DOI: 10.20378/irb-58030

Kapitel VIII

Zur rationalen Fundierung einer neo-utilitaristischen Wohlfahrtsökonomie

JOHANNES SCHMIDT

1. Einführung

Die Geschichte der Wohlfahrtsökonomie ist — überspitzt formuliert — das Ergebnis einer dauerhaften Auseinandersetzung mit dem Utilitarismus. Gingen die bedeutendsten Vertreter der ‚alten' Wohlfahrtsökonomie noch ohne weiteres davon aus, daß sich die ethische Bewertung gesellschaftlicher Zustände an der Maximierung der Nutzensumme zu orientieren habe (vgl. Edgeworth 1881; Marshall 1890; Pigou 1920), so wandte sich die Disziplin im Zuge der Arbeiten von Pareto (1897), Robbins (1932), Bergson (1938) und Samuelson (1947) entschieden von ihrer utilitaristischen Tradition ab. Diese Wendung zu einer ‚neuen' Wohlfahrtsökonomie wurde mit einer umfassenden Kritik untermauert, die sich sowohl gegen die — meist nur implizit eingebrachten — ethischen Prämissen der ‚alten' Wohlfahrtsökonomie als auch gegen deren deskriptive Annahmen zur kardinalen Meßbarkeit und interpersonellen Vergleichbarkeit der individuellen Nutzenwerte richtete. Als konstruktives Resultat dieser Kritik präsentierten Bergson und Samuelson das Konzept einer ‚individualistischen' Sozialen Wohlfahrtsfunktion (SWF), das lediglich von ordinalen (sowie interpersonell nicht vergleichbaren) Nutzenindikatoren ausgeht und die ethische Bewertung gesellschaftlicher Zustände nur an das Paretoprinzip bindet.

Dem dezidierten Paretianismus der ‚neuen' Wohlfahrtsökonomie steht eine ganze Reihe von Versuchen gegenüber, die utilitaristische Ethik im konzeptuellen Rahmen einer individualistischen SWF wiederzubeleben

(vgl. Vickrey 1945, 1960; Fleming 1952; Harsanyi 1953, 1955; Ng 1975; Hammond 1987). Das gemeinsame Anliegen dieser neo-utilitaristischen Ansätze besteht darin, die umfangreiche Klasse der mit dem Bergson-Samuelson Konzept grundsätzlich vereinbaren ethischen Prinzipien durch die Formulierung zusätzlicher normativer Restriktionen auf das Nutzensummenprinzip zuzuspitzen. John C. Harsanyi, der einflußreichste Vertreter dieser Richtung, präsentiert sowohl einen axiomatischen als auch einen entscheidungstheoretischen Ansatz zur Fundierung einer utilitaristischen SWF. Vor dem Hintergrund einer metaethischen Konzeption, die moralische Prinzipien als hypothetische Imperative rekonstruiert und die Moralphilosophie als Teildisziplin einer allgemeinen Theorie des rationalen Verhaltens betrachtet, erhebt Harsanyi mit beiden Ansätzen den Anspruch, daß sich eine äquivalente Fundierung des utilitaristischen Prinzips auf schwache ethische Bedingungen stützen läßt, wenn diese Bedingungen mit einem speziellen Rationalitätskonzept verknüpft werden.

Die folgende Untersuchung beschäftigt sich mit der Frage, ob das von Harsanyi favorisierte Rationalitätskonzept tatsächlich die Leistung erbringt, die er sich von ihm verspricht. Zu Beginn werden die methodologischen, konzeptuellen und technischen Voraussetzungen des Arguments skizziert (Abschnitt 2). Im Mittelpunkt steht dann eine immanente Kritik des axiomatischen (Abschnitt 3) bzw. entscheidungstheoretischen Fundierungsansatzes (Abschnitt 4). Einige allgemeine Bemerkungen zu Harsanyis Unternehmung beschließen den Beitrag (Abschnitt 5).

2. Grundlagen

Die Mengen $N = \{1, ..., i, ..., n\}$, $M = \{1, ..., j, ..., m\}$ und $X = \{x, y, ..., w\}$ sollen die n Individuen i einer Gesellschaft, die m in dieser Gesellschaft verfügbaren Güter j und die r realisierbaren gesellschaftlichen Zustände x, y, etc. repräsentieren ($\#X = r$). Die gesellschaftlichen Zustände $x \in X$ seien als alternative Güterallokationen spezifiziert: $x = (x_1, ..., x_i, ..., x_n)$, $x_i = (x_{i1}, ..., x_{ij}, ..., x_{im})$. Über der Menge X seien eine gesellschaftliche Präferenzordnung R (mit der symmetrischen bzw. asymmetrischen Komponente I bzw. P) und n individuelle Präferenzordnungen R_i (mit ihren Komponenten I_i und P_i) definiert. Während die binären Präferenzrelationen R („mindestens so gut wie"), I („ebensogut wie") und P („besser als") die ‚neutralen' Bewertungen eines imaginären gesellschaftlichen Planers (bzw. ethischen Beobachters) modellieren, repräsentieren die entsprechenden Relationen R_i, I_i und P_i die subjektiven Präferenzen

der einzelnen Gesellschaftsmitglieder. Vor diesem Hintergrund läßt sich das von Bergson (1938) und Samuelson (1947) entwickelte Konzept einer individualistischen SWF auf eine gesellschaftliche Präferenzordnung R reduzieren, die die individuellen Präferenzordnungen R_i insofern respektiert, als sie der folgenden Paretobedingung genügt:

*Bedingung P**. $\forall$ x, y $\in$ X:
(1) $(\forall\, i \in N\colon xI_iy) \rightarrow xIy$
(2) $[(\forall\, i \in N\colon xR_iy) \,\&\, (\exists\, i \in N\colon xP_iy)] \rightarrow xPy.$

Über die bisher skizzierten Vorgaben hinaus geht man in der Bergson-Samuelson Tradition regelmäßig davon aus, daß eine reellwertige SWF W und n reellwertige Nutzenfunktionen U_i existieren, die die gesellschaftliche Präferenzordnung R bzw. die individuellen Präferenzordnungen R_i numerisch repräsentieren:

(1) $\forall$ x, y $\in$ X: $xRy \leftrightarrow W(x) \geq W(y)$
(2) $\forall$ x, y $\in$ X: $xR_iy \leftrightarrow U_i(x) \geq U_i(y)$, $(i = 1, \ldots, n)$.

Verknüpft man diese beiden Repräsentationsprämissen, die eine ordinale SWF W und n ordinale Nutzenfunktionen U_i definieren, mit der Substanz der Bedingung P*, so erhält man die in der Wohlfahrtsökonomie übliche Spezifizierung einer individualistischen SWF:

*Bedingung W**. (1) $\forall$ x $\in$ X: $W(x) = W(U_1(x), \ldots, U_n(x))$
(2) $\forall$ i $\in$ N: $\delta W/\delta U_i > \emptyset$.

Während Bergson und Samuelson an einer eindeutigen Spezifizierung von W keinerlei Interesse zeigen, geht es den Neoutilitaristen darum, das breite Spektrum der mit der Bedingung W* vereinbaren Sozialen Wohlfahrtsfunktionen auf eine additive SWF zu reduzieren:

$$\forall\, x \in X\colon W(x) = \sum_{i=1}^{n} U_i(x).$$

Die von Harsanyi zur Fundierung einer utilitaristischen SWF vorgeschlagenen Ansätze beruhen auf einer metaethischen Konzeption, die der Moralphilosophie eine doppelte Beschränkung auferlegt (vgl. Harsanyi 1958, 1977a). Da Harsanyi ethische Prinzipien als hypothetische Imperative interpretiert, weist er der Moralphilosophie zunächst einmal die relativ bescheidene Aufgabe zu, den Gesellschaftsmitgliedern eine Menge analytisch wahrer Sätze zur Verfügung zu stellen, die jeweils ein spezielles ethisches Prinzip mit einer logisch äquivalenten Menge allgemeiner mo-

ralischer Bedingungen untermauern. Die Beschränkung der Moralphilosophie auf den Beweis analytisch wahrer Sätze hat den Vorzug, daß jedes — als hypothetischer Imperativ rekonstruierte — ethische Prinzip zumindest insofern objektive Gültigkeit beanspruchen kann, als der in ihm enthaltene logische Zusammenhang von jedermann akzeptiert werden muß. Während die Gewährleistung dieser ‚hypothetischen' objektiven Gültigkeit ausschließlich von den logischen Fertigkeiten des Moralphilosophen abhängt, kann ihm die Fundierung eines ‚faktisch' allgemeingültigen — d. h.: von allen Individuen als verbindlich zu betrachtenden — ethischen Prinzips nur dann gelingen, wenn eine Menge elementarer moralischer Bedingungen existiert, die zum einen von jedermann akzeptiert werden, und zum anderen einen eindeutigen Grundsatz erzeugen.

Die zweite Beschränkung, die Harsanyi der Moralphilosophie auferlegt, resultiert aus einer zusätzlichen metaethischen Prämisse, derzufolge das Fällen moralischer Urteile als spezielle Form des rationalen individuellen Verhaltens zu betrachten ist. Diese Prämisse, mit der die Ethik — neben der Entscheidungstheorie und Spieltheorie — als dritte Disziplin einer umfassenden Theorie des rationalen Verhaltens etabliert wird, hat die bemerkenswerte Konsequenz, daß sich die äquivalente Fundierung eines moralischen Prinzips nicht ausschließlich auf eine Reihe genuin ethischer Bedingungen stützen kann, sondern darüber hinaus immer auf ein formales Kriterium der individuellen Rationalität zurückgreifen muß. Die Unverzichtbarkeit einer rationalen Begründung ethischer Prinzipien ist für Harsanyis Unternehmung insofern von größter Bedeutung, als sie auf eine universelle Fundierung des Neoutilitarismus zielt. Da die skizzierte metaethische Konzeption unter diesen Umständen nur die Verwendung schwacher ethischer Bedingungen zuläßt, für die zumindest allgemeine Akzeptierbarkeit reklamiert werden kann, ruht die Last der Deduktion einer utilitaristischen SWF zwangsläufig zu einem großen Teil auf dem favorisierten Rationalitätskriterium.

Harsanyi stützt seine beiden Fundierungsansätze auf das Rationalitätskonzept der Bayes'schen Entscheidungstheorie. Das dieser Theorie zugrundeliegende Entscheidungsproblem läßt sich mit der Menge $X^* = \{x^*, y^*, z^*, \ldots\}$ aller Wahrscheinlichkeitsverteilungen modellieren, die über der Menge X definiert werden können ($X \subset X^*$). Während alle $x \in X$ als gesellschaftliche Handlungsmöglichkeiten zu interpretieren sind, die mit Sicherheit eine bestimmte Güterallokation $(x_1, \ldots, x_n)$ erzeugen, repräsentiert jede Alternative $x^* \in X^*$ eine mögliche kollektive Aktion, die mit angebbaren objektiven Wahrscheinlichkeiten p_1, p_2, etc. zu unterschiedlichen Allokationen x, y, etc. führen kann. Die Elemente der Menge X^*

sind demnach grundsätzlich als riskante Alternativen der folgenden Form zu kennzeichnen:

$$x^* = (x, p_1; \ldots; z, p_k; \ldots; w, p_r),$$

$$\text{wobei } p_k \geq 0 \text{ für } k = 1, \ldots, r \text{ und } \sum_{k=1}^{r} p_k = 1.$$

Um die Präsentation zu vereinfachen, werden im folgenden lediglich riskante Alternativen der Form $x^* = (x, p; y, 1 - p)$ betrachtet. Sieht man nun von der speziellen Interpretation der Alternativenmenge X^* zunächst einmal ab, so läßt sich der Ausgangspunkt der Bayes'schen Entscheidungstheorie mit dem Problem eines Individuums i beschreiben, aus der Menge X^* eine — gemessen an seinen persönlichen Präferenzen — optimale Aktion auszuwählen. Zur Lösung dieses Problems bietet die Theorie eine Reihe von Axiomen an, die — folgt man Harsanyis Intentionen — als normative Anforderungen an die Rationalität des individuellen Entscheidungsverhaltens bei Risiko zu interpretieren sind (vgl. Harsanyi 1987). Die folgende Formulierung der Bayes'schen Rationalitätspostulate kommt mit drei Axiomen aus (vgl. Herstein/Milnor 1953):

Axiom I. Über der Menge X^* existiert eine Präferenzordnung R_i.

Axiom II. Für alle $x, y, z \in X$ sind die Mengen $\{p|\ (x, p; y, 1 - p)R_i z\}$ und $\{p|\ zR_i(x, p; y, 1 - p)\}$ abgeschlossen.

Axiom III. $\forall\ x, y, z \in X$: $xI_iy \rightarrow (x, p; z, 1 - p)I_i(y, p; z, 1 - p)$.

Mit dem ersten Postulat wird das elementare Rationalitätskonzept der ökonomischen Theorie von der ursprünglichen Alternativenmenge X auf die Menge X^* ausgedehnt. Das zweite Axiom formuliert eine Stetigkeitsbedingung, die dafür sorgt, daß geringe Veränderungen des Wahrscheinlichkeitswertes p nur geringe Veränderungen in der Bewertung einer riskanten Alternative $(x, p; y, 1 - p)$ nach sich ziehen. Das in dieser Stetigkeitsbedingung enthaltene Rationalitätspostulat ist deutlich zu erkennen, wenn man eine spezielle Variante des Axioms II betrachtet, die in der Bayes'schen Entscheidungstheorie häufig verwendet wird (vgl. von Neumann/Morgenstern 1944; Marschak 1950; Luce/Raiffa 1957):

Axiom II'. Für alle $x, y, z \in X$ mit xP_iyP_iz gilt: Es existiert genau ein Wahrscheinlichkeitswert p, $0 < p < 1$, so daß $(x, p; z, 1 - p)I_iy$.

Das dritte Axiom, das unter den Bezeichnungen *sure-thing principle* (vgl. Savage 1954) bzw. *strong independence axiom* (vgl. Samuelson 1952) geführt

wird, ist als zentrales Rationalitätspostulat der Theorie zu betrachten. Mit ihm wird die Forderung erhoben, daß sich die Bewertung einer riskanten Alternative (x, p; y, 1 − p) nicht verändern darf, wenn die Komponente x durch eine als gleich gut erachtete Alternative y ersetzt wird.

Die wesentliche Leistung der Bayes'schen Entscheidungstheorie besteht in dem Nachweis, daß ihre Postulate sowohl die Existenz einer kardinalen Repräsentation der individuellen Präferenzen als auch eine spezielle Maxime für den rationalen Umgang mit riskanten Situationen begründen. Genügen die Präferenzen eines Individuums i den Axiomen I–III, so läßt sich über der Menge X* eine kardinale Nutzenfunktion U_i definieren, die die Präferenzordnung R_i numerisch repräsentiert und alle riskanten Alternativen x* in der folgenden Weise bewertet:

$$U_i(x^*) = U_i(x, p; y, 1 - p) = p \cdot U_i(x) + (1 - p) \cdot U_i(y)$$

Eine Nutzenfunktion dieses Typs wird üblicherweise als ‚von Neumann/Morgenstern- (vNM-) Nutzenfunktion' bezeichnet. Da die Axiome I–III weder den Nullpunkt noch die Einheit einer vNM-Nutzenfunktion eindeutig festlegen, wird mit der Definition von U_i implizit eine Klasse U_i' von kardinalen Nutzenfunktionen beschrieben, die die Präferenzen eines Individuums i ebensogut repräsentieren wie U_i:

$$U_i'(\cdot) = \alpha_i + \beta_i \cdot U_i(\cdot),\ \beta_i > 0$$

Mit der Existenz einer vNM-Nutzenfunktion wird offensichtlich auch ein formales Kriterium des rationalen Verhaltens bei Risiko begründet. Ein im Bayes'schen Sinne rationales Individuum wird nämlich aus einer Menge riskanter Optionen immer diejenige auswählen, die seinen erwarteten Nutzen maximiert.

3. Axiomatischer Ansatz

Der von Harsanyi (1955, 1977 c) zur axiomatischen Fundierung einer utilitaristischen SWF vorgeschlagene Ansatz operiert über die Bayes'schen Rationalitätspostulate hinaus nur mit einer einzigen ethischen Bedingung, deren allgemeine Akzeptierbarkeit sich wohl kaum bestreiten läßt:

Postulat a. Die Präferenzen des gesellschaftlichen Planers erfüllen die Axiome I–III.

Postulat b. Die Präferenzen aller Individuen i (i = 1, ..., n) erfüllen die Axiome I–III.

Postulat c. $\forall\, x^*, y^* \in X^*: (\forall\, i \in N: x^* I_i y^*) \rightarrow x^* I y^*$.

Mit der Bedingung a wird das Bayes'sche Rationalitätskonzept auf die neutralen Bewertungen des gesellschaftlichen Planers übertragen. Genügen die Präferenzen dieses ethischen Beobachters den Axiomen I—III, so läßt sich über der Menge X^* eine kardinale SWF W definieren, die die gesellschaftliche Präferenzordnung R numerisch repräsentiert und die formalen Eigenschaften einer vNM-Nutzenfunktion besitzt, sich bei der Bewertung einer riskanten Alternative x^* also immer an der erwarteten sozialen Wohlfahrt orientiert (‚vNM-SWF'):

$$W(x^*) = W(x, p; y, 1 - p) = p \cdot W(x) + (1 - p) \cdot W(y)$$

Die Bedingung b hat die technische Konsequenz, daß über der Menge X^* n vNM-Nutzenfunktionen U_i existieren, die die persönlichen Präferenzen der Gesellschaftsmitglieder abbilden. Während die Postulate a und b die von Bergson und Samuelson favorisierten Repräsentationsprämissen in spezifischer Weise verschärfen, wendet das Postulat c lediglich die Indifferenz-Komponente der Bedingung P* auf die Bewertung riskanter gesellschaftlicher Alternativen an. Es ist ohne weiteres zu erkennen, daß dieses schwache Individualismus-Postulat die vNM-SWF W als eindeutige Funktion der vNM-Nutzenindikatoren U_i spezifiziert:

$$\forall\, x^* \in X^*: W(x^*) = W(U_1(x^*), \ldots, U_n(x^*))$$

Daß die Postulate a—c über diese triviale Implikation hinaus stark genug sind, um alle nicht-linearen Sozialen Wohlfahrtsfunktionen zu eliminieren, zeigt das folgende, von Harsanyi (1955) bewiesene

Theorem. Erfüllt eine vNM-SWF W für ein gegebenes Profil individueller vNM-Nutzenfunktionen $(U_1, \ldots, U_n)$ das Postulat c, dann existieren n reelle Zahlen a_i, so daß für alle $x^* \in X^*$ gilt:

$$W(x^*) = \sum_{i=1}^{n} a_i \cdot U_i(x^*).$$

Dieses Ergebnis ist zweifellos erstaunlich stark. Sobald man individuelle und kollektive Entscheidungen an das Bayes'sche Rationalitätskonzept bindet, reicht eine harmlose ethische Bedingung bereits aus, um eine sehr spezielle Klasse von Sozialen Wohlfahrtsfunktionen zu erzeugen. Im Zuge der axiomatischen Fundierung einer über X^* definierten linearen SWF bietet Harsanyi überdies drei Bedingungen an, die eine identische ex-ante- und ex-post-Bewertung riskanter gesellschaftlicher Alternativen garantie-

ren (vgl. etwa Hammond 1983). Mit dieser spezifischen Leistung der Postulate a—c ist es zu erklären, daß sich die normative Kritik des Harsanyi-Theorems bisher auf das Postulat a konzentriert hat.[1] Die folgenden Überlegungen zielen demgegenüber auf eine rein immanente Kritik des Harsanyi-Theorems. Es soll gezeigt werden, daß die Postulate a—c, selbst wenn man sie allesamt akzeptiert, von einer substantiell interessanten Fundierung des utilitaristischen Prinzips noch weit entfernt sind.

Die begrenzte Reichweite des Harsanyi-Theorems ist zunächst einmal daran zu erkennen, daß es per se weder eine negative Gewichtung ($a_i < 0$) noch eine völlige Ignorierung der Interessen bestimmter Individuen ($a_i = 0$) ausschließt. Um eine strikt positive Gewichtung aller individuellen Nutzenwerte ($\forall\, i \in N\colon a_i > 0$) — und damit: eine notwendige Voraussetzung für die utilitaristische Ermittlung der sozialen Wohlfahrt — zu gewährleisten, sind Harsanyis Postulate um zwei zusätzliche Vorkehrungen zu ergänzen. In diesem Zusammenhang ist zu beachten, daß der von Harsanyi (1955, 1977 c) präsentierte Beweis des Theorems über die Postulate a—c hinaus zwei technische Prämissen enthält:

(1) Es existiert ein gesellschaftlicher Zustand $x^0 \in X$, so daß
$W(x^0) = U_1(x^0) = \ldots = U_n(x^0) = 0$.

(2) Für jedes Individuum h ($h = 1, \ldots, n$) existiert eine Alternative $x^h \in X$, so daß
$U_h(x^h) = 1$ und $U_i(x^h) = 0$ für alle $i \neq h$.

Da diese beiden Annahmen dafür sorgen, daß neben dem Nutzenvektor $(0, \ldots, 0)$ jeder der n Einheitsvektoren $[(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)]$ einen gesellschaftlichen Zustand repräsentiert, operiert Harsanyi implizit mit der folgenden substantiellen Restriktion (vgl. Jeffrey 1971):

(3) Es existiert ein gesellschaftlicher Zustand $x^0 \in X$, und es existiert für jedes Individuum h ($h = 1, \ldots, n$) eine Alternative $x^h \in X$, so daß
$x^h P_h x^0$ und $x^h I_i x^0$ für alle $i \neq h$.

Nun ist zwar vielfach gezeigt worden, daß die logischen Implikationen der Postulate a—c in keiner Weise von den Prämissen (1) und (2) bzw. (3)

[1] Im Mittelpunkt der Kritik steht dabei die Anwendung des *sure-thing principle* auf gesellschaftliche Entscheidungen. Vgl. dazu grundlegend Diamond (1967) und Sen (1970), S. 143—145 sowie Deschamps/Gevers (1977), McClennen (1981), Broome (1984) und Schmidt (1991), S. 140—151.

abhängen.[2] Der in (3) enthaltenen Normierung der Alternativenmenge kommt aber eine erhebliche Bedeutung zu, wenn das Postulat c zur üblichen Paretobedingung verschärft wird:

Postulat d. $\forall\, x^*, y^* \in X^*$:
(1) $(\forall\, i \in N: x^* I_i y^*) \rightarrow x^* I y^*$
(2) $[(\forall\, i \in N: x^* R_i y^*) \,\&\, (\exists\, i \in N: x^* P_i y^*)] \rightarrow x^* P y^*$.

Während nämlich Harsanyi (1977 c, 1978) davon ausgeht, daß die Postulate a, b und d unmittelbar eine positive Gewichtung aller individuellen Nutzenindikatoren erzeugen, haben Resnik (1983) und Fishburn (1984) gezeigt, daß die Paretobedingung d mit einer speziellen Normierung der Alternativenmenge verknüpft werden muß, um die ihr zugedachte Funktion erfüllen zu können. Der Zusammenhang ist deutlich zu erkennen, wenn man die Implikationen der Postulate a, b und d vor dem Hintergrund der Annahmen (1) und (2) betrachtet:

$$W(x^h) = \sum_{i=1}^{n} a_i \cdot U_i(x^h) = a_h > 0 = W(x^0),\ (h = 1, \ldots, n)$$

Nun ist mit dem Nachweis, daß die Postulate a, b und d eine lineare SWF erzeugen, die unter den Prämissen (1) und (2) bzw. (3) eine positive Gewichtung aller individuellen Interessen garantiert, für die Fundierung des utilitaristischen Prinzips wenig gewonnen, solange es nicht gelingt, die — abgesehen von der Normalisierungsannahme (1) — willkürlich gewählten vNM-Indikatoren U_i in einer interpersonell vergleichbaren Nutzeneinheit auszudrücken. Eine triviale ‚Lösung' dieses Problems besteht offensichtlich darin, die Gewichtungsparameter a_i kurzerhand als Vergleichsoperatoren zu interpretieren und das Ergebnis des Theorems in der folgenden Weise zu reformulieren:

$$W(\cdot) = \sum_{i=1}^{n} V_i(\cdot), \text{ wobei } V_i(\cdot) = a_i \cdot U_i(\cdot) \text{ für } i = 1, \ldots, n.$$

Dieses Verfahren (vgl. Harsanyi 1955) ignoriert allerdings die Tatsache, daß der Beweis des Theorems ohne jede Vergleichbarkeitsprämisse auskommt (vgl. Harsanyi 1978), die a_i also lediglich als formale, substantiell nicht näher spezifizierte Gewichtungsparameter ausweist. Während die schlichte Etikettierung der a_i als Vergleichsoperatoren lediglich ein triviales

[2] Zu alternativen Beweisen des Harsanyi-Theorems, die ohne diese zusätzlichen Prämissen auskommen, vgl. Camacho/Sonstelie (1974), Domotor (1979), Fishburn (1984), Border (1985), Selinger (1986) und Coulhon/Mongin (1989).

Repräsentationstheorem erzeugt, lassen sich die Postulate a, b und d für eine substantiell interessante Fundierung des utilitaristischen Prinzips nutzen, wenn man jeden Gewichtungsparameter a_i in eine deskriptive Komponente c_i ('Vergleichsgewicht') und eine normative Komponente e_i ('ethisches Gewicht') zerlegt, so daß Harsanyis lineare SWF die folgende Form annimmt (vgl. Harsanyi 1977 c; Brock 1980):

$$W(\cdot) = \sum_{i=1}^{n} e_i \cdot c_i \cdot U_i(\cdot),$$

wobei $c_i > 0$, $e_i > 0$ und $a_i = e_i \cdot c_i$ für $i = 1, \ldots, n$.

Die Spezifizierung des Vektors $(c_1, \ldots, c_n)$ ist als Resultat einer Serie von deskriptiven interpersonellen Nutzenvergleichen zu interpretieren, deren Ziel darin besteht, das willkürlich gewählte vNM-Profil $(U_1, \ldots, U_n)$ in ein vNM-Profil $(U_1^0, \ldots, U_n^0)$ zu transformieren, das die Nutzenwerte aller Individuen in einer gemeinsamen Einheit mißt. Während die c_i auf der Grundlage empirischer Informationen über die relativen interpersonellen Präferenzintensitäten zu wählen sind, wird mit jeder Spezifizierung des Vektors $(e_1, \ldots, e_n)$ auf der Basis alternativer ethischer Kriterien über die relative Gewichtung der relevanten Nutzenindikatoren U_i^0 entschieden. Da die Festsetzung der c_i lediglich die Gewähr dafür bieten soll, daß das zur ethischen Bewertung gesellschaftlicher Zustände herangezogene vNM-Nutzenprofil die für die Verwendung einer linearen SWF (zumindest) erforderlichen interpersonellen Informationen enthält, läßt sich die normative Substanz des Harsanyi-Theorems mit der folgenden Klasse von Sozialen Wohlfahrtsfunktionen beschreiben:

$$W(\cdot) = \sum_{i=1}^{n} e_i \cdot U_i^0(\cdot),$$

wobei $U_i^0(\cdot) = c_i \cdot U_i(\cdot)$ und $e_i > 0$ für $i = 1, \ldots, n$.

Mit diesem substantiellen Ergebnis ist eine wesentliche Vorentscheidung zugunsten des Neoutilitarismus verbunden. Da die gesellschaftliche Rangordnung zweier Alternativen x* und y* ausschließlich von der Frage abhängt, ob der numerische Wert der Summe

$$\sum_{i=1}^{n} e_i \cdot [U_i^0(x^*) - U_i^0(y^*)]$$

positiv (x*Py*), negativ (y*Px*) oder gleich null ist (x*Iy*), werden mit Harsanyis Theorem alle ethischen Prinzipien eliminiert, die bei der Bewertung gesellschaftlicher Zustände die Verteilung der individuellen Nut-

zenniveaus berücksichtigen.[3] Selbst wenn man also das Postulat b mit der empirischen Annahme interpersonell völlig vergleichbarer Nutzenindikatoren verknüpft, sorgt die gleichzeitige Verwendung der Postulate a und d unweigerlich dafür, daß sich die kollektive Präferenzordnung ausschließlich auf interpersonelle Vergleiche der individuellen Nutzendifferenzen stützt.

Obwohl die Postulate a, b und d alle ethischen Prinzipien ausschließen, die auf interpersonelle Vergleiche von Nutzenniveaus zurückgreifen, sind sie noch viel zu schwach, um eine utilitaristische Berechnung der sozialen Wohlfahrt zu garantieren. Solange den ethischen Gewichten e_i über die Positivitätsrestriktion hinaus keine weitere Beschränkung auferlegt wird, ist das Harsanyi-Theorem mit einem breiten Spektrum anti-utilitaristischer Prinzipien vereinbar. So ist es z. B. überhaupt kein Problem, durch eine geeignete Spezifizierung des Vektors $(e_1, \ldots, e_n)$ eine diktatorische SWF zu erzeugen, die die gesellschaftliche Rangordnung der Alternativen faktisch ausschließlich von den Präferenzen eines einzigen Individuums abhängig macht. Um zu einer äquivalenten Fundierung des utilitaristischen Prinzips zu kommen, ist Harsanyi daher gezwungen, über das Pareto-Kriterium hinaus eine weitere ethische Bedingung ins Spiel zu bringen (vgl. Harsanyi 1975, 1977 c):

Postulat e. Die vNM-SWF W ist eine symmetrische Funktion der — in der gleichen Einheit ausgedrückten — vNM-Nutzenindikatoren $(U_1^0, \ldots, U_n^0)$.

Erst diese zusätzliche Symmetrie-Bedingung bietet die Gewähr dafür, daß alle individuellen Nutzenwerte mit dem gleichen Gewicht (pro Nutzeneinheit) in die Ermittlung der sozialen Wohlfahrt eingehen:

$$W(\cdot) = \sum_{i=1}^{n} e \cdot U_i^0(\cdot) = e \cdot \sum_{i=1}^{n} U_i^0(\cdot),\ e > 0$$

Da das Postulat e nicht nur die typische (deskriptive) Vergleichbarkeitsprämisse der utilitaristischen Ethik, sondern — zumindest vor dem Hintergrund der skizzierten Vorentscheidung — auch deren charakteristische (normative) Gewichtungsmaxime auf direktem Wege einführt (vgl. Stras-

[3] Diese Eigenschaft des Harsanyi-Theorems gab den Anstoß zu einer umfangreichen Debatte, die sich mit den moralischen Qualitäten nicht-linearer Sozialer Wohlfahrtsfunktionen beschäftigt. Vgl. dazu vor allem Sen (1973, 1976, 1977) und Harsanyi (1975, 1977 b), aber auch Nunan (1981) und Gauthier (1982). Zur Fundierung einer Klasse von Sozialen Wohlfahrtsfunktionen, die die gesellschaftliche Präferenzordnung sowohl von der Summe der individuellen Nutzendifferenzen als auch von den Eigenschaften der Nutzenverteilung abhängig machen, vgl. Trapp (1988), Kap. II.1.

nick 1981), läßt sich die Harsanyis Unternehmung abschließende Bedingung wohl kaum noch als Repräsentant einer schwachen, allgemein akzeptierbaren moralischen Forderung interpretieren.

4. *Entscheidungstheoretischer Ansatz*

Der zweite Ansatz, den Harsanyi (1953, 1955) zur universellen Fundierung einer utilitaristischen SWF präsentiert, wendet das Bayes'sche Rationalitätskonzept auf ein hypothetisches Entscheidungsproblem an, dessen spezielle Konstruktion die moralische Qualität einer individuellen Entscheidung garantieren soll. Um dieses Entscheidungsproblem zu modellieren, greift Harsanyi (1977 c) auf die Konzepte einer ‚erweiterten Alternative' bzw. einer ‚erweiterten Präferenzordnung' zurück (vgl. Suppes 1966; Sen 1970). Unter einer erweiterten Alternative (x, i) wird eine vollständige Beschreibung der Position verstanden, die ein Individuum i in einem gesellschaftlichen Zustand x einnimmt. Diese Beschreibung umfaßt zum einen die objektiven Bedingungen des Individuums i (also etwa sein Güterbündel x_i) und zum anderen seine subjektiven Merkmale (wie z. B. seine persönlichen Präferenzen R_i). Dem Konzept einer erweiterten Präferenzordnung liegt die Annahme zugrunde, daß jedes Individuum h in der Lage ist, zwei beliebige hypothetische Alternativen (x, i) und (y, j) in eine Rangordnung zu bringen. Die erweiterte Präferenzordnung eines Individuums h läßt sich formal mit einer binären Präferenzrelation $\tilde{R}_h$ („mindestens so gut wie") beschreiben, die über der Menge $X \times N$ aller individuellen Positionen definiert ist. Mit seinem entscheidungstheoretischen Fundierungsansatz geht Harsanyi insofern über das übliche Konzept einer erweiterten Präferenzordnung hinaus, als er den Bewertungen des Individuums h riskante erweiterte Alternativen der Form $\hat{x} = [(x, i), p; (y, j), 1 - p]$ zugrundelegt. Das relevante Entscheidungsproblem läßt sich daher auf allgemeinstem Niveau mit der Menge $\hat{X} = \{\hat{x}, \hat{y}, \hat{z}, \ldots\}$ aller Wahrscheinlichkeitsverteilungen modellieren, die über der Menge $X \times N$ definiert werden können.

Um die Lösung dieses hypothetischen Entscheidungsproblems für die Fundierung einer utilitaristischen SWF zu nutzen, formuliert Harsanyi je eine Rationalitäts-, Unparteilichkeits- und Sympathiebedingung. Im ersten Schritt der Argumentation werden die erweiterten Präferenzen des Individuums h an die Bayes'schen Rationalitätspostulate gebunden. Diese Annahme hat die technische Konsequenz, daß sich über der Alternativenmenge $\hat{X}$ eine ‚erweiterte Nutzenfunktion' $\tilde{U}_h$ definieren läßt, die die

erweiterte Präferenzordnung $\tilde{R}_h$ numerisch repräsentiert und die Eigenschaften einer vNM-Nutzenfunktion besitzt. Ein im Bayes'schen Sinne rationales Individuum h wird demnach jeder individuellen Position (x, i) einen kardinalen Nutzenwert $\tilde{U}_h(x, i)$ zuordnen und riskante erweiterte Alternativen in der folgenden Weise bewerten:

$$\begin{aligned}\tilde{U}_h(\hat{x}) &= \tilde{U}_h[(x, i), p; (y, j), 1 - p] \\ &= p \cdot \tilde{U}_h(x, i) + (1 - p) \cdot \tilde{U}_h(y, j)\end{aligned}$$

Mit der Unparteilichkeitsbedingung wird derjenige Teil des Entscheidungsproblems isoliert, den Harsanyi als moralisch relevant erachtet. Folgt man diesem zweiten Schritt der Argumentation, so erzeugen die erweiterten Präferenzen eines Individuums h genau dann eine unparteiische Entscheidung, wenn sie sich ausschließlich auf Alternativen der Form [(x, 1), 1/n; ...; (x, n), 1/n], [(y, 1), 1/n; ...; (y, n), 1/n], etc. beziehen. Da exakt r riskante erweiterte Alternativen dieser Form existieren, führt Harsanyi mit der Unparteilichkeitsbedingung ein spezielles Kriterium für die ethische Bewertung der Alternativenmenge X ein. Bei der Bewertung der gesellschaftlichen Zustände $x \in X$ kommen demnach die ‚ethischen Präferenzen' eines Individuums h genau dann zur Geltung, wenn es seinen erweiterten Präferenzen die Annahme zugrundelegt, es könne mit der gleichen Wahrscheinlichkeit die n Positionen aller Gesellschaftsmitglieder i einnehmen. Um das abgeleitete Konzept einer ethischen Präferenz vom allgemeineren Konzept einer erweiterten Präferenz formal zu unterscheiden, schlägt Harsanyi vor, die ethischen Präferenzen eines Individuums h mit einer SWF W_h zu modellieren. Da der Definitionsbereich einer vNM-Nutzenfunktion $\tilde{U}_h$ mit der skizzierten Unparteilichkeitsbedingung auf alle riskanten erweiterten Alternativen der Form [(x, 1), 1/n; ...; (x, n), 1/n] reduziert wird, lassen sich die ethischen Präferenzen eines im Bayes'schen Sinne rationalen Individuums h offensichtlich in der folgenden Weise repräsentieren:

$$\forall\, x \in X\colon W_h(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{U}_h(x, i)$$

Der abschließenden Sympathiebedingung fällt die Aufgabe zu, dieses formale Resultat in eine substantiell utilitaristische SWF zu verwandeln. Ein Individuum h, das Harsanyis Sympathiekriterium genügt, wird bei der Bewertung jeder erweiterten Alternative (x, i) vollkommen von seiner eigenen Identität abstrahieren und sich sowohl die objektiven Bedingungen als auch die subjektiven Merkmale anverwandeln, die die Position des Individuums i im gesellschaftlichen Zustand x charakterisieren. Während

die Annahme einer vollkommenen Empathie meist nur verwendet wird, um aus dem Konzept einer erweiterten Präferenzordnung ein geschlossenes System ordinaler interpersoneller Nutzenvergleiche abzuleiten (vgl. Arrow 1951, 1977; Sen 1970), benutzt Harsanyi das Sympathiekriterium, um die Informationsgrundlagen einer utilitaristischen SWF zu erzeugen. Sein hypothetisches Entscheidungsmodell operiert nämlich mit der Prämisse, daß das Individuum h in der Lage ist, den $r \cdot n$ Positionen (x, i) auf dem Wege einer vollkommenen Introspektion $r \cdot n$ kardinale und interpersonell völlig vergleichbare Nutzenwerte $U_i(x)$ zuzuordnen, die die subjektiven Präferenzen der Gesellschaftsmitglieder i repräsentieren. Diese Prämisse hat offensichtlich die Konsequenz, daß sich ein Individuum h bei der ethischen Bewertung jeder Alternative x an den n gleich wahrscheinlichen Nutzenniveaus $U_i(x)$ zu orientieren hat. Aus dieser zusätzlichen Spezifizierung des Entscheidungsproblems zieht Harsanyi (1977 a, 1977 c) den Schluß, daß einem rationalen Individuum h keine andere Wahl bleibt, als die vNM-Nutzenwerte $\tilde{U}_h(x, i)$ mit den introspektiven Nutzenwerten $U_i(x)$ gleichzusetzen:

$$\forall\, x \in X, \forall\, i \in N: \tilde{U}_h(x, i) = U_i(x)$$

Folgt man diesem letzten Schritt der Argumentation, so bietet bereits die Sympathiebedingung die Gewähr dafür, daß die ethischen Präferenzen jedes rationalen Individuums h in einer genuin utilitaristischen SWF zum Ausdruck kommen, die die Wohlfahrt der Gesellschaft mit dem durchschnittlichen Nutzenniveau aller Individuen identifiziert:

$$\forall\, x \in X: W_h(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} U_i(x), \; (h = 1, \ldots, n)$$

Da Harsanyi mit der Sympathiebedingung die benötigte interpersonelle Nutzeninformation endogenisiert und überdies mit dem Unparteilichkeitskriterium für eine grundsätzliche Gleichgewichtung der individuellen Interessen sorgt, kann es keinen Zweifel daran geben, daß die substantiellen Prämissen seines Entscheidungsmodells weit über die Postulate a—c hinausgehen. Diese weitreichenden Vorkehrungen erweisen sich jedoch immer noch als zu schwach, um in Verbindung mit den Bayes'schen Rationalitätspostulaten eine äquivalente Fundierung des utilitaristischen Prinzips zu erzeugen. Selbst wenn man nämlich davon ausgeht, daß sich das Individuum h bei der ethischen Bewertung jeder Alternative x ausschließlich auf den Nutzenvektor $(U_1(x), \ldots, U_n(x))$ stützt und jeder Position (x, i) grundsätzlich das gleiche Gewicht einräumt, zwingt das Bayes'sche Rationalitätskonzept per se keineswegs — wie Harsanyi behauptet — dazu, die

vNM-Indikatoren $\tilde{U}_h(x, i)$ in Höhe der introspektiven Nutzenwerte $U_i(x)$ anzusetzen. Dieses spezielle Problem ist auf die allgemeine Tatsache zurückzuführen, daß jede vNM-Nutzenfunktion unweigerlich die Risikoneigung des betreffenden Individuums widerspiegelt (vgl. Arrow 1951; Harsanyi 1987). Unter den hypothetischen Prämissen des Harsanyi-Modells hat diese Eigenschaft des Bayes'schen Rationalitätskonzepts die bemerkenswerte Konsequenz, daß jedes Individuum h beim Ansatz der vNM-Indikatoren $\tilde{U}_h(x, i)$ grundsätzlich nicht nur die einzelnen Nutzenniveaus $U_i(x)$, sondern auch deren Verteilung berücksichtigen wird (vgl. Pattanaik 1968).

Um die begrenzte Reichweite des entscheidungstheoretischen Ansatzes zu verdeutlichen, genügt es, je ein risikofreudiges (j), risikoscheues (k) und risikoneutrales Individuum (l) zu betrachten, die jeweils vor dem Problem stehen, zwei gesellschaftliche Zustände x und y unter den folgenden Annahmen in eine ethische Rangordnung zu bringen:

$$U_1(x) = U_2(x) = \ldots = U_n(x)$$
$$U_1(y) > U_2(y) > \ldots > U_n(y)$$
$$\sum_{i=1}^{n} U_i(x) = \sum_{i=1}^{n} U_i(y)$$

Da der gesellschaftliche Zustand x weder Risiken noch Chancen birgt, gibt es für keines der drei Individuen einen Grund, bei der ethischen Bewertung der ersten Alternative vom Durchschnittsnutzenprinzip abzuweichen:

$$W_h(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{U}_h(x, i) = \frac{1}{n} \cdot \sum_{i=1}^{n} U_i(x), \; (h = j, k, l)$$

Die Tatsache, daß der gesellschaftliche Zustand y zwar einerseits das gleiche durchschnittliche Nutzenniveau erzeugt wie die Alternative x, andererseits aber das Risiko (bzw. die Chance) enthält, diesen Durchschnittswert zu unterschreiten (bzw. zu übertreffen), hat indes die Konsequenz, daß nur das risikoneutrale Individuum l der utilitaristischen Maxime treu bleibt:

$$W_l(y) = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{U}_l(y, i) = \frac{1}{n} \cdot \sum_{i=1}^{n} U_i(y) = W_l(x)$$

Während das Individuum l die unterschiedlichen Verteilungseigenschaften der Vektoren $(U_1(x), \ldots, U_n(x))$ und $(U_1(y), \ldots, U_n(y))$ völlig ignoriert, werden die Individuen j und k die ethische Rangordnung der Alternativen x und y ausschließlich von den resultierenden Nutzenverteilungen abhän-

gig machen. Da ein risikofreudiges (bzw. risikoscheues) Individuum ceteris paribus — d. h.: bei gleicher Summe der $U_i(\cdot)$ — eine ungleiche (bzw. gleiche) Nutzenverteilung bevorzugt, wird das Individuum j (bzw. k) bei der Bewertung der Alternative y in jedem Fall einem nicht-utilitaristischen Kalkül folgen:

$$W_j(y) = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{U}_j(y, i) > \frac{1}{n} \cdot \sum_{i=1}^{n} U_i(y) = W_j(x)$$

$$W_k(y) = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{U}_k(y, i) < \frac{1}{n} \cdot \sum_{i=1}^{n} U_i(y) = W_k(x)$$

Das Beispiel zeigt, daß die Prämissen des Harsanyi-Modells neben dem utilitaristischen Prinzip auch ethische Prinzipien zulassen, die die relative Gewichtung der individuellen Interessen von den Eigenschaften der Nutzenverteilung abhängig machen. Wie breit das Spektrum der mit dem entscheidungstheoretischen Ansatz vereinbaren Sozialen Wohlfahrtsfunktionen tatsächlich ist, wird deutlich, wenn man dem Individuum k (bzw. j) eine unendlich große Risikoaversion (bzw. Risikofreude) unterstellt. Diese extreme Annahme hat nämlich die Konsequenz, daß sich das Individuum k (bzw. j) bei der ethischen Bewertung gesellschaftlicher Zustände am rawlsianischen Maximin-Prinzip (bzw. am Maximax-Prinzip) orientieren wird, das die soziale Wohlfahrt mit dem Nutzenniveau des am schlechtesten (bzw. besten) gestellten Individuums indentifiziert (vgl. Rawls 1971; Arrow 1973; Alexander 1974):

$$W_k(\cdot) = \min_i U_i(\cdot)$$

$$W_j(\cdot) = \max_i U_i(\cdot)$$

Solange der Risikoneigung des bewertenden Individuums keine Beschränkung auferlegt wird, ist der entscheidungstheoretische Ansatz von einer äquivalenten Fundierung des utilitaristischen Prinzips offenbar noch weit entfernt. Um eine utilitaristische SWF als exklusives Resultat einer rationalen, unparteiischen und sympathetischen Entscheidung zu sichern, sind Harsanyis Prämissen um die Annahme zu ergänzen, daß das Konzept einer ethischen Präferenz eine risikoneutrale Bewertung der gesellschaftlichen Zustände erfordert.[4] Diese zusätzliche Prämisse ist — wie zuvor schon das

[4] Verzichtet man auf diese Normierung der individuellen Risikopräferenzen, so steht man vor dem Problem, die divergierenden ethischen Präferenzen der Gesellschaftsmitglieder zu einer allgemein akzeptierbaren SWF zu aggregieren. Zu unterschiedlichen Lösungen dieses Aggregationsproblems vgl. Pattanaik (1968), Mueller/Tollison/Willett (1974), Svensson (1989) und Segerstrom (1990).

Symmetrie-Postulat — als starke ethische Bedingung zu interpretieren, die die relative Gewichtung der individuellen Interessen im Sinne der utilitaristischen Ethik festlegt. Trotz der probabilistischen Modellierung des Unparteilichkeitskriteriums bietet nämlich erst eine risikoneutrale Bewertung der Alternativen die Gewähr dafür, daß jedes Individuum h allen individuellen Nutzenwerten — völlig unabhängig von ihrer Verteilung — das gleiche Gewicht beimißt.

5. Schlußbemerkungen

Harsanyis Anspruch, den Utilitarismus als einzig rationale und zugleich allgemein akzeptierbare Moral auszuweisen, wird weder mit dem axiomatischen noch mit dem entscheidungstheoretischen Ansatz eingelöst. Solange das Bayes'sche Rationalitätskonzept lediglich mit schwachen ethischen Bedingungen verknüpft wird, kann es die Last der Deduktion nicht tragen. Da beide Fundierungsansätze die relative Gewichtung der individuellen Interessen offen lassen, ist jeweils ein starkes ethisches Postulat vonnöten, um alle Rivalen des utilitaristischen Prinzips auszuschließen. Harsanyis Unternehmung sind daher insofern klare Grenzen gesetzt, als entweder das Ziel einer äquivalenten Fundierung oder das Kriterium der allgemeinen Akzeptierbarkeit geopfert werden muß.

Nun hat sich zwar gezeigt, daß das Bayes'sche Rationalitätskonzept keineswegs die von Harsanyi erhoffte Leistung erbringt. Zugleich ist aber auch deutlich geworden, daß es bereits in Verbindung mit einer unumstrittenen ethischen Prämisse drastische Konsequenzen nach sich zieht. So genügt die Anwendung der Bayes'schen Rationalitätspostulate auf individuelle und kollektive Präferenzen, um das umfassende Konzept einer individualistischen SWF auf eine spezielle Klasse ethischer Prinzipien zu reduzieren, die bei der Bewertung gesellschaftlicher Zustände alle Informationen über die Nutzenverteilung ausblenden. Angesichts dieses Ergebnisses wäre im Rahmen einer fundamentalen Kritik der Harsanyischen Unternehmung die Frage zu stellen, ob das Bayes'sche Rationalitätskonzept — gemessen an unseren moralischen Überzeugungen — nicht schon zuviel leistet. Sollen wir uns, so wäre dann zu fragen, an ein formales Rationalitätskriterium, das wir — isoliert betrachtet — uneingeschränkt akzeptieren mögen, auch dann noch gebunden fühlen, wenn es uns zwingt, substantielle ethische Probleme zu ignorieren?

Literatur

Alexander, S. S.: Social evaluation through notional choice. *Quarterly Journal of Economics* 88 (1974), S. 597–624.

Arrow, K. J.: *Social Choice and Individual Values.* New York 1951. 2. Aufl. 1963.

Arrow, K. J.: Some ordinalist-utilitarian notes on Rawls's theory of justice. *Journal of Philosophy* 70 (1973), S. 245–263.

Arrow, K. J.: Extended sympathy and the possibility of social choice. *American Economic Review: Papers and Proceedings* 67 (1977), S. 219–225.

Bergson, A.: A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52 (1938), S. 310–334.

Border, K. C.: More on Harsanyi's utilitarian cardinal welfare theorem. *Social Choice and Welfare* 1 (1985), S. 279–281.

Brock, H. W.: The problem of ‚utility weights' in group preference aggregation. *Operations Research* 28 (1980), S. 176–187.

Broome, J.: Uncertainty and fairness. *Economic Journal* 94 (1984), S. 624–632.

Butts, R. E./Hintikka, J. (Hg.): *Foundational Problems in the Special Sciences.* Dordrecht 1977.

Camacho, A./Sonstelie, J.: Cardinal welfare, individualistic ethics, and interpersonal comparisons of utilities: A note. *Journal of Political Economy* 82 (1974), S. 607–611.

Coulhon, T./Mongin, P.: Social choice theory in the case of von Neumann-Morgenstern utilities. *Social Choice and Welfare* 6 (1989), S. 175–187.

Deschamps, R./Gevers, L.: Separability, risk-bearing and social welfare judgements. *European Economic Review* 10 (1977), S. 77–94.

Diamond, P. A.: Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy* 75 (1967), S. 765–766.

Domotor, Z.: Ordered sum and tensor product of linear utility structures. *Theory and Decision* 11 (1979), S. 375–399.

Edgeworth, F. Y.: *Mathematical Psychics.* London 1881.

Fishburn, P. C.: On Harsanyi's utilitarian cardinal welfare theorem. *Theory and Decision* 17 (1984), S. 21–28.

Fleming, M.: A cardinal concept of welfare. *Quarterly Journal of Economics* 66 (1952), S. 366–384.

Gauthier, D.: On the refutation of utilitarianism. In: *The Limits of Utilitarianism.* Hg. von H. B. Miller/W. H. Williams. Minneapolis 1982, S. 144–163.

Hammond, P. J.: Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty. In: *Social Choice and Welfare.* Hg. von P. K. Pattanaik/M. Salles. Amsterdam 1983, S. 175–205.

Hammond, P. J.: On reconciling Arrow's theory of social choice with Harsanyi's fundamental utilitarianism. In: *Arrow and the Foundations of the Theory of Economic Policy.* Hg. von G. R. Feiwel. New York 1987, S. 179–221.

Harsanyi, J. C.: Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy* 61 (1953), S. 434–435.

Harsanyi, J. C.: Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63 (1955), S. 309–321.

Harsanyi, J. C.: Ethics in terms of hypothetical imperatives. *Mind* 47 (1958), S. 305–316.

Harsanyi, J. C.: Nonlinear social welfare functions: Do welfare economists have a special exemption from Bayesian rationality? *Theory and Decision* 6 (1975), S. 311–332.

Harsanyi, J. C.: Morality and the theory of rational behavior. *Social Research* 44 (1977), S. 623–656 (= 1977 a).

Harsanyi, J. C.: Nonlinear social welfare functions: A rejoinder to Professor Sen. In: *Butts/Hintikka* (1977), S. 293–296 (= 1977 b).

Harsanyi, J. C.: *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge 1977 (= 1977 c).

Harsanyi, J. C.: Bayesian decision theory and utilitarian ethics. *American Economic Review: Papers and Proceedings* 68 (1978), S. 223–228.

Harsanyi, J. C.: Von Neumann-Morgenstern utilities, risk taking, and welfare. In: *Arrow and the Ascent of Modern Economic Theory*. Hg. von G. R. Feiwel. London 1987, S. 545–558.

Herstein, I. N./Milnor, J.: An axiomatic approach to measurable utility. *Econometrica* 21 (1953), S. 291–297.

Jeffrey, R. C.: On interpersonal utility theory. *Journal of Philosophy* 68 (1971), S. 647–656.

Luce, R. D./Raiffa, H.: *Games and Decisions*. New York 1957.

Marschak, J.: Rational behavior, uncertain prospects, and measurable utility. *Econometrica* 18 (1950), S. 111–141.

Marshall, A.: *Principles of Economics*. London 1890, 9. Aufl. 1961.

McClennen, E. F.: Constitutional choice: Rawls versus Harsanyi. In: *Pitt* (1981), S. 93–109.

Mueller, D. C./Tollison, R. D./Willett, T. D.: The utilitarian contract: A generalization of Rawls' theory of justice. *Theory and Decision* 4 (1974), S. 345–367.

Neumann, J. von/Morgenstern, O.: *Theory of Games and Economic Behavior*. Princeton 1944, 3. Aufl. 1953.

Ng, Y.-K.: Bentham or Bergson? Finite sensibility, utility functions and social welfare functions. *Review of Economic Studies* 42 (1975), S. 545–569.

Nunan, R.: Harsanyi vs. Sen: Does social welfare weigh subjective preferences? *Journal of Philosophy* 78 (1981), S. 586–600.

Pareto, V.: *Cours d'Economie Politique*. Lausanne 1897.

Pattanaik, P. K.: Risk, impersonality, and the social welfare function. *Journal of Political Economy* 76 (1968), S. 1152–1169.

Pigou, A. C.: *The Economics of Welfare*. London 1920, 4. Aufl. 1932.

Pitt, J. C. (Hg.): *Philosophy in Economics*. Dordrecht 1981.

Rawls, J.: *A Theory of Justice*. Cambridge, Mass. 1971.

Resnik, M. D.: A restriction on a theorem of Harsanyi. *Theory and Decision* 15 (1983), S. 309–320.

Robbins, L.: *An Essay on the Nature and Significance of Economic Science*. London 1932, 2. Aufl. 1935.

Samuelson, P. A.: *Foundations of Economic Analysis*. Cambridge, Mass. 1947.

Samuelson, P. A.: Probability, utility, and the independence axiom. *Econometrica* 20 (1952), S. 670–678.

Savage, L. J.: *The Foundations of Statistics*. New York 1954, 2. Aufl. 1972.

Schmidt, J.: *Gerechtigkeit, Wohlfahrt und Rationalität*. Freiburg 1991.

Segerstrom, P. S.: Moral efficiency. A new criterion for social choice. *Social Choice and Welfare* 7 (1990), S. 109–129.

Selinger, S.: Harsanyi's aggregation theorem without selfish preferences. *Theory and Decision* 20 (1986), S. 53–62.

Sen, A.: *Collective Choice and Social Welfare*. San Francisco 1970.

Sen, A.: *On Economic Inequality*. Oxford 1973.

Sen, A.: Welfare inequalities and Rawlsian axiomatics. *Theory and Decision* 7 (1976), S. 243–262.

Sen, A.: Non-linear social welfare functions: A reply to Prof. Harsanyi. In: *Butts/Hintikka* (1977), S. 297–302.

Strasnick, S.: Neo-utilitarian ethics and the ordinal representation assumption. In: *Pitt* (1981), S. 63–92.

Suppes, P.: Some formal models of grading principles. *Synthese* 16 (1966), S. 284–306.

Svensson, L.-G.: Fairness, the veil of ignorance and social choice. *Social Choice and Welfare* 6 (1989), S. 1–17.

Trapp, R. W.: *‚Nicht-klassischer' Utilitarismus*. Frankfurt a. M. 1988.

Vickrey, W.: Measuring marginal utility by reactions to risk. *Econometrica* 13 (1945), S. 319–333.

Vickrey, W.: Utility, strategy, and social decision rules. *Quarterly Journal of Economics* 74 (1960), S. 507–535.