Charakter ihrer Schätzungen. [Nordhaus u. Boyer 2000c, 98] Doch darf dies nicht darüber hinwegtäuschen, daß *konzeptionell* ihr Vorgehen bedeutet, daß – *insoweit* die Schätzungen korrekt sind – die monetisierten Schadensbeträge und die darauf aufbauenden Schadensfunktionen die ursprüngliche Information *ersetzen*. Wenn der „willingness to pay approach“ korrekt ist und die Zahlungsbereitschaften korrekt geschätzt sind, dann wäre es ein Fehler, die Ergebnisse einer darauf aufbauenden Kosten-Nutzenanalyse mit dem Hinweis auf die ursprüngliche, nicht monetisierte Schadensinformation in Frage stellen zu wollen. Ein solcher Zweifel muß entweder heißen, daß die Zahlungsbereitschaften nicht korrekt geschätzt worden sind, oder daß dem „willingness to pay approach“ eben *das* nicht (vollständig) gelingt, wozu er eigentlich da ist: nicht marktlich bewertete Güter, Werte und Zielvorstellungen untereinander und mit marktlichen Gütern vergleichbar zu machen.

Damit bleibt der ersten Eindruck einer Harmlosigkeit des Treibhauseffekts im RICE-Modell, die schlecht mit der Idee von Katastrophen und tödlichen Konsequenzen vereinbar zu sein scheint, auch auf den zweiten Blick bestehen, und auch, wenn man berücksichtigt, daß das RICE-Modell durchaus in gewisser Weise das Risiko katastrophaler Entwicklungen, die Zerstörung von menschlichen Siedlungsgebieten und menschliches Sterben aufgrund des Treibhauseffektes modelliert und erfaßt. Das geschieht in eben der Weise, daß ein Modell konstruiert wird, in dem nur noch sichere monetäre und Nutzengrößen vorkommen, und das (bei allen praktischen Abstrichen) der ursprünglichen Welt mit Risiko, Zerstörung und Tod für Zwecke der Beurteilung von verschiedenen Umweltstrategien äquivalent sein soll. Nur insoweit diese Äquivalenz gilt, können die normativen Konsequenzen des Modells auch für die Realität Relevanz besitzen. Das Modell für die real anstehenden Entscheidungen relevant zu halten, heißt deshalb nichts anderes, als auch die moderaten Umweltschäden dieses Modell für eine adäquate Erfassung der realen Gefahren zu halten.

⁹³ Faktisch sind die konkreten Bewertungsmethoden für die unterschiedlichen Schadensformen sehr verschieden. Darin sehen die Autoren kein prinzipielles Problem: „The advantage of the WTP approach is that it can encompass different approaches to measuring impacts (including surveys as well as statistical impact measures).“ [Nordhaus u. Boyer 2000c, 71]

7.3.8.1 Gesundheit

Bei der Behandlung gesundheitlicher Folgen des Treibhauseffekts nehmen Nordhaus u. Boyer [2000c] nur eine kleine Liste von Krankheiten (Malaria, Dengue und sechs tropische Krankheiten) und die direkten Folgen der Luftverschmutzung in den Blick und modellieren, wie die Entwicklung der aufgrund dieser Faktoren verlorenen Lebensjahre (YLL, Years of Life Lost) durch die atmosphärische Temperatur beeinflusst wird. Genauer gesagt: Sie gehen von den von Murray u. Lopez [1996] geschätzten regionalen Werten und Projektionen der YLLs für die berücksichtigten (neun) Faktoren für 1990 und 2020 aus und bauen darauf eine eigene Schätzung des Prozentsatzes des Brutto-sozialprodukts auf, der bei einer globalen Erwärmung um 2,5 °C verlorenggeht, wenn man ein verlorenes Lebensjahr mit dem zweifachen Jahreseinkommen pro Kopf der jeweiligen Region bewertet. Diese Schätzungen übertragen sie dann direkt von den acht Regionen der Studie von Murray und Lopez auf die dreizehn Regionen einer früheren Version des RICE-Modells und schätzen dann die Parameter einer halblogarithmischen Gleichung, die den prozentualen Verlust des Brutto-sozialprodukts in Abhängigkeit von der atmosphärischen Temperatur bestimmt.⁹⁴

Die Vorgehensweise der Autoren kann aus einer Reihe von Gründen kritisiert werden:

1. Das genaue Verfahren der Bestimmung der gesundheitlichen Folgen des Treibhauseffekts ist in vielen Details angreifbar. Es gehen eine Reihe von willkürlichen und/oder für den Leser nicht nachvollziehbare Annahmen in dieses Verfahren ein, und im ganzen wirkt das Verfahren doch eher grob.⁹⁵ Bevor man sich dies allerdings als eine Kritik an der Vorgehensweise der Autoren zu eigen macht (statt als ein Hinweis auf den sehr vorläufigen Charakter solcher Schätzungen), sollte man berücksichtigen, daß die Autoren ihre Vorgehensweise als einen – praktika-blen – Ersatz für – fehlende – „systematic estimates of health impacts“ zu verstehen scheinen [Nordhaus u. Boyer 2000c, 80], und auch die inhärenten Schwierigkeiten der Erstellung verlässlicherer Schätzungen gebührend gewichten. [Vgl. Nordhaus u. Boyer 2000c, 71]
2. Die Autoren berücksichtigen nur einen Teil der gesundheitlichen Schäden, die vom Treibhauseffekt zu erwarten sind.
 - Sie erwähnen zwar die gesundheitlichen Folgen von häufigeren Überschwem-mungen und möglichen Auswirkungen auf die Wasserqualität [Nordhaus u.

⁹⁴ Zu weiteren Details der Vorgehensweise vgl. Nordhaus u. Boyer [2000c, 78-82, 93].

⁹⁵ Die Schätzung des Verlusts bei 2,5 °C ist z. B. der Durchschnitt dreier verschiedener Schätzungen. Eine dieser Schätzungen „judgmentally adjusts the change in YLLs for each subregion to approximate the difference among subregions that is climate related“, d. h. sie beruht auf nicht weiter nachvollziehbaren subjektiven Einschätzungen der Autoren. Eine der Schätzungen nimmt – willkürlich – an, daß durch eine Temperaturerhöhung von 2,5 °C 50 % der von Murray u. Lopez [1996] bis 2020 als Baseline erwarteten Verringerung der YLLs verlorenggeht. Der Prozentsatz für Afrika wird von den Autoren – mit einer plausiblen Begründung – nach unten angepaßt, der gewählte Wert (3 % statt der „eigentlich“ geschätzten 4,6 %) wird nicht weiter begründet. [Vgl. Nordhaus u. Boyer 2000c, 80-83]

Boyer 2000c, 80], beziehen sie aber – wie die gesundheitlichen Folgen von häufigeren Sturmkatastrophen, von Hunger und Dürre und von Verlusten menschlichen Habitats – in ihre Berechnungen nicht mit ein. Das könnte, was Überschwemmungen und Sturmkatastrophen angeht, vielleicht mit der Datenlage begründet werden. Unverständlich ist es, was die Faktoren Unterernährung und Verfügbarkeit von sicherem Wasser angeht, da Murray u. Lopez [1996, 312] für Unterernährung (nur Kinder) und für „bad water supply, sanitation and personal and domestic hygiene practices“ ähnliche Daten zur Verfügung stellen, wie die, die Nordhaus für Luftverschmutzung benutzt hat. [Murray u. Lopez 1996, 315] Alternativ hätte man wenigstens Proteinmangel und Durchfallerkrankungen in die Kategorie der (u.a.) klimaabhängigen Krankheiten aufnehmen können.⁹⁶

- Sie berücksichtigen nur den Verlust an Lebensjahren durch Todesfälle, und lassen den (zum Teil sehr großen) Verlust an Lebensqualität aufgrund des Lebens mit Krankheiten und ihren Folgen außer acht.
3. Die Weise, in der Nordhaus verlorene Lebensjahre bewertet und mit anderen Klimaschäden verrechnet (z. B. mit dem Freizeitwert von Wintersport oder Golf [Nordhaus u. Boyer 2000c, 84]) kann als problematisch angesehen werden. Jedenfalls scheint es mit egalitären ethischen Ansätzen kaum vereinbar, wenn die verlorenen Lebensjahre von Menschen verschiedener Regionen mit dem jeweiligen Pro-Kopf-Einkommen bewertet werden, und so das Lebensjahr eines Bürgers der USA im Referenzfall im Jahr 1995 mehr als das 56fache wert ist, wie das Lebensjahr eines Bürgers der ärmsten Region, und 2105 noch immer das 17fache.

Entsprechend unserer Fragestellung werde ich nur den letzten Punkt weiterführen und fragen, wie die aufgrund des Treibhauseffekts zu erwartenden Verluste an Lebensjahren in einer rawlsianischen Perspektive zu bewerten wären.⁹⁷ Dabei werde ich voraussetzen, daß die von Nordhaus verwendeten Schätzmethode(n) (hinreichend) korrekt die zu erwartenden Verluste bestimmen.

⁹⁶ Auch wenn die Gesamteinschätzung der Auswirkungen des Treibhauseffekts auf die landwirtschaftliche Produktion weltweit heute eher optimistisch ausfällt, kann ein verstärktes Auftreten von Unterernährung in besonders gefährdeten Gruppen und Regionen erwartet werden. Häufiger auftretende extreme Wetterereignisse werden in Entwicklungsländern die Gefahr von Hunger und Unterernährung erhöhen. Vgl. dazu IPCC [2001, 12, 43, 85, 258-270, 410, 460, 473, 500-507, 708, 721 f.]

Für den Zusammenhang zwischen Durchfallerkrankungen und Klimawandel vgl. IPCC [2001, 42, 50, 397, 454, 459, 472, 513, 571, 720 f.].

Die Zahl der Menschen, die in einer Situation des Wassermangels leben müssen, wird durch den Treibhauseffekt steigen und erhöhte Wassertemperaturen reduzieren die Wasserqualität. [IPCC 2001, 5, 31, 213 f.] Davon, wie auch von extremen Wetterereignissen gehen Risiken für die menschliche Gesundheit aus, auch wenn die genauere quantitative Zuschreibung aufgrund des Einflusses von anderen (sozioökonomischen) Faktoren schwierig bleibt. [IPCC 2001, 225, 471 f.]

⁹⁷ Für einen egalitären Utilitarismus würde man erwarten, daß verlorenen Lebensjahre gleich bewertet würden. Schwieriger ist die Frage zu beantworten, wie genau ein verlorenes Lebensjahr zu bewerten wäre. Abhängig von dieser Wahl können die gesundheitlichen Folgen mit einem sehr verschiedenen Gewicht in den utilitaristischen Kalkül eingehen.

Zur Beantwortung dieser Frage wird es notwendig, Rawls zu ergänzen: Teil seiner idealen Theorie ist das Verständnis einer Person als „someone who can be a citizen, that is a normal and fully cooperating member of society over a complete life“. [Rawls 1993b, 18] Die Probleme schwerer Krankheiten und Unfälle, und die damit einhergehende vorübergehende oder permanente Einschränkung der für die volle Kooperation notwendigen Fähigkeiten, werden von Rawls zunächst beiseitegelegt, er betrachtet sie als Probleme der Extension der idealen Theorie, die – zum Teil⁹⁸ – später behandelt werden können. [Rawls 1993b, 20 f., 182-185, 244 f., 272] Tatsächlich hat er das Problem der medizinischen Versorgung dann in Rawls [2001, 171-175] aufgegriffen, um in einem kurzen Abschnitt seinen Vorschlag einer Extension der Theorie der Gerechtigkeit auf das Problem von Krankheiten und Unfällen vorzulegen.

Demnach ist Frage der medizinischen und gesundheitlichen Bedürfnisse derjenigen, denen zeitweilig die Fähigkeiten fehlen, die zur normalen und vollen Kooperation notwendig sind, im gesetzgeberischen Stadium⁹⁹ zu entscheiden, da erst da die notwendigen Informationen zur Verfügung stehen. Auf der Grundlage von Informationen über die wahrscheinlichen medizinischen Bedürfnisse der am wenigsten begünstigten Gruppe und ihrer Kosten auf verschiedenen Niveaus der Behandlung und Versorgung, soll der erwartete Index an Grundgütern – einschließlich der erwarteten (man möchte ergänzen: *notwendigen*) Kosten der Behandlung und Versorgung – maximiert werden, wobei die anderen notwendigen Aufwendungen der Gesellschaft („einschließlich derjenigen [Ansprüche], die in sämtlichen Phasen des Lebens von der Kindheit bis ins Alter geltend gemacht werden“) gebührend zu berücksichtigen sind. Verschiedene Ansprüche sind also gegeneinander abzuwägen¹⁰⁰, wobei die Dringlichkeit verschiedener medizinischer und gesundheitlicher Ansprüche untereinander, und medizinischer und gesundheitlicher Ansprüche gegenüber anderen Ansprüchen (u.a.) dadurch bestimmt wird, inwieweit ihre Erfüllung dafür notwendig ist, die Gesundheit von Menschen wiederherzustellen, um ein „normales Leben als kooperierende Gesellschaftsmitglieder weiterzuführen“. Maßnahmen, die dazu dienen, besitzen eine „hohe Dringlichkeit – genauer gesagt, ihre Dringlichkeit ist durch das Prinzip der fairen Chancengleichheit bestimmt“¹⁰¹ [Rawls 2003, 265-268]

Rawls hält bei diesen Überlegungen an der Vorstellung von Bürgern „als Personen mit einer lebenslangen öffentlichen (politischen) Identität sowie als lebenslang normale und voll kooperierende Mitglieder der Gesellschaft“ fest¹⁰². Die Ansprüche sämtlicher Lebensphasen sind „vom Standpunkt *einer* Person“ zu „betrachten, die alle diese Lebensstadien durchläuft“. [Rawls 2003, 267] Die Frage von Menschen, die einen vorzeitigen Tod sterben, wird also zumindest explizit nicht behandelt. Es liegt jedoch nahe,

⁹⁸ Rawls hält es durchaus für möglich, daß Probleme, die über die „normale Gesundheitsfürsorge“ hinausgehen, im Rahmen seiner Theorie der Gerechtigkeit überhaupt nicht adäquat behandelt werden können. Da die Idee der politischen Gerechtigkeit nicht alles abdecken muß, spricht das nicht notwendigerweise gegen die Theorie. [Rawls 1993b, 21]

⁹⁹ Zum gesetzgeberischen Stadium s. o. S. 144.

¹⁰⁰ „Abwägen“ schließt hier die Möglichkeit strikter Vorrangsregeln ein.

¹⁰¹ Hier schließt sich Rawls Daniels [1985] an. Vgl. auch Rawls [1993b, 184].

¹⁰² Dieser Zustand wird durch Krankheiten und Unfall sozusagen nur „kurz“ unterbrochen.

zumindest ein Element aus seinen Überlegungen auch für die Frage eines vorzeitigen Todes zu übernehmen: Auch seine Verhinderung besitzt eine hohe Dringlichkeit, *zumindest* die der durch das Prinzip der fairen Chancengleichheit bestimmten. Wird doch durch den Tod ein Mensch in ähnlicher, wenn nicht radikalerer Weise daran gehindert, ein volles Leben als kooperierendes Mitglied einer Gesellschaft zu leben.

Aus zwei weiteren Gründen sind die Rawlsschen Überlegungen zur medizinischen Versorgung nicht direkt auf unsere Fragestellung zu übertragen:

1. Sie beziehen sich auf die interne Situation einer Gesellschaft und nicht auf die internationale Kooperation. Die Frage des Treibhauseffekts ist aber wesentlich eine internationale Frage, da keine Nation allein das Ausmaß der Temperaturveränderungen und damit des zu erwarteten Verlustes an Lebensjahren bestimmen kann. Wenn wir an der Rawlsschen *systematischen* Aufteilung einer auf einzelne Völker bezogenen Theorie der Gerechtigkeit und eines internationalen Rechts der Völker festhalten, gehört unsere Fragestellung also in der Bereich des Rechts der Völker, und es ist zu fragen, wie – angesichts des Verlusts an Lebensjahren – in einem zweiten Urzustand von Vertretern von Völkern die Kooperation in der Bekämpfung des Treibhauseffekts geregelt würde.
2. Bei Rawls geht es um die Befriedigung medizinischer und gesundheitlicher Ansprüche einer Generation in der Abwägung gegenüber Ansprüchen derselben Generation. Bei unserer Fragestellung geht es jedoch (im wesentlichen) um eine Frage der intergenerationellen Gerechtigkeit, da Ansprüche von über mehrere Jahrhunderte verteilten Generationen gegeneinander abgewogen werden müssen.

Für unseren Kontext nehme ich an, daß auch die Parteien im zweiten Urzustand der Verhinderung unzeitigen Sterbens eine hohe Dringlichkeit zuordnen werden. Genauer nehme ich an, daß die Erreichung einer wohlgeordneten Gesellschaft bzw. ihrer Bedingungen höchste Priorität hat, gefolgt von der Verhinderung eines vorzeitigen Todes. Erst an dritter Stelle kommt – wenn überhaupt – das Ziel einer Maximierung des Pro-Kopf-Konsums.¹⁰³ Ich nehme weiter an, daß den Parteien die relevanten Informationen über die verschiedenen Strategien des Umgangs mit dem Treibhauseffekt bekannt seien.

Zur Begründung für die Vorrangsregel möchte ich vor allem darauf hinweisen, daß dies plausiblerweise als eine rawlsnahe Ergänzung (Extension) seiner Theorie gelten kann: Durch die Vorordnung des Ziels der Erreichung der Bedingungen einer wohlgeordneten Gesellschaft wird sowohl der Vorrang der Freiheit wie auch das Gut einer wohlgeordneten Gesellschaft (d. h. das Gemeinwohl, das eben in der Verwirklichung und Fortentwicklung einer wohlgeordneten Gesellschaft besteht) respektiert. [Vgl. Rawls 2001, 198-202] Hinzu kommt, daß soziale Ungleichheiten ein wesentlicher Faktor für gesundheitliche Ungleichheiten zu sein scheint, sodaß man erwarten kann,

¹⁰³ Weitere Ziele eines Rechts der Völker spielen in unserem Zusammenhang keine Rolle.

daß soziale Gerechtigkeit im Sinne Rawls', Gerechtigkeit also, die gleiche Grundfreiheiten – inklusive einer effektiven Gleichheit politischer Grundfreiheiten – mit Chancengleichheit und dem Differenzprinzip verbindet, selbst wesentlich zu einem besseren Gesundheitszustand der Bevölkerung und einer höheren Lebenserwartung beitragen kann.¹⁰⁴ [Daniels u. a. 1999; Daniels 2001] Die (sekundäre) Vorordnung der Minimierung von verlorenen Lebensjahren vor der Erhöhung des Pro-Kopf-Konsums ist gut vereinbar mit der Verneinung des ständigen Wachstums als eines von der Gerechtigkeit geforderten Ziels und mit der hohen Dringlichkeit, die Rawls dem Ziel zumißt, Menschen die Fähigkeiten wiedererlangen zu lassen, die sie als kooperierende Mitglieder einer Gesellschaft brauchen. Die gewählte Einordnung des Zieles der Verhinderung frühzeitigen Sterbens im Rahmen des Rechts der Völker korreliert gut mit der Zwischenstellung, die Rawls selbst für Gesundheitsfürsorge innerhalb einer Gesellschaft – nachgeordnet der Sicherung der gleichen Grundfreiheiten, vorgeordnet der Bereitstellung der übrigen Grundgüter für die am meisten Benachteiligten – vorsieht.

Auch wenn wenig Zweifel daran bestehen kann, daß in einer rawlsianischen Perspektive die Verhinderung vorzeitigen Sterbens vieler Menschen ein dringliches Ziel darstellen wird, bleibt die Frage, ob die von Nordhaus in seiner Bestimmung der Schäden des Treibhauseffekts zugrundegelegten „verlorenen Lebensjahre“ (YLL) dafür den geeigneten Indikator darstellen. Zwei Eigenschaften lassen dies zweifelhaft erscheinen: Bei der Berechnung der YLL diskontieren Murray u. Lopez [1996] zukünftige Lebensjahre mit einem Diskontfaktor von 3 % ab und gewichten verschiedene Lebensjahre unterschiedlich.¹⁰⁵

Die Hauptgründe für die Diskontierung sind zum einen die (geringe) Wahrscheinlichkeit der Ausrottung einer ganzen Gesellschaft und zum anderen, daß ohne Diskontierung von früheren Generationen zu große Opfer verlangt werden könnten. [Murray u. Lopez 1996, 47, 51-53] Das erste Argument ist mit der Rawlsschen Sicht von Rationalität grundsätzlich vereinbar, wenigstens solange eine zuverlässige Basis für die Abschätzung der entsprechenden Wahrscheinlichkeiten vorliegt (vgl. Rawls [2001, 98, 101] und Rawls [1999g, 260, 362 f.]), doch ist eine Diskontierungsrate von 3 % dafür viel zu hoch. [So auch Murray u. Lopez 1996, 47] Das zweite Argument setzt einfach voraus, daß das Ziel von Gesundheitspolitik die Minimierung der intertemporalen, möglicherweise abdiskontierten Summe von DALYs¹⁰⁶ bzw. in unserem Zusammenhang, wo wir Nordhaus folgend von den Verlusten an Lebensqualität aufgrund von Krankheit absehen, von YLLs ist. Wenn das nicht schon als gegeben vorausgesetzt wird, dann kann auf die Frage der Abdiskontierung künftiger Lebensjahre dasselbe Argument angewandt werden, das Rawls gegen die Abdiskontierung im Fall der Bestimmung der Sparrate

¹⁰⁴ Vgl. dazu auch den Zusammenhang zwischen demokratischen Freiheiten und der Verhinderung von Hungersnöten. [Sen 1999, 187, 196-229]

¹⁰⁵ Murray u. Lopez [1996, 61] benutzen die kontinuierliche Gewichtungsfunktion $Cxe^{-\beta x}$, mit $C = 0,1658$ und $\beta = 0,04$. x ist das Lebens- bzw. Sterbealter. Ein Lebensjahr eines Neugeborenen hat demnach das Gewicht 0, das eines Einjährigen das Gewicht 0,159. Den höchsten Wert erreicht die Funktion für das Alter von 25 Jahren mit 1,525. Das Lebensjahr eines 75-jährigen hat immerhin noch das Gewicht 0,619, das eines 100-jährigen das Gewicht 0,304.

¹⁰⁶ „Disability-Adjusted Life Year“, die Summe von YLLs und YLDs (Years Lived with Disability).

und in der Folge auch gegen die intertemporale Nutzensumme als Kriterium vorgebracht hat:

We should observe that to reject pure time preference as a first principle is compatible with recognizing that a certain discounting of the future may improve otherwise defective criteria. For example, I have already remarked that the utilitarian principle may lead to an extremely high rate of saving which imposes excessive hardships on earlier generations. This consequence can be to some degree corrected by discounting the welfare of those living in the future. [...] Unhappily I can only express the opinion that these devices simply mitigate the consequences of mistaken principles. [...] having started with the idea that the appropriate rate of saving is the one which maximizes social utility over time (maximizes some integral), we may obtain a more plausible result if the welfare of future generations is weighted less heavily; [...] We may find that to achieve justice between generations, these modifications in the principle of utility are required. Certainly introducing time preference may be an improvement in such cases; but I believe that its being invoked in this way is an indication that we have started from an incorrect conception. [...] time preference has no intrinsic ethical appeal. It is introduced in a purely ad hoc way to moderate the consequences of the utility criterion. [Rawls 1999g, 262]

Die Hauptgründe für die Altersgewichtung sind zum einen ein empirisches Argument, nämlich eine in verschiedenen Studien und Befragungen beobachtete Präferenz für unterschiedliche Altersgewichtungen, und zum anderen der Verweis auf den altersabhängig unterschiedlichen Beitrag zur sozialen Wohlfahrt. [Murray u. Lopez 1996, 54-60] Das erste „Argument“ ist natürlich *ethisch* als solches überhaupt kein Argument¹⁰⁷: Es mag eine solche weitverbreitete Tendenz geben, aber auch die Mehrheit kann sich irren. Das zweite Argument hat in jedem Fall keinen Platz in einer rawlsianischen Theorie der Gerechtigkeit, in der unterschiedliche Beiträge zur sozialen Wohlfahrt über das Differenzprinzip, die Chancengleichheit, und die Erfüllung legitimer Erwartungen berücksichtigt werden, aber „Verdienst“ keine Ansprüche generiert. [Rawls 1999g, 88 f., 273-277] Lebensjahre verschieden zu gewichten, ist in dieser Sicht der falsche Weg, unterschiedliche Beiträge zu berücksichtigen, da die Grundstruktur einer Gesellschaft ja gerade die Kooperation von Bürgern organisieren soll, die sich als *gleich* und frei verstehen. Aber selbst wenn man in dem utilitaristischen Rahmen verbleibt, in dem Murray u. Lopez [1996] argumentieren, hängen die so rechtfertigbaren genauen Gewichte von

¹⁰⁷ Dies gilt insbesondere auch, wenn man sich die Rawlsche metaethische Theorie des reflektiven Gleichgewichts zu eigen macht. Ein weites reflektives Gleichgewicht verlangt zwar eine Auseinandersetzung mit alternativen Konzeptionen der Gerechtigkeit und dem Gewicht der verschiedenen Argumente, die für sie sprechen. [Rawls 2001, 31] Aber eine „alternative Konzeption der Gerechtigkeit“ ist mehr als ein empirisch beobachtete Präferenz (Rawls selbst beschränkt den Kreis zu betrachtende Alternativen sogar auf die „leading conceptions of political justice found in our philosophical tradition“) und eine sinnvolle Auseinandersetzung kann erst beginnen, wenn auch Argumente für die Alternative vorgelegt werden. Im übrigen kann das Resultat der Auseinandersetzung dann auch eine entschiedene Ablehnung sein.

den konkreten Annahmen über die Produktion von Gütern und anderen Beiträgen zur sozialen Wohlfahrt und vielen weiteren Faktoren (Wachstum, Verteilung, Nutzenfunktionen, Diskontierungsraten) in einer Weise ab, daß man nicht erwarten kann, daß sie zu den verschiedenen Kontexten, in denen man die Schätzungen der YLLs gebrauchen möchte, und den Annahmen der diese Kontexte darstellenden Modelle auch passen. Dies gilt natürlich auch für die von den Autoren gewählten tatsächlichen Gewichte, die so gewählt sind, daß sie die empirisch erhobenen Gewichte in etwa replizieren. [Murray u. Lopez 1996, 61]

Die oben gestellte Frage ist also zu verneinen: Die von Nordhaus zur Abschätzung der gesundheitlichen Folgen gewählten YLLs passen nicht eigentlich in einen rawlsianischen Rahmen.

Demgegenüber scheinen mir ungewichtete und nicht diskontierte YLLs gut in diesen Rahmen zu passen. Gegenüber der Möglichkeit eines frühzeitigen Todes ist es das wichtigste Ziel, Menschen ein *volles Leben* als kooperierende Mitglieder einer Gesellschaft zu ermöglichen. Da der Tod Teil des menschlichen Lebens ist und wir alle unvermeidlicherweise sterben, scheint es unvernünftig, jeden Tod als einen *vorzeitigen* Tod zu werten, der ein volles Leben verhindert hat. Das gilt insbesondere für den Tod eines Menschen, der alle Phasen des menschlichen Lebens („from childhood to old age“ [Rawls 2001, 174]) durchlaufen hat. Wenn wir die aber die Frage, ob ein menschliches Leben ein „normales vollständiges Leben“ war, nicht nur als eine binäre Frage verstehen, sondern Grade der Vollständigkeit zulassen, liegt es nahe, diese Grade von der durchlaufenen Lebenszeit abhängig zu machen. Dann erscheint es zumindest plausibel, den Verlust, das was zu einem vollen Leben fehlt, wenigstens für aggregative „Makro“-Fragestellungen mit der verbleibenden Lebenserwartung zu bewerten. Dies ist (gerade im Kontext der übrigen Rawlsschen Theorie der Gerechtigkeit) kein Freibrief, das Leben eines alten Menschen als nicht mehr schützenswert zu betrachten, oder es gar aktiv zu beenden. Auf der anderen Seite heißt es sehr wohl, daß das Ziel der Lebensverlängerung von alten Menschen in der Gesundheitspolitik gegenüber etwa dem Ziel der Bekämpfung der Kindersterblichkeit (solange es um dieselbe Zahl der Betroffenen geht) ein geringeres Gewicht hat.

Die gerade vorgenommene Verteidigung der Verwendung von YLLs im Rahmen einer rawlsianischen Konzeption ist unvollständig, solange nicht auch geklärt ist, *wie* diese YLLs verwendet werden sollen. Wenn die intertemporale Summe der YLLs ein ungeeignetes Kriterium sind, bleiben immer noch die Periodensummen als möglicherweise ethisch relevante Größen. Diese Periodensummen berücksichtigen allerdings nicht die Verteilung des Verlustes von Lebensjahren auf verschiedene Gruppen bzw. hier Regionen, von der man erwarten würde, daß sie in einer rawlsianischen Konzeption eine Rolle spielt. Im Sinne der Rawlsschen Sorge um die am schlechtesten gestellte Gruppe schlage ich deshalb vorläufig vor, sowohl die globalen Periodensummen wie die regionalen YLLs pro tausend Einwohner zu berücksichtigen, und jedenfalls dann von einer eindeutigen Verbesserung (bezüglich des Ziels der Verhinderung eines vorzeitigen Todes) zu sprechen, wenn in allen Perioden sowohl die globale Summe der YLLs als auch die YLLs pro (tausend) Einwohner in der am schlechtesten gestellten Region

abnimmt.¹⁰⁸

Die Verwendung der Summen ist aus drei Gründen weniger problematisch, als dies generell im Utilitarismus (wenigstens aus rawlsianischer Sicht) der Fall ist:

1. Die Verringerung der Summe von verlorenen Lebensjahren in den jeweiligen Perioden erscheint deshalb als ein vernünftiges Ziel, weil es als unvermeidlich angenommen wird, daß eine substantielle Zahl von Menschen aufgrund klimatischer Gegebenheiten eines verfrühten Todes sterben. Wenn wir annehmen würden, daß es möglich wäre, die Bedingungen für gerechte Gesellschaften langfristig zu verwirklichen und gleichzeitig klimabedingte YLLs vollständig zu eliminieren, dann müßte angesichts des Gewichtes, das eine vollständiges Leben bei Rawls besitzt, einfach diese Möglichkeit verwirklicht werden. Die Entscheidung über verschiedene Klimastrategien ist in gewissem Sinn eine besondere Form von Triage-Entscheidung, die zwar eine große Zahl von Menschen betrifft, aber nicht unter dem Zeitdruck eines Schlachtfeldes, einer aktuellen Katastrophe. In einer solchen Triage-Situation hat die Maxime, die größte Zahl von Menschen zu retten, auch unabhängig vom Utilitarismus eine große Attraktivität und wird z.T. einfachhin als *das* Ziel der Triage verstanden. [Vgl. Schultze 1998, Dressler 1998 und Baker u. Strosberg 1992]
2. Mit der Vorordnung des Ziels des Erreichens der Bedingungen für wohlgeordnete Gesellschaften und in einer Völkergemeinschaft, die alle Gesellschaften dazu führen möchte, respektable Mitglieder der Völkergemeinschaft zu werden, die das Völkerrecht und die Menschenrechte respektieren und – wenn es sich nicht um Demokratien handelt – zumindest allen gesellschaftlichen Gruppen eine Form der Mitsprache in gesellschaftlichen Entscheidungen einräumen, wird eine willkürliche Mißachtung der Rechte der Einzelnen zumindest erschwert, oder wenn die ethische Zielvorstellung verwirklicht ist, sogar verhindert. Auch wenn für viele Menschen ein vorzeitiger Tod unvermeidlich sein sollte, erleiden sie diesen Tod nicht, weil ihre fundamentalen Rechte nicht geschützt würden. (Sondern weil es unvermeidlich ist, daß das geschieht.)
3. Die Verknüpfung des Summenkriteriums mit dem Kriterium der verlorenen Lebensjahre pro 1000 Einwohner für die am schlechtesten gestellte Region (allgemeiner: Gruppe), verhindert, daß den (sowieso) am schlechtesten Gestellten zusätzlich übermäßige Lasten auferlegt werden.

Ich habe zu begründen versucht, weshalb YLLs mit Altersgewichten und einem Diskontierungsfaktor von 3 % eigentlich nicht in den Rahmen einer rawlsianischen Theorie der Gerechtigkeit passen. Angesichts des sowieso sehr groben Charakters der Abschätzung der gesundheitlichen Folgen, und da auch diese YLLs ausreichen, um deutlich zu machen, was aus der von mir vorgeschlagenen Extension der Rawlsschen Theorie

¹⁰⁸ Auch dieses *Dominanzkriterium* ist noch unvollständig, reicht aber aus, um wenigstens für einige Fälle definitive Aussagen zu treffen. Ich nehme an, daß es wenig umstritten sein wird.

folgt, und allgemeiner, welche Bedeutung die explizite, gesonderte Beachtung der gesundheitlichen Folgen des Treibhauseffekts (einschließlich des durch ihn verursachten Sterbens) für die Beurteilung von verschiedenen Klimastrategien bedeutet, werde ich sie für das Folgende beibehalten.

Bevor wir allerdings zu einer Anwendung der eben vorgeschlagenen Kriterien auf die verschiedenen Umwelt- und Umverteilungsstrategien schreiten, ist eine weitere Schwierigkeit in den Blick zu nehmen: Der Klimaeffekt ist sicher nicht der einzige Grund, wegen dem Menschen krank werden und sterben. Wir haben schon auf den Zusammenhang von Gerechtigkeit und Gesundheit hingewiesen, und außerdem besteht natürlich auch ein Zusammenhang zwischen Armut und Krankheit bzw. vorzeitigem Sterben. Die Frage ist, ob nur die YLLs aufgrund der Klimaerwärmung berücksichtigt werden sollten, oder nicht vielmehr (im Prinzip) alle YLLs, die durch die Wahl der verschiedenen politischen Strategien beeinflusst werden, insbesondere die, die sich aus dem unterschiedlichen Reichtum der Regionen in den verschiedenen Strategien ergeben.

Wenn wir die Rawlssche (nicht universalistische) Grundkonzeption des Rechts der Völker im wesentlichen beibehalten wollen, und sie nur um eine internationale Regelung für das internationale Problem des Sterbens aufgrund des Treibhauseffekts ergänzen wollen, dann scheint aber die einzige konsequente Vorgehensweise die zu sein, nur die YLLs aufgrund des Treibhauseffekts zu berücksichtigen, wenigstens von dem Moment an, wo die (u.a. materiellen) Bedingungen einer wohlgeordneten Gesellschaft erreicht sind. Die Verantwortung der tatsächlichen Errichtung einer wohlgeordneten Gesellschaft liegt dann bei den einzelnen Völkern. In einer wohlgeordneten, gerechten Gesellschaft würden zwar noch Menschen sterben, aber nicht aufgrund von Ungerechtigkeit, nicht deshalb weil ihnen Unrecht geschieht. Wenn wir der Meinung wären, daß das von uns als Zielpunkt der Beistandspflicht gewählte Volkseinkommen nicht ausreicht, um Krankheiten und Sterben, die Folge von Ungerechtigkeit sind, auszuschließen, dann hätten wir es eben falsch bestimmt, und müßten ein höheres wählen. Mit Rawlsschen Vorstellung, daß auch relativ arme Gesellschaften gerechte Gesellschaften sein können, ist es aber in jedem Fall unvereinbar, die Minimierung der Summe aller YLLs (unabhängig von ihren kausalen Faktoren) zu einer Forderung der Gerechtigkeit zu erheben, da es zum einen plausibel erscheint, daß durch vermehrte Aufwendungen für Gesundheit YLLs immer weiter reduzierbar sind (das Ziel ist also unausschöpfbar), und zum anderen ein solches Ziel mit dem vernünftigen Pluralismus der Lebensentwürfe (die keineswegs alle eine Maximierung der Lebenszeit zum Ziel haben) unvereinbar erscheint.

Wir beschränken uns also auf die YLLs, die Resultat des Treibhauseffekts sind, und berechnen sie aus der von Nordhaus u. Boyer [2000c, 93] angegebenen (korrigierten¹⁰⁹) Beziehung zwischen regionaler Durchschnittstemperatur und prozentualem Verlust an Bruttosozialprodukt zurück, wobei wir annehmen, daß diese Gleichung korrekt den

¹⁰⁹ Bei der tatsächlich angegebenen Gleichung – $Q_{health,j}(T_j) = 0.002721(T_j)^{0.2243}$ – handelt es sich offensichtlich um einen Druckfehler. Korrekt muß es heißen: $Q_{health,j}(T_j) = 0.002721 \exp(0.2243T_j)$. Diese Gleichung ergibt sich in der Tat, wenn man mit den Daten von Nordhaus u. Boyer [2000c, 73, 91] die entsprechende semilogarithmische Beziehung schätzt.

entsprechenden Verlust schätzt, wenn ein verlorenes Lebensjahr den Wert von zwei Jahreseinkommen hat. [Vgl. Nordhaus u. Boyer 2000c, 82] Vom Ergebnis ziehen wir zuletzt die Zahl der klimabezogenen YLLs ab, die sich aufgrund der geographischen Gegebenheiten, unabhängig vom Treibhauseffekt bei vorindustriellen Temperaturen ergeben.

Abbildung 7.3 zeigt für die bislang betrachteten Strategien (ohne Geoengineering) die verlorenen Lebensjahre (in Tausend). Es zeigt sich, daß die strikteren Umweltstrategien (im Modell) vor allem in späteren Perioden in einem erheblichen Umfang Leben retten. Die erfolgreichste Strategie ist die Beschränkung des Temperaturanstiegs auf 1,5 °C, die schlechteste der Referenzfall. Von den weniger radikalen Strategien ist die beste die Optimalstrategie, oder die verschobene Optimalstrategie. Die großzügige Umverteilungsstrategie ist mit höheren Temperaturen verbunden als die Optimalstrategie ohne Umverteilung, und führt deshalb auch zu mehr YLLs. Im Jahr 2105 etwa würden im Referenzfall 15 520 Tausend Lebensjahre verlorengehen, mit der Optimalstrategie 14 742, mit der verschobenen Optimalstrategie 14 764, mit der Umverteilungsstrategie 15 049 und mit der Beschränkung des Temperaturanstiegs auf 1,5 °C nur 7 778,7 Tausend.¹¹⁰ Gegenüber dem Referenzfall könnten also mit der letzteren Strategie 7 741,3 Tausend Lebensjahre gerettet werden. Mit höheren Temperaturen werden in späteren Jahren die Unterschiede noch deutlicher: Im Jahr 2305 beträgt der Unterschied zwischen der Optimalstrategie und der 1,5 °C Strategie schon 31 481 Tausend Lebensjahre¹¹¹, gegenüber der Referenzstrategie können sogar 33 839 Tausend Lebensjahre gerettet werden.¹¹²

Entsprechend der von mir vorgeschlagenen Ergänzung der Rawlsschen Theorie des Rechts der Völker haben diese Informationen aber keinen Einfluß auf die Entscheidung zwischen den bisher betrachteten Strategien. Nach wie vor ist demnach die beste Strategie die der Umverteilung mit „optimalen“ CO₂ Steuern, die zu vergleichsweise hohen YLLs in späteren Jahren führt. Dieses Ergebnis resultiert aus der hohen Bedeutung, die der Erreichung der Bedingungen eine wohlgeordneten Gesellschaft zugerechnet wird. Bei der Beurteilung dieses Ergebnisses ist auch zu beachten, daß wir argumentiert haben, daß in einer wohlgeordneten Gesellschaft die Gesundheit der Bürger gefördert wird und deshalb die Errichtung einer solchen Gesellschaft als solches ein Beitrag zur Verringerung vorzeitigen Sterbens ist. Außerdem können wir erwarten, daß ein höherer

¹¹⁰ Zum Vergleich: Nachdem Indonesien die Zahl der aufgrund des Tsunamis vom 26.12.04 Vermißten nach unten revidiert hatte, waren am 8.4.05 217 000 Menschen als tot oder vermißt gemeldet [AFP 2005]. Wenn wir annehmen, daß das der Zahl der aufgrund des Tsunamis Verstorbenen entspricht, und die Altersverteilung der Opfer ungefähr mit der der Gesamtbevölkerung Indonesiens übereinstimmt (163 978 Tote und Vermißte), dann sind durch den Tsunami 5 526,5 Tausend Lebensjahre verloren gegangen. Wenn wir annehmen, daß (wie verschiedentlich gemeldet) Kinder unter den Toten überproportional vertreten waren, läge die Zahl der durch den Tsunami verschuldeten YLLs noch unter dieser Schätzung.

¹¹¹ Das entspräche unter den Annahmen der vorangehenden Anmerkung dem 5,4fachen des Verlusts durch den Tsunami vom Dezember 2004 in Indonesien.

¹¹² Abgesehen von minimalen Unterschieden in der ersten Periode ordnet der zweite Teil des vorgeschlagenen Kriteriums – YLLs pro tausend Einwohner der am schlechtesten gestellten Region – die verschiedenen Strategien gleich, weshalb ich nicht weiter darauf eingehe.

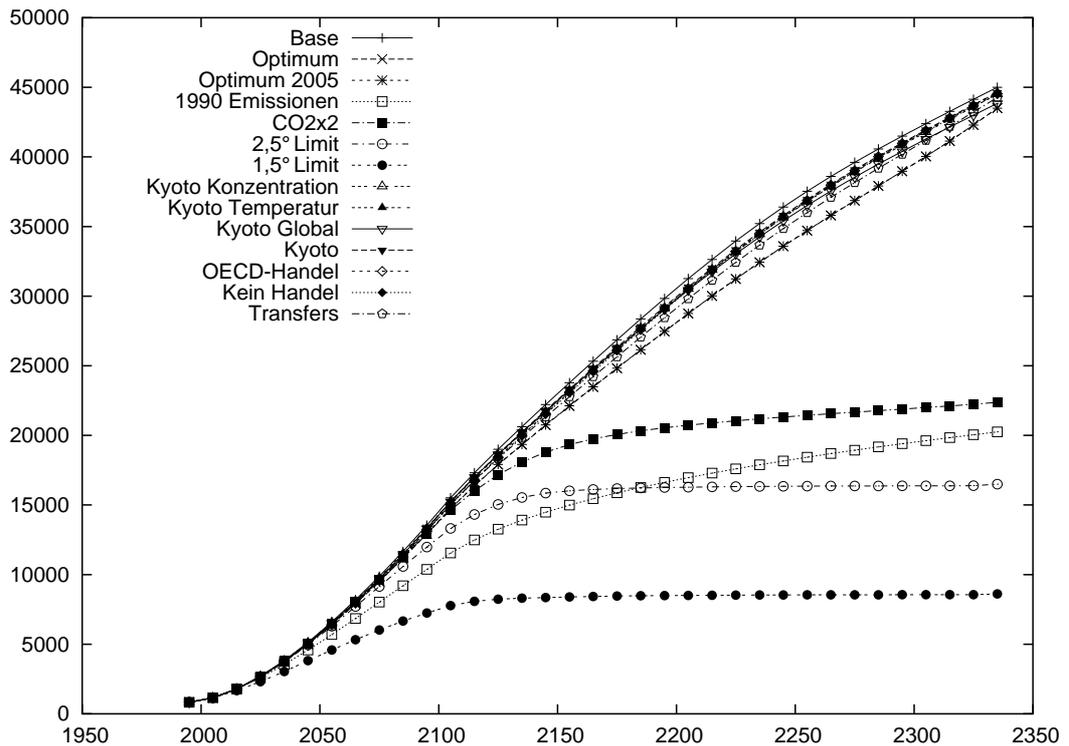


Abbildung 7.3: Dem Treibhauseffekt zurechenbare YLLs in Tausend für vierzehn Strategien. Eigene Berechnung.

Reichtum für arme Gesellschaften sehr effektiv zur Bekämpfung von Krankheit und vorzeitigen Sterben beiträgt.

Den letzteren Punkt möchte ich noch illustrieren, indem ich mit einer ähnlichen Methode, wie sie Nordhaus u. Boyer [2000c] einsetzen, die Zahl der durch Armut bedingten YLLs bestimme und dann die Gesamtzahl der YLLs aufgrund des Klimaeffekts und aufgrund von Armut in den Blick nehme.

Für die Berechnung der YLLs aufgrund von Armut gehen wir also so ähnlich wie Nordhaus vor, beschränken uns aber auf nur eine der Methoden, die er in der Form eines Durchschnitts in die Bestimmung der YLLs aufgrund des Treibhauseffekts eingehen lässt, und schätzen direkt eine Gleichung für die YLLs und nicht für den Verlust des Bruttosozialprodukts aufgrund der YLLs. Genauer nehmen wir an, daß die YLLs, die der Unterernährung und schlechtem Wasser, unzureichenden Sanitäreinrichtungen und unzureichender Hygiene zuzurechnen sind [Murray u. Lopez 1996, 312], armutsbezogen sind, und schätzen eine halblogarithmische Gleichung, die den Logarithmus der YLLs pro Kopf in Beziehung zum Pro-Kopf-Einkommen setzt.¹¹³

¹¹³ Ich lasse dabei die Schadensfunktionen unverändert. Abgesehen von dem sowieso nur überschlagsmäßigen Charakter der Berechnungen entspricht das der Annahme, daß der Teil der Schadensfunktion,

Abbildung 7.4 zeigt für die vierzehn bisher betrachteten Strategien (ohne Geoengineering) für jede Periode die Summe „aller“ (d. h. der klimabezogenen und der armutsbezogenen) YLLs und Abbildung 7.5 jeweils die – über die acht Regionen – maximalen YLLs (klima- und armutsbezogen) pro Tausend Einwohner. Beide Abbildungen zeigen, wie aufgrund des höheren Wohlstands der armen Regionen bei großzügigen Transfers in den frühen Perioden YLLs in einem erheblichen Ausmaß reduziert werden kann, ohne daß der Effekt der Errichtung wohlgeordneter Gesellschaften, der, wie wir argumentiert haben, als solcher einen erheblichen Beitrag zur Gesundheit der Bürger leisten würde, eingerechnet wurde.

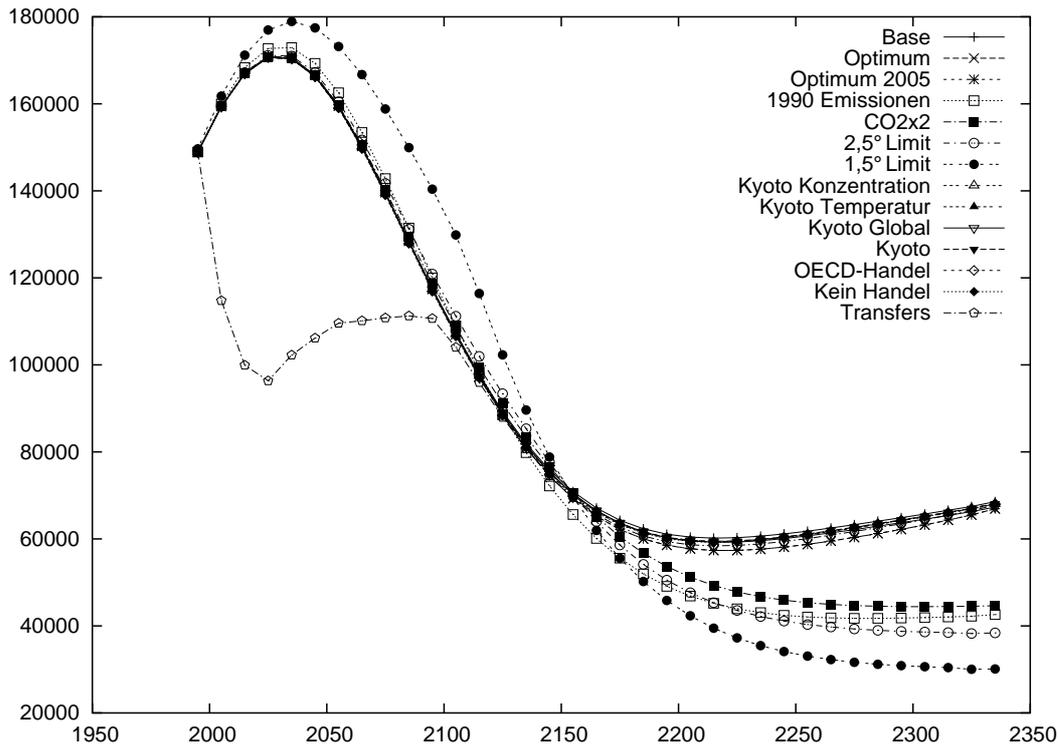


Abbildung 7.4: Summe aller YLLs für vierzehn Strategien. Eigene Berechnung.

Es könnte nun scheinen, als sei die Hinzufügung eines weiteren Grundsatzes des Völkerrechts verbunden mit dem Vorrang der Erreichung der Bedingungen einer wohlgeordneten Gesellschaft ein sinnloses Unterfangen, da auf diese Weise der neue Grundsatz praktisch nie zum Tragen komme. Das ist aber nicht der Fall. Wenn die Bedingungen einer wohlgeordneten Gesellschaft einmal erreicht sind, kommt die die Beistandspflicht zu ihrem Ende, und der neue Grundsatz entfaltet seine volle (ethische) Kraft, solange seine Erfüllung das Ziel der Beistandspflicht nicht gefährdet.

Zur Illustration erwähne ich noch kurz weit radikalere Umverteilungsstrategien, die

der den YLLs aufgrund des Klimaeffekts entspricht, korrekt den direkt wirtschaftlichen Schaden des frühzeitigen Sterbens modelliert.

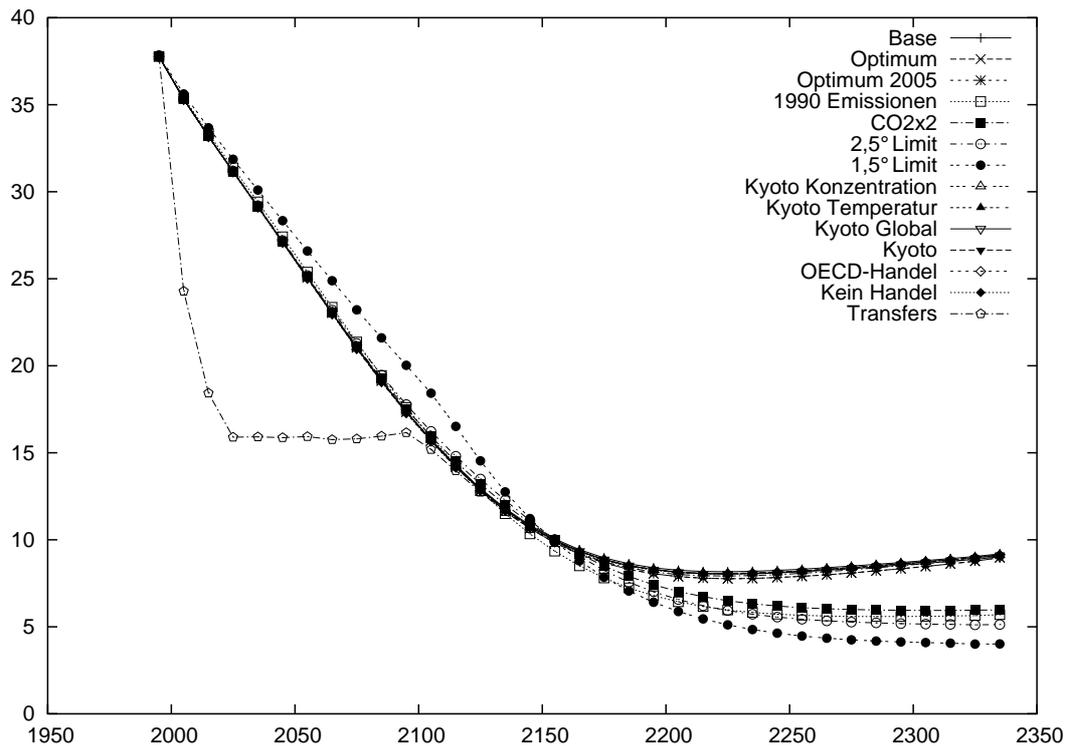


Abbildung 7.5: Maximum der YLLs pro Tausend für vierzehn Strategien. Eigene Berechnung.

dazu führen, daß schon von 2005 an für alle Regionen die Bedingungen einer wohlgeordneten Gesellschaft erfüllt sind. Dabei nehme ich an, daß dies der früheste Zeitpunkt ist, an dem das Ziel überhaupt noch erreicht werden kann.¹¹⁴ Um diesen früheren Termin der Zielerreichung zu verwirklichen, ist es notwendig, von der – wie ich argumentiert habe – ethisch willkürlichen¹¹⁵ Begrenzung der Transfers auf einen Anteil des Zuwachses der Produktion abzugehen. Für Kombinationen solcher Umverteilungsstrategien und ehrgeiziger Klimaziele gilt also, daß ehrgeizigere Klimaziele (in der Form etwa von Temperaturlimits) aufgrund des zusätzlichen Grundsatzes solange als ethisch besser eingeordnet werden, wenn gleichzeitig (etwa durch entsprechend angepaßte Transfers) die materiellen Bedingungen eine wohlgeordneten Gesellschaft erhalten bleiben.

Aber selbst wenn man in der Perspektive einer unvollständigen Konformität verbleibt, kann der zusätzliche Grundsatz seine Wirkung entfalten, wenn man berücksichtigt, daß die (reale) Bereitschaft, für benachteiligte Gesellschaft, und die (reale) Be-

¹¹⁴ In einer etwas lockeren Weise könnte man argumentieren, daß dies der früheste Zeitpunkt ist, da wir schon dieses Jahr erreicht haben. Im Prinzip wäre es aber problemlos möglich, eine Zielerreichung schon 1995 zu modellieren, ohne daß sich an den *qualitativen* Ergebnissen etwas ändern würde.

¹¹⁵ S. o. S. 228 u. 230.

reitschaft, für zukünftige Generationen und eine gesunde Umwelt Einschränkungen in Kauf zu nehmen, eine gewisse Unabhängigkeit besitzen: Wenn die erstere an ihrer Grenzen gelangt ist, würde in dieser Perspektive der neue Grundsatz in Kraft treten.

Die hier angestellten Überlegungen zur Behandlung verlorener Lebensjahre in einer rawlsianischen Perspektive sollten nicht überbewertet werden: Die Berechnung der aufgrund des Treibhauseffekts verlorenen Lebensjahre ist sicher alles andere als eine verlässliche Prognose und das vorgeschlagene Kriterium ist nicht vollständig, nicht systematisch begründet und sicher diskussionsbedürftig. Soviel scheint man allerdings meines Erachtens als Resümee festhalten zu können:

1. Die Beurteilung verschiedener Strategien zum Umgang mit dem Treibhauseffekt wird wesentlich davon abhängen, welche Auswirkungen auf die menschliche Gesundheit, insbesondere auch auf das vorzeitige Sterben von Menschen zu erwarten sind, und wie diese ethisch zu bewerten sind.
2. Angesichts der Bedeutung, die ein vollständiges menschliches Leben – als eine notwendige Vorbedingung der Verwirklichung fast eines jeden vernünftigen Lebensplans – bei Rawls hat, ist es mit einer rawlsianischen Perspektive wohl unvereinbar, verlorene Lebensjahre mit anderen Grundgütern (insbesondere Einkommen und Wohlstand) zu verrechnen. Die weitgehende Sicherung der Möglichkeit für jeden Menschen, ein vollständiges Leben in einer wohlgeordneten Gesellschaft zu leben, ist der fortlaufenden Maximierung des Wohlstands vorzuordnen.
3. Die Maximierung der Lebenslänge oder die „Maximierung“ der Gesundheit während des ganzen Lebens ist nicht Teil eines jeden vernünftigen Lebensentwurfes. Deshalb wird in einer rawlsianischen Perspektive der gesellschaftliche Aufwand (national und international) für die Gesundheit und die Lebensverlängerung in einer sinnvollen Weise zu begrenzen sein.

7.3.8.2 Unsicherheit und Katastrophen

Ein wesentliches Element der aktuellen Diskussion über die Auswirkungen des Treibhauseffekts ist der hohe Grad der Unsicherheit, der mit vielen Aussagen und Projektionen verbunden ist. Auch der rohe und unzuverlässige Charakter der Schätzung verlorener Lebensjahre bei Nordhaus u. Boyer [2000c] und im vorhergehenden Unterabschnitt kann als Reflex dieser Unsicherheit verstanden werden. Es handelt sich gewissermaßen um den verzweifelte Versuch, trotz und angesichts unseres mangelnden Wissens, trotz der „enormous uncertainties about underlying physical and biological impacts and about the potential for adaption“ [Nordhaus u. Boyer 2000c, 71] zu Zahlen zu kommen, mit denen wir dann rechnen können. Auch in der Sicht von Nordhaus und Boyer handelt es sich dabei um einen Ersatz „systematischer Schätzungen“ der Auswirkungen auf die Gesundheit. [Vgl. Nordhaus u. Boyer 2000c, 80]

In ähnlicher oder noch radikalerer Weise sind auch Aussagen über mögliche katastrophische Klimaentwicklungen mit Unsicherheiten verbunden. Hierbei geht es um die Sorge, daß es aufgrund des globalen Temperaturanstiegs zu schnellen, extremen, nicht

reversiblen, möglicherweise unvorhersehbaren und in ihren lokalen, regionalen oder globalen Auswirkungen katastrophalen Klimaänderungen kommt. (Vgl. z. B. Nordhaus u. Boyer [2000c, 69, 87 f.], IPCC [2001, 945-952] und Cline [2004, 8 f.].)

Wenn man diese Sorge ernst zu nehmen hat, wie u.a. die im letzten Absatz zitierte Literatur nahe legt, dann stellt sich die Frage, wie man mit der Möglichkeit schwer vorhersagbarer sehr großer Katastrophen umgehen soll. Als erste Antwort auf diese Frage ist wohl darauf hinzuweisen, daß es selbst aus der Sicht von Autoren, die ansonsten einem Kosten-Nutzen Ansatz verbunden sind, keine etablierte Antwort auf diese Frage gibt. (Nordhaus u. Boyer [2000c, 71], IPCC [2001, 946])

Nordhaus u. Boyer [2000c, 87-89, 93 f.] integrieren mögliche katastrophische Schäden in ihren Zahlungsbereitschaftsansatz.¹¹⁶ In einer früheren Studie [Nordhaus 1994] hat Nordhaus 15 Experten u.a. nach ihrer subjektiven Einschätzung der Wahrscheinlichkeit für das Eintreten eines großen Schadenfalls, bei dem 25 % des Weltbruttosozialprodukts vernichtet würden, unter verschiedenen Szenarien befragt. Für das Szenario eines Temperaturanstiegs von 3 °C bis 2090 war der Durchschnitt der Schätzungen 0,6 %, für das eines Temperaturanstiegs von 6 °C bis 2175 3,4 %. Um die „wachsende Besorgnis“ wegen der Möglichkeit katastrophischer Entwicklungen zu berücksichtigen, verdoppeln die Autoren dann diese beiden Prozentsätze und beziehen sie auf einen (höheren) Verlust von 30 % des Weltbruttosozialprodukts, den sie dann entsprechend ihrer subjektiven Einschätzung der regionalen Gefährdung mit einem relativ groben Schema (es gibt nur die drei Stufen von 22,1 %, 33,2 % und 44,2 %) auf die zwölf Weltregionen der früheren Version des RICE-Modells verteilen. Unter der Annahme einer Rate der relativen Risikoaversion von 4 berechnen sie dann für jede Region und die beiden Szenarien einer Temperaturerhöhung von 3 °C und 6 °C die jeweilige Zahlungsbereitschaft für die Vermeidung des Katastrophenfalls. Auf dieser Basis bestimmen sie zuletzt für jede Region die Schadensfunktion aufgrund des katastrophischen Risikos, von der sie annehmen, daß sie zwischen 0 °C und 3 °C linear verläuft, um dann die Form einer Potenzfunktion anzunehmen.

Für die meisten Regionen ist der auf diese Weise für die globalen Temperaturanstiege von 2,5 °C und 6 °C geschätzte Schaden aufgrund des Risikos eines katastrophischen Ereignisses die bei weitem wichtigste Komponente der gesamten Schadenssumme und dominiert deshalb auch die geschätzte quadratische Gesamtschadensfunktion. [Nordhaus u. Boyer 2000c, 98]

Wie schon für die gesundheitlichen Auswirkungen des Treibhauseffekts ist es auch für das Risiko katastrophischer Entwicklungen mehr als fraglich, ob die von Nordhaus u. Boyer [2000c] gewählte Vorgehensweise in den Rahmen einer rawlsianischen ethischen Konzeption paßt. Noch weniger als bei der Frage der Gesundheit, gibt es allerdings bei Rawls *direkte* Hinweise darauf, wie auf dieses Risiko national oder international reagiert werden sollte.¹¹⁷ Wie in der Frage der gesundheitlichen Auswirkungen wird es also im Folgenden um einen von Rawls inspirierten vorläufigen Vorschlag einer

¹¹⁶ Cline [2004, 9] nennt ihre Vorgehensweise „heroisch“.

¹¹⁷ Abgesehen von einigen kursorischen Bemerkungen hat sich Rawls an keiner Stelle seines Werkes mit dieser Frage direkt auseinandergesetzt. Zu dieser Lücke in Rawls' Theorie und Vorschlägen, die Theorie zu ergänzen vgl. u. a. Manning [1981], Bell [2003] und Felipe [2005].

Ergänzung der Rawlsschen Theorie der Gerechtigkeit und des Völkerrechts gehen. Da wir auch hier an der Rawlsschen Trennung zwischen Fragen der internen Gerechtigkeit und des Völkerrechts festhalten, und es auch weiterhin um eine Frage der internationalen Kooperation angesichts der Gefahren des Treibhauseffekts geht, suchen wir genauer um eine Ergänzung des Rechts der Völker.

Wenn wir genauer fragen, weshalb aus einer rawlsianischen Perspektive die Vorgehensweise in Nordhaus u. Boyer [2000c] unbefriedigend ist, so ist wohl vor allem auf die Kombination von zwei Aspekten hinzuweisen:

1. Nordhaus betrachtet in seiner Behandlung des katastrophischen Risikos schon im Ansatz nur den möglichen (großen) strikt ökonomischen Verlust in einer Periode. Wenn wir aber von katastrophischen Risiken sprechen, geht es von vornherein um Entwicklungen, die nicht nur die wirtschaftliche Produktion, sondern alle Schadenskategorien direkt betreffen können, einschließlich der Destabilisierung sozialer, wirtschaftlicher und politischer Systeme. (Vgl. etwa IPCC [2001, 946, 949 f.])
2. Aus mehreren Gründen scheint die Anwendung der Methode des Erwartungsnutzen auf diese Fragestellung aus einer rawlsianischen Perspektive als fragwürdig: Wenigstens zum gegenwärtigen Zeitpunkt ist es nicht möglich, für die verschiedenen Kategorien möglicher katastrophischer Entwicklungen [IPCC 2001, 946 f.] *verlässliche* deterministische oder stochastische Modelle vorzulegen. Das gilt insbesondere auch für nicht ausschließbare „non-implausible ‚nightmare‘ or ‚doomsday‘ scenarios [which] could result from the speculative but consistent concatenation of individually possible causal relationships“. [IPCC 2001, 947] Mit seinen Überlegungen zur rationalen Reue (s. o. Abschnitt 6.2) macht Rawls deutlich, daß wenigstens *dann* eine ex post Perspektive sinnvoll ist, wenn eine unerträgliche, inakzeptable Situation droht. Dies ist z. B. im Fall eines Albraumszenarios plausiblerweise anzunehmen. Zusammengenommen sind damit schon zwei der Bedingungen, die nach Rawls die Anwendung einer Maximinregel der Entscheidung begründen (s. o. S. 142), erfüllt. Wenn die Völkergemeinschaft, wie wir im Rahmen der idealen Theorie erwarten dürfen, ihre Beistandspflicht erfüllt hat, und damit für alle die Völker die Bedingungen einer wohlgeordneten Gesellschaft gegeben sind, dann kann man auch die dritte Bedingung der Anwendung eines Maximinregel in gewisser Weise als erfüllt ansehen: Für die Völker ist zumindest ein „recht befriedigendes“ Niveau [Rawls 2003, 157 f.] der Versorgung mit für die Realisierung einer wohlgeordneten Gesellschaft notwendigen „Gütern“ (im weiten Sinn) zunächst einmal *gegeben*.

Eine Schwierigkeit ergibt sich noch daraus, daß es in der Rawlsschen Formulierung Voraussetzung für die Anwendung der Maximinregel ist, daß das befriedigende Niveau *garantiert* werden kann. Eine solche Garantie ist aber wenigstens beim gegebenen Stand der Klimaforschung nicht möglich, wenn auch angenommen wird, daß ein niedrigerer Temperaturanstieg tendenziell mit einem geringeren Risiko katastrophischer Entwicklungen verbunden ist:

The vague evidence provided by the present state of research supports the notion that even relatively small changes in mean climate could lead to large changes in the occurrence of stochastic extreme events. Furthermore, it suggests that large-scale discontinuities are unlikely below a 2 °C warming but relatively plausible for a sustained warming of 8-10 °C. The relatively small set of investigations discussed above lead to the conclusion that a warming range of 4-5 °C seems to represent a critical disturbance regime where macro-discontinuities may start to emerge. This temperature threshold appears to be sensitive to the rate of change at which this level is reached. [IPCC 2001, 952]

Eine plausible Extension der Rawlsschen Überlegungen zur Maximinregel für Situationen, in denen das Risiko des Verlustes eines befriedigenden Zustandes nicht vollständig eliminierbar ist, wäre es, dieses Risiko wenigstens möglichst klein zu halten. Eine solche Extension legt sich auch deshalb nahe, weil man ja wohl überhaupt (auch für den Fall der Wahl von Grundsätzen der Gerechtigkeit für die Grundstruktur einer Gesellschaft) einen vollständigen Ausschluß des Risikos des Verlustes eines befriedigenden Zustandes in der Realität nie erreichen können. Auch für den Kernbereich der Rawlsschen Theorie der Gerechtigkeit ist eine solche Extension also notwendig.

Wenn man die vorgestellten Überlegungen zur internationalen Umverteilung im Rahmen eine Beistandspflicht, zur Gefährdung von Leben und Gesundheit aufgrund des Treibhauseffekts und zur Gefahr katastrophischer Entwicklungen kombiniert, ergibt sich also als Konsequenz ein in rawlsianischer Sichtweise plausibles Argument dafür,

1. mit Priorität im Rahmen einer internationalen Beistandspflicht darauf hinzuwirken, daß alle Völker möglichst schnell die kulturellen, politischen und materiellen Bedingungen für eine wohlgeordnete Gesellschaft erreichen, was u.a. erhebliche Transfers erfordern kann;
2. den Weg dahin und die weitere Entwicklung aller Gesellschaften, die das für eine wohlgeordnete Gesellschaft notwendige Niveau materiellen Wohlstands erreicht haben, so zu gestalten, daß langfristig katastrophale Entwicklung möglichst unwahrscheinlich werden, d. h. der Temperaturanstieg aufgrund des Treibhauseffekts minimiert wird.

Unter den betrachteten Strategien des RICE-Modells wäre dies ein Argument für eine Strategie, die ein hohes Maß an Umverteilung mit einer Beschränkung des Temperaturanstiegs in jedem Fall unter 2 °C oder gar 1,5 °C kombiniert.

Literaturverzeichnis

AFP 2005

AFP: *Death toll in Asian tsunami disaster revised down to 217,000.* Version: April 2005. <http://newsunleashed.com/death-toll-in-asian-tsunami-disaster-revised-down-to-217000.html>.
– Online-Ressource, Abruf: 20.11.05

Ahlheim u. Rose 1992

AHLHEIM, M. ; ROSE, M.: *Messung individueller Wohlfahrt*. Berlin, 1992

Allais 1989

ALLAIS, M.: *La théorie générale des surplus*. Grenoble, 1989

Arrow 1977a

ARROW, K. J.: Einige ordinal-utilitaristische Bemerkungen über Rawls' Theorie der Gerechtigkeit. In: *Höffe, O. (Hrsg.), Über John Rawls Theorie der Gerechtigkeit*. Frankfurt, 1977, S. 199–223

Arrow 1977b

ARROW, K. J.: Extended Sympathy and the Possibility of Social Choice. In: *American Economic Review* 67 (1977), S. 219–225

Arrow u. Scitovsky 1969

ARROW, K. J. (Hrsg.) ; SCITOVSKY, T. (Hrsg.): *Readings in Welfare Economics*. London, 1969

Atkinson u. Stiglitz 1980

ATKINSON, A. B. ; STIGLITZ, J. E.: *Lectures on Public Economics*. Singapore, 1980

Baker u. Strosberg 1992

BAKER, R. ; STROSBERG, M.: Triage and equality: an historical reassessment of utilitarian analyses of triage. In: *Kennedy Institute of Ethics Journal* 2 (1992), Juni, Nr. 2, S. 103–123

Beitz 2002

BEITZ, Ch. R.: *Political theory and international relations*. Princeton, 2002

Bell 2003

BELL, D.: *Political Liberalism and Ecological Justice*. Version: September 2003. <http://www.essex.ac.uk/ECPR/events/generalconference/marburg/papers/8/2/Bell.pdf>. – Paper presented at ECPR General Conference 2003

Belli 2001

BELLI, P.: *Ten Years of Health Reform in Former Socialist Economies: Lessons Learned And Options for the Future*. Version: Mai 2001. <http://www.hsph.harvard.edu/hcpds/wpweb/Belliwp1106.pdf>. – HCPDS Working Paper Series

Belli u. a. 2003

BELLI, P. ; HORVATHOVA, Z. ; MALY, I. ; GAAL, P. ; KROL, Z. ; BERMAN, P. ; BOSSERT, Th.: *Formal and Informal Household Spending on Health*. Version: Juli 2003. <http://www.hsph.harvard.edu/ihs/publications/pdf/lac/MerckReportFINAL.pdf>. – Harvard School of Public Health, International Health Systems Program Report

Bhagwati 1969

BHAGWATI, J. (Hrsg.): *Internation Trade*. Harmondsworth, 1969

Binmore 1994

BINMORE, K. G.: *Playing Fair*. Cambridge, 1994

Binmore 1998

BINMORE, K. G.: *Just Playing*. Cambridge, 1998

Blackorby u. Donaldson 1985

BLACKORBY, C. ; DONALDSON, D.: Consumers' Surplus and Consistent Cost-Benefit Tests. In: *Social Choice and Welfare* (1985), S. 251–262

Blackorby u. Donaldson 1988

BLACKORBY, C. ; DONALDSON, D.: Money Metric Utility: A Harmless Normalization? In: *Journal of Economic Theory* 46 (1988), S. 120–129

Bormann u. Schröder 2004

BORMANN, F.-J. (Hrsg.) ; SCHRÖDER, Ch. (Hrsg.): *Abwägende Vernunft*. Berlin, 2004

Brieskorn u. Wallacher 1998

BRIESKORN, N. (Hrsg.) ; WALLACHER, J. (Hrsg.): *Homo oeconomicus: Der Mensch der Zukunft?* Stuttgart, 1998

Böventer u. Illing 1995

BÖVENTER, E. v. ; ILLING, G.: *Einführung in die Mikrotheorie*. München, 1995

Caplan 1999

CAPLAN, B.: The Austrian Search for Realistic Foundations. In: *Southern Economic Journal* 65 (1999), Nr. 4, S. 823–838

Chen u. Ravallion 2004

CHEN, S. ; RAVALLION, M.: *Global Poverty Monitoring Database*. Version: Mai 2004. <http://www.worldbank.org/research/povmonitor/index.htm>. – Online-Ressource, Abruf: 27.11.04

Chipman 1987

CHIPMAN, J. S.: Compensation Principle. In: *The New Palgrave* Bd. 1. London, 1987, S. 524–531

Chipman u. Moore 1978

CHIPMAN, J. S. ; MOORE, J. C.: The New Welfare Economics 1939-1974. In: *International Economic Review* 19 (1978), S. 547–584

Cline 2004

CLINE, W. R.: *Meeting the Challenge Of Global Warming*. Version: März 2004. http://www.cgdev.org/docs/Climate_Change_300404.pdf. – Copenhagen Consensus Challenge Paper

Coughlin 2002

COUGHLIN, C. C.: The controversy over free trade: the gap between economists and the general public. In: *The Federal Reserve Bank of St. Louis Review* 84 (2002), Nr. 1, S. 1–22

Daniels 1985

DANIELS, N.: *Just Health Care*. Cambridge, 1985

Daniels 2001

DANIELS, N.: Justice, Health, and Health Care. In: *American Journal of Bioethics* 1 (2001), Nr. 2, S. 3–15

Daniels u. a. 1999

DANIELS, N. ; KENNEDY, B. ; KAWACHI, I.: Why Justice is Good for Our Health. In: *Daedalus* 128 (1999), Nr. 4, S. 215–51

Deschamps u. Gevers 1978

DESCHAMPS, R. ; GEVERS, L.: Leximin and Utilitarian Rules: A Joint Characterization. In: *Journal of Economic Theory* 17 (1978), S. 143–163

Deutsche Gesellschaft für die Vereinten Nationen e.V. 1998

DEUTSCHE GESELLSCHAFT FÜR DIE VEREINTEN NATIONEN E.V.: *Bericht über die menschliche Entwicklung 1998*. Bonn, 1998

Deutsche Gesellschaft für die Vereinten Nationen e.V. 2001

DEUTSCHE GESELLSCHAFT FÜR DIE VEREINTEN NATIONEN E.V.: *Bericht über die menschliche Entwicklung 2001*. Bonn, 2001

Diamond u. Mirrlees 1971

DIAMOND, P. A. ; MIRRLEES, J. A.: Optimal Taxation and Public Production. In: *American Economic Review* 61 (1971), S. 8–27, 261–278

Dressler 1998

DRESSLER, S.: Triage. In: *Historisches Wörterbuch der Philosophie* Bd. 10. 1998, S. 1478 f.

Elster 1987

ELSTER, J. (Hrsg.): *The Multiple Self*. Cambridge, 1987

Elster u. Roemer 1991

ELSTER, J. (Hrsg.) ; ROEMER, J. E. (Hrsg.): *Interpersonal Comparisons of Well-Being*. Cambridge, 1991

Enquete-Kommission „Schutz der Erdatmosphäre“ des Deutschen Bundestages 1995

ENQUETE-KOMMISSION „SCHUTZ DER ERDATMOSPHÄRE“ DES DEUTSCHEN BUNDESTAGES (Hrsg.): *Mehr Zukunft für die Erde: Nachhaltige Energiepolitik für dauerhaften Klimaschutz*. Bonn, 1995

Feldman 1987

FELDMAN, A. M.: Welfare Economics. In: *The New Palgrave* Bd. 4. London, 1987, S. 889–895

Felipe 2005

FELIPE, S. T.: Rawls' Legacy: A Limited Possibility of a Non-speciesist Environmental Justice. In: *ethic@* 4 (2005), Juni, Nr. 1, S. 23–37

Freedom House 2001

FREEDOM HOUSE: *Civil Liberties Check List*. Version: Juni 2001. <http://www.freedomhouse.org/research/freeworld/2001/methodology3.htm>. – Online-Ressource, Abruf: 9.5.05

Freedom House 2004

FREEDOM HOUSE: *Freedom in the World Country Ratings: 1972 through 2003*. Version: Januar 2004. <http://www.freedomhouse.org/ratings/allscore04.xls>. – Online-Ressource, Abruf: 27.11.04

Fuller u. Geide-Stevenson 2003

FULLER, D. ; GEIDE-STEVENSON, D.: Consensus Among Economists: Revisited. In: *Journal of Economic Education* 34 (2003), Nr. 4, S. 369–387

Gaechter u. Riedl 2002

GAECHTER, S. ; RIEDL, A: Moral Property Rights in Bargaining / CESifo. 2002 (697). – Forschungsbericht

Gorman 1953

GORMAN, W. M.: Community Preference Fields. In: *Econometrica* 21 (1953), S. 63–80

Gorman 1955

GORMAN, W. M.: The Intransitivity of Certain Criteria Used in Welfare Economics. In: *Oxford Economic Papers* 7 (1955), S. 25–35

Gorman 1961

GORMAN, W. M.: On a Class of Preference Fields. In: *Metroeconomica* 13 (1961), S. 53–56

Habermas 1995

HABERMAS, J.: Reconiliation Through the Public use of Reason: Remarks on John Rawls's Political Liberalism. In: *The Journal of Philosophy* 92 (1995), März, Nr. 3, S. 109–131

Harsanyi 1976a

HARSANYI, J. C.: Can the Maximin Principle Serve as a Basis for Morality? In: *Harsanyi, J. C., Essays on Ethics, Social Behaviour and Scientific Explanation*. Dordrecht, 1976, S. 37–63

Harsanyi 1976b

HARSANYI, J. C.: *Essays on Ethics, Social Behavior, and Scientific Explanation*. Dordrecht, 1976

Harsanyi 1977a

HARSANYI, J. C.: Morality and the Theory of Rational Behaviour. In: *Sen, A.; Williams, B. (Hrsg.), Utilitarianism and Beyond*. Cambridge, 1977, S. 39–62

Harsanyi 1977b

HARSANYI, J. C.: *Rational Behavior and bargaining equilibrium in games and social situations*. Cambridge, 1977

Harsanyi 1978

HARSANYI, J. C.: Bayesian Decision Theory and Utilitarian Ethics. In: *American Economic Review* 68 (1978), Nr. 2, S. 223–228

Hausman 1992

HAUSMAN, D. M.: *The inexact and separate science of economics*. Cambridge, 1992

Hertel u. Reimer 2004

HERTEL, Th. W. ; REIMER, J. J.: *Predicting the Poverty Impacts of Trade Reform*. Version: November 2004. http://wdsbeta.worldbank.org/external/default/WDSContentServer/IW3P/IB/2004/12/17/000160016_20041217172335/Rendered/PDF/wps3444.pdf. – World Bank Policy Research Working Paper

Hicks 1939

HICKS, J. R.: The Foundations of Welfare Economics. In: *Economic Journal* 69 (1939), S. 696–712

Hicks 1981

HICKS, J. R.: *Wealth and Welfare*. Oxford, 1981

Höffe 1977

HÖFFE, O.: *Über John Rawls Theorie der Gerechtigkeit*. Frankfurt, 1977

IDEA 2004

IDEA, International: *Funding of Political Parties and Election Campaigns*. Stockholm, 2004

ILO 2003

ILO: *Yearbook of Labor Statistics*. Genf, 2003

IPCC 2001

IPCC: *Climate Change 2001: Impacts, Adaptation and Vulnerability*. Cambridge, 2001

Jaeger 1981

JAEGER, K.: Gleichgewicht, ökonomisches. In: *Handwörterbuch der Wirtschaftswissenschaft* Bd. 3. Stuttgart, 1981, S. 671–699

Kahnemann u. a. 1986

KAHNEMANN, D. ; KNETSCH, J. L. ; THALER, R.: Fairness as a Constraint on Profit Seeking. In: *American Economic Review* 76 (1986), S. 728–741

Kaldor 1939

KALDOR, N.: Welfare propositions of economics and interpersonal comparisons of utility. In: *The Economic Journal* 49 (1939), S. 549–552

Kaldor 1946

KALDOR, N.: A Comment. In: *Review of Economic Studies* 14 (1946), S. 49

Keen u. Wildasin 2000

KEEN, M. ; WILDASIN, D. E.: *Pareto Efficiency In International Taxation*. Version: November 2000. http://www.cesifo.de/DocCIDL/cesifo_wp371.pdf. – CESifo Working Paper

Kern u. Nida-Rümelin 1994

KERN, L. ; NIDA-RÜMELIN, J.: *Logik kollektiver Entscheidungen*. München, 1994

Kutschera 1982

KUTSCHERA, F. v.: *Grundlagen der Ethik*. Berlin, 1982

Külp 1982

KÜLP, B.: Wohlfahrtsökonomik I: Grundlagen. In: *Handwörterbuch der Wirtschaftswissenschaft* Bd. 9. Stuttgart, 1982, S. 469–486

Maddison 2004

MADDISON, A.: *The World Economy: Historical Statistics*. Version: August 2004. http://www.eco.rug.nl/~Maddison/Historical_Statistics/horizontal-file.xls. – Online-Ressource, Abruf: 8.12.04

Manning 1981

MANNING, R.: Environmental Ethics and John Rawls' Theory of Justice. In: *Environmental Ethics* 3 (1981), S. 155–165

Murray u. Lopez 1996

MURRAY, C. J. L. (Hrsg.) ; LOPEZ, A. D. (Hrsg.): *The Global Burden of Disease*. Cambridge, 1996

Nordhaus 1991

NORDHAUS, W. D.: To slow or not to slow: The economics of the greenhouse effect. In: *The Economic Journal* 101 (1991), S. 920–937

Nordhaus 1994

NORDHAUS, W. D.: Expert opinion on climatic change. In: *American Scientist* 82 (1994), S. 45–52

Nordhaus 1999

NORDHAUS, W. D.: *Roll the DICE Again: The Economics of Global Warming*. Version: Januar 1999. <http://www.econ.yale.edu/~nordhaus/homepage/rice98pap121898.PDF>. – Online-Ressource, Abruf: 19.04.04

Nordhaus u. Boyer 2000a

NORDHAUS, W. D. ; BOYER, J.: *Rice99_web_version.xls*. Version: November 2000. <http://www.econ.yale.edu/~nordhaus/homepage/Rice99webversion.xls>. – Online-Ressource, Abruf: 18.4.04

Nordhaus u. Boyer 2000b

NORDHAUS, W. D. ; BOYER, J.: *Roll the DICE Again: Economic Models of Global Warming*. Version: November 2000. <http://www.econ.yale.edu/~nordhaus/homepage/webtableofcontents102599.htm>. – Online-Ressource, Abruf: 18.4.04

Nordhaus u. Boyer 2000c

NORDHAUS, W. D. ; BOYER, J.: *Warming the World: Economic Models of Global Warming*. Cambridge, 2000

Nordhaus u. Boyer 2001

NORDHAUS, W. D. ; BOYER, J.: *RICE_2001_Bonn_base.xls*. Version: jun 2001. http://www.econ.yale.edu/~nordhaus/homepage/RICE_2001_Bonn_base.xls. – Online-Ressource, Abruf: 18.4.04

Nordhaus u. Yang 1996

NORDHAUS, W. D. ; YANG, Z.: A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies. In: *The American Economic Review* 86 (1996), S. 741–765

OECD 2004

OECD: *Entwicklungszusammenarbeit: Bericht 2003*. Paris, 2004

Olson 1958

OLSON, E.: Udenrigschandelens Gevinst. In: *Nationaløkonomisk Tidsskrift* (1958), Nr. 1-2, S. 76–79

Otani 1972

OTANI, Y.: Gains from Trade Revisited. In: *Journal of International Economics* 2 (1972), S. 127–156

Pareto 1917

PARETO, V.: *Traité de sociologie générale*. Paris, 1917

Pareto 1966

PARETO, V.: *Manuel D'Économie Politique*. Genève, 1966

Pareto 1972

PARETO, V.: *Manual of Political Economy*. London, 1972

Pogge 1989

POGGE, Th. W.: *Realizing Rawls*. Ithaca, 1989

Priddat 1998

PRIDDAT, B. P.: Moral Based Rational Man. In: *Brieskorn, N.; Wallacher, J. (Hrsg.), Homo oeconomicus: Der Mensch der Zukunft?* Stuttgart, 1998, S. 1–31

Putnam 1981

PUTNAM, H.: *Reason, Truth and History*. Cambridge, 1981

Rademacher 2004

RADEMACHER, F. J.: *Global Marshall Plan*. Hamburg, 2004

Rawls 1973

RAWLS, J.: *A Theory of Justice*. Oxford, 1973

Rawls 1974

RAWLS, J.: Some Reasons for the Maximin Criterion. In: *American Economic Review* 64 (1974), S. 141–146

Rawls 1975

RAWLS, J.: *Eine Theorie der Gerechtigkeit*. Frankfurt, 1975

Rawls 1992a

RAWLS, J.: *Die Idee des politischen Liberalismus*. Frankfurt, 1992

Rawls 1992b

RAWLS, J.: Der Vorrang der Grundfreiheiten. In: *Rawls, J., Die Idee des politischen Liberalismus*. Frankfurt, 1992, S. 159–254

Rawls 1993a

RAWLS, J.: The Basic Liberties and Their Priority. In: *Rawls, J., Political Liberalism*. New York, 1993, S. 289–371

Rawls 1993b

RAWLS, J.: *Political Liberalism*. New York, 1993

Rawls 1993c

RAWLS, J.: Priority of Right and Ideas of the Good. In: *Rawls, J., Political Liberalism*. New York, 1993, S. 173–211

Rawls 1995

RAWLS, J.: Political Liberalism: Reply to Habermas. In: *The Journal of Philosophy* 92 (1995), März, Nr. 3, S. 132–180

Rawls 1999a

RAWLS, J.: *Collected Papers*. Cambridge : Harvard University Press, 1999

Rawls 1999b

RAWLS, J.: A Kantian Conception of Equality. In: *Rawls, J., Collected Papers*. Cambridge, 1999, S. 254–266

Rawls 1999c

RAWLS, J.: Kantian Constructivism in Moral Theory. In: *Rawls, J., Collected Papers*. Cambridge, 1999, S. 303–358

Rawls 1999d

RAWLS, J.: *The Law of Peoples*. Cambridge, 1999

Rawls 1999e

RAWLS, J.: Outline of a Decision Procedure for Ethics. In: *Rawls, J., Collected Papers*. Cambridge, 1999, S. 1–19

Rawls 1999f

RAWLS, J.: Reply to Alexander and Musgrave. In: *Rawls, J., Collected Papers*. Cambridge, 1999, S. 232–253

Rawls 1999g

RAWLS, J.: *A Theory of Justice*. Oxford : Oxford University Press, 1999

Rawls 2001

RAWLS, J.: *Justice as Fairness*. Cambridge, 2001

Rawls 2002

RAWLS, J.: *Das Recht der Völker*. Berlin, 2002

Rawls 2003

RAWLS, J.: *Gerechtigkeit als Fairneß*. Frankfurt, 2003

Rescher 1991

RESCHER, N.: *Baffling Phenomena*. Savage, 1991

Ricken 1993a

RICKEN, F.: Endliche abwägende Vernunft. In: *Ethik und Sozialwissenschaften* 4 (1993), Nr. 4, S. 647–654

Ricken 1993b

RICKEN, F.: Ethik der abwägenden Vernunft. In: *Ethik und Sozialwissenschaften* 4 (1993), Nr. 4, S. 587–594

Robbins 1984

ROBBINS, L.: *An Essay on the Nature and Significance of Economic Science*. London, 1984

Roberts 2000

ROBERTS, R.: Speaking about Trade to the Open-Minded Skeptic. In: *Cato Journal* 19 (2000), Nr. 3, S. 439–448

Roemer 1996

ROEMER, J. E.: *Theories of Distributive Justice*. Cambridge, 1996

Rose u. Sauernheimer 1992

ROSE, K. ; SAUERNHEIMER, K.: *Theorie der Außenwirtschaft*. München, 1992

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2003

SACHVERSTÄNDIGENRAT ZUR BEGUTACHTUNG DER GESAMTWIRTSCHAFTLICHEN ENTWICKLUNG: *Staatsfinanzen Konsolidieren - Steuersystem Reformieren. Jahresgutachten 2003/04*. 2003

Samuelson 1950

SAMUELSON, P. A.: Evaluation of Real National Income. In: *Oxford Economic Papers* 1 (1950), S. 1–29

Samuelson 1956

SAMUELSON, P. A.: Social Indifference Curves. In: *Quarterly Journal of Economics* 70 (1956), S. 1–22

Samuelson 1962

SAMUELSON, P. A.: The Gains from International Trade Once Again. In: *Economic Journal* 72 (1962), S. 820–829

Samuelson 1963

SAMUELSON, P. A.: *Foundations of Economic Analysis*. Cambridge, 1963

Samuelson 1977

SAMUELSON, P. A.: Reaffirming The Existence of „Reasonable“ Bergson-Samuelson Social Welfare Functions. In: *Economica* 44 (1977), S. 81–88

Scapparone 1997

SCAPPARONE, P.: The Maximum of Collective Ophelimity in the Cours D'Économie Politique. In: *History of Economic Ideas* V (1997), Nr. 3, S. 35–48

Schlieper 1982

SCHLIEPER, U.: Wohlfahrtsökonomik II: Theorie des Zweitbesten. In: *Handwörterbuch der Wirtschaftswissenschaft*. Stuttgart, 1982, S. 486–493

Schultze 1998

SCHULTZE, R.: Katastrophenmedizin/Katastrophenschutz. In: *Lexikon der Bioethik* Bd. 2. Gütersloh, 1998, S. 342–346

Scitovsky 1941

SCITOVSKY, T.: A note on welfare propositions in economics. In: *The Review of Economics Studies* 9 (1941), S. 77–88

Scitovsky 1993

SCITOVSKY, T.: *Welfare and Competition*. Aldershot, 1993

Sen 1999

SEN, A.: *Ökonomie für den Menschen*. München, 1999

Sen u. Williams 1977

SEN, A. (Hrsg.) ; WILLIAMS, B. (Hrsg.): *Utilitarianism and Beyond*. Cambridge, 1977

Sen 1970

SEN, A. K.: *Collective Choice and Social Welfare*. San Francisco, 1970

Sen 1986

SEN, A. K.: Social Choice Theory. In: *Handbook of Mathematical Economics* Bd. 3. Amsterdam, 1986, S. 1073–1181

Shahriari u. a. 2001

SHAHRIARI, H. ; BELLI, P. ; LEWIS, M.: *Institutional Issues in Informal Health Payments in Poland*. Version: Februar 2001. http://www1.worldbank.org/hnp/Pubs_Discussion/Shahriari-InstitutionalIssues-whole.pdf. – HNP Discussion Paper

Slaughter 1999

SLAUGHTER, M. J.: Globalization and Wages: A Tale of Two Perspectives. In: *World Economy* 22 (1999), Nr. 5, S. 609–630

Social Watch 2003

SOCIAL WATCH: *Social Watch Report 2003*. Montevideo, 2003

Sohmen 1976

SOHMEN, E.: *Allokationstheorie und Wirtschaftspolitik*. Tübingen, 1976

Stringham 2001

STRINGHAM, E.: Kaldor-Hicks Efficiency and the Problem of Central Planning. In: *The Quarterly Journal Of Austrian Economics* 4 (2001), Nr. 2, S. 41–50

Trapp 1988

TRAPP, R. W.: *„Nicht-klassischer“ Utilitarismus*. Frankfurt, 1988

UNAIDS, UNICEF, WHO 2004

UNAIDS, UNICEF, WHO: *Epidemiological Fact Sheets on HIV/AIDS and Sexually Transmitted Infections, 2004 Update*. Version: 2004. <http://www.who.int/GlobalAtlas/PDFFactory/HIV/index.asp?strSelectedCountry=AL>. – Online-Ressource, Abruf: 30.11.2004

UNESCO 1998

UNESCO: *World Culture Report 1998*. Paris, 1998

UNSP 2004

UNSP: *National Accounts Main Aggregates Database*. Version: 2004. <http://unstats.un.org/unsd/snaama/introduction.asp>. – Online-Ressource, Abruf: 9.12.04

Varian 1992

VARIAN, H. R.: *Microeconomics Analysis*. New York, 1992

Weber 1973a

WEBER, M.: *Gesammelte Aufsätze zur Wissenschaftslehre*. Tübingen, 1973

Weber 1973b

WEBER, M.: Die „Objektivität“ sozialwissenschaftlicher und sozialpolitischer Erkenntnis. In: *Weber, M., Gesammelte Aufsätze zur Wissenschaftslehre*. Tübingen, 1973, S. 146–214

Weber 1973c

WEBER, M.: Der Sinn der „Wertfreiheit“ der soziologischen und ökonomischen Wissenschaften. In: *Weber, M., Gesammelte Aufsätze zur Wissenschaftslehre*. Tübingen, 1973, S. 489–540

Weinhardt 1989

WEINHARDT, C.: The Trade Off between „Real“ Equity and „Real“ Efficiency in Welfare Measurement Theory / Institut für Wirtschaftstheorie und OR, Universität Karlsruhe. Karlsruhe, 1989 (366). – Discussion Paper

Weltbank 1997

WELTBANK: *Weltentwicklungsbericht 1997*. Bonn, 1997

Weltbank 2004

WELTBANK: *World Development Indicators*. Washington, 2004

Weymark 1991

WEYMARK, J. A.: A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism. In: *Elster, J.; Roemer, J. E. (Hrsg.), Interpersonal Comparisons of Well-Being*. Cambridge, 1991, S. 255–320

Winters u. a. 2004

WINTERS, L. A. ; MCCULLOCH, N. ; MCKAY, A.: Trade Liberalization and Poverty: The Evidence So Far. In: *Journal of Economic Literature* 42 (2004), Nr. 1, S. 72–115

Wood 1995

WOOD, A.: How Trade Hurt Unskilled Workers. In: *Journal of Economic Perspectives* 9 (1995), Nr. 3, S. 57–80

Yamashige 1995

YAMASHIGE, S.: *Subjectively Envy-Free Allocation*. Toronto, 1995. – Working Paper / Department of Economics and Institute of Policy Analysis, University of Toronto

Zajac 1995

ZAJAC, E. E.: *Political Economy of Fairness*. Cambridge, 1995

Curriculum Vitae

Persönliche Daten

P. Andreas Gösele SJ
Kaulbachstr. 31a
80539 München

Tel.: (0 89) 23 86 23 58
E-Mail: goesele@hfph.mwn.de

Geb. am 29. 04. 1959 in Singen am Hohentwiel

Lebenslauf

1965-1969	Grundschule in Singen und Engen (Hegau)
1969-1979	Gymnasium in St. Blasien und Konstanz, Abitur am 22.5.1979
1979-1980	Freiwilliges Soziales Jahr in Singen
1980-1982	Noviziat der Gesellschaft Jesu in Nürnberg
1982-1984	Grundstudium der Philosophie München, Hochschule für Philosophie
1984-1986	Interstiz in Nürnberg (Caritas) und Äthiopien (JRS)
1986-1990	Studium der Theologie in Belo Horizonte, Brasilien; Abschluß mit dem Bakkalaureat in Theologie
1990	Priesterweihe
1990-1994	Studium der Volkswirtschaft an der LMU, München; Abschluß mit dem Diplom-Volkswirt Univ.
1995-2007	Doktoratsstudium an der LMU
seit 1995	Wissenschaftlicher Mitarbeiter des Instituts für Gesellschaftspolitik an der Hochschule für Philosophie, München
2001-2005	Lehrbeauftragter an der Hochschule für Philosophie, München
seit 2006	Lehrbeauftragter an der Hochschule für Philosophie, München

München, 1. März 2007