hwp
w&u

UNIVERSITÄT HOHENHEIM

Hohenheimer Working Papers
Wirtschafts- & Unternehmensethik

Lehrstuhl 560 D

No. 9 (2009)

Utilitarismus und Gerechtigkeit. Eine Analyse des equiprobability model von John C. Harsanyi

Matthias Strifler

Universität Hohenheim | Lehrstuhl 560 D
www.theology-ethics.uni-hohenheim.de

Abstract

This paper attempts to track the notion of justice in utilitarianism, especially in Harsanyi's equiprobability model. After glancing briefly at the roots the focus is on the assumptions of the model with which the model itself stands or falls. It is argued, that on the one hand similarity of humans is an already proven fact and on the other a logical basis for any ethical theory. By fencing off the model from other utilitarian theories, preference and rule utilitarianism are proposed as the only serious utilitarian ethical theories. The idea of justice is discussed in confrontation to Rawls' theory of justice, where the main focus is on the voting modalities in the original position. In contrast to the prevailing doctrine the paper concludes, that a slightly modified version of Harsanyi's equiprobability model is superior to Rawls' theory of justice and should therefore be given more attention in ethics as well as in political philosophy.

1. Einleitung

Die heutzutage herrschende Lehrmeinung in der Ethik behandelt die Theorie der Gerechtigkeit von RAWLS als den modernen Stand der Dinge und sieht damit auch den, vor allem von HARSANYI fortentwickelten Utilitarismus als inferiore Alternative dazu an. In ihm sei es möglich, „daß ein einzelner oder eine Minderheit für das Glück der Mehrheit aufgeopfert wird“ (PIEPER 2007, S. 287), was als ethisch unzulässig anzusehen sei. So stellt auch HOMANN (1988, S. 212) die Frage, „ob der Utilitarismus bei einem grundsätzlich opportunistischen Ansatzpunkt die Verbindlichkeit denken kann, die man gemäß der intuitiven moralischen Urteile von aller Moral erwartet und die allein jeden einzelnen vor Diskriminierung sicherzustellen vermag.“ So zeigt sich, dass sich in der politischen Philosophie, die ja der Ethik als Basiswissenschaft bedarf, der Utilitarismus als möglicher Gegenentwurf zur Theorie der Gerechtigkeit von RAWLS zum Teil nicht einmal Erwähnung findet (vgl. BECKER, SCHMIDT, ZINTL 2006).

Im Gegensatz zu dieser Lehrmeinung versucht dieses Paper Argumente dafür zu finden, dass es möglicherweise verkehrt ist, den Utilitarismus nach der Spielart HARSANYIs als eine inferiore Alternative zur RAWLSschen Gerechtigkeitstheorie anzusehen. Vielmehr entpuppt sich nach einer intensiven Auseinandersetzung mit dem Gerechtigkeitsentwurf von HARSANYI dieser als dem RAWLSschen überlegen.

2. Die ethische Theorie von JOHN C. HARSANYI

In einem ersten Schritt wird kurz auf die Wurzeln der Theorie eingegangen sowie auf ihre Einbettung in die Entscheidungstheorie. In einem zweiten Schritt wird der Kern der Theorie, nämlich das equiprobability model, eingehend dargestellt.

2.1 Herkunft und Wurzeln

Die ethische Theorie von JOHN C. HARSANYI basiert zum einen, so sagt er selbst (HARSANYI 1977a, S. 623 f.), auf drei moralphilosophischen Denktraditionen. Die erste führt er auf ADAM

SMITH zurück, „who equated the moral point of view with that of an impartial but sympathetic spectator“ (HARSANYI 1977a, S. 623). Soll eine Situation nach ihrer moralischen Richtigkeit bewertet werden, so muss dies vom Standpunkt eines „impartial spectator“ (SMITH 1790, S.239) aus getan werden, also eines unparteiischen Zuschauers.

Die zweite Denktradition beruht auf IMMANUEL KANT. So stammt das Prinzip des Universalismus von KANT, das als eine der fünf Eigenschaften des Utilitarismus eine wichtige Rolle spielt. Drittens führt er die utilitaristischen Vordenker BENTHAM, J. S. MILL, SIDGWICK und EDGEWORTH an, „which made maximization of social utility the basic criterion of morality – social utility being defined either as the sum, or the arithmetic mean, of the utiltiy levels of all individuals in the society“ (HARSANYI 1977a, S. 624).

Zum anderen basieren seine Überlegungen auf der bayesianischen Theorie rationalen Verhaltens. Genauer, HARSANYI behauptet „that moral behavior itself is a special form of rational behavior“ (HARSANYI 1977a, S. 625). Diesen Umstand führt er darauf zurück, dass die moderne Entscheidungtheorie die Ethik zu einem „organic part of the general theory of rational behavior” gemacht hat. Innerhalb dieser 'allgemeinen Theorie rationalen Verhaltens' unterscheidet er die Entscheidungstheorie, die Spieltheorie und die Ethik voneinander. Wobei er die Ethik als eine „theory of rational behavior in the service of the common interests of society as a whole“ (HARSANYI 1977a, S. 629) definiert. Als Begründung, warum er diese drei Ansätze unter einer allgemeinen Theorie zusammenfasst, führt er drei Argumente an.

Erstens benutzen alle drei Disziplinen dieselbe Vorgehensweise. So wird zuerst eine „primary definition of rationality” vorgenommen, von der dann eine „secondary definition of rationality” abgeleitet wird. Wobei die erste Definition in der jeweiligen Disziplin „some set of axioms or [...] a constructive decision model” beinhaltet. Die Entdeckung einer passenden „primary definition is always essentially a philosophical – that is, a conceptual – problem“ (HARSANYI 1977a, S. 630). Des Weiteren ist die Ableitung der „secondary definition“ immer ein mathematisches Problem.

Zweitens sind die Axiome, die von den drei Disziplinen benutzt werden, mathematisch ähnlich (vgl. HARSANYI 1977a, S.630). In allen spielt die Nutzenmaximierung, wenn auch auf unterschiedliche Weise, die zentrale Rolle.

Als dritten Grund führt HARSANYI an, dass es möglich sei, einige grundlegende Probleme der Spieltheorie und der Ethik auf entscheidungstheoretische Probleme zu reduzieren. Diesen

Grund sieht er als „the most important link among the three disciplines“ (HARSANYI 1977a, S. 630f).

2.2 Das equiprobability model

Die Grundlage dieser Theorie bildet ein „model for moral value judgments“ (HARSANYI 1977a, S. 631). Dabei unterscheidet HARSANYI zwischen zwei Arten von Präferenzen: sozialen und individuellen Präferenzen (HARSANYI 1955, S. 312) oder ethischen und subjektiven Präferenzen (HARSANYI 1955, S. 315) beziehungsweise moralischen und persönlichen Präferenzen (HARSANYI 1977a, S. 631 und ausführlich S. 635). So ist jede Entscheidung, die nach moralischen Präferenzen gefällt wird („moral value judgment”), „a judgment of preference, but it is a judgment of preference of a very special kind“.

Beispielhaft unterscheidet der Autor die Präferenzen wie folgt: Bevorzugt eine Person eine Einkommensverteilung, die die Armen hinreichend berücksichtigt, aus dem einzigen Grund, weil sie selbst arm ist, so liegt eine persönliche Präferenz vor. Hat eine Person jedoch diese Präferenz obwohl sie reich ist, oder bevorzugt sie diese Einkommensverteilung unabhängig davon, dass sie arm ist, so liegt eine soziale/moralische Präferenz vor (vgl. HARSANYI 1953, S. 434). Mit anderen Worten sind Entscheidungen, die nach moralischen Präferenzen gefällt werden „nonegoistic impersonal judgments of preference“ (HARSANYI 1953, S. 434).

Deutlicher wird der Unterschied bei folgendem Gedankenexperiment: Angenommen, es versammeln sich alle Individuen am Nullpunkt der Geschichte und diskutieren, welche Regeln sie sich für das Zusammenleben geben wollen[1]. Genauer, Sie diskutieren, welches Wirtschaftssystem, beziehungsweise welches Gesellschaftssystem sie sich geben wollen. Dabei stehen hier zwei beispielhaft zur Debatte: Kapitalismus und Sozialismus. Wobei im Kapitalismus die Einkommensunterschiede enorm sind (die schlechteste Position hat ein Einkommen von 0) und im Sozialismus eine Gleichverteilung herrscht. Dabei sei der Nutzen eine Funktion des Einkommens.

Hat nun jemand eine Vorliebe für ein kapitalistisches System, obwohl er im Vorhinein nicht weiß, welche genaue Position er in diesem System innehaben wird, und es sei weiter

angenommen, diese Person hätte zwischen den zwei Systemen unter der Voraussetzung zu wählen, dass sie dieselbe Wahrscheinlichkeit hätte jede mögliche Position zu besetzen, so ist in just diesem Moment, so argumentiert HARSANYI, die Wahl unabhängig von „morally irrelevant selfish considerations", und „would now become a genuine moral value judgment" (HARSANYI 1977a, S. 632). „Moralisch ist ein Präferenzurteil also nur dann, wenn es lediglich die allgemeinen Vor- und Nachteile im Blick hat" (HOMANN 1988, S. 221).

Natürlich ist die beschriebene fiktive Situation, dass sich alle Individuen am Nullpunkt der Geschichte treffen und die beschriebene Person tatsächlich nichts über ihre (jetzige) oder zukünftige gesellschaftliche Position weiß, keine notwendige Vorraussetzung, um eine solche Entscheidung herbeiführen zu können. Es ist aber (für die Praxis) notwendig, dass die Person bei der Entscheidung - so gut es geht - versucht, jeglichen moralisch irrelevanten Part von Information zu ignorieren. HARSANYI „denkt, dass alle Menschen – zumindest grundsätzlich – zwischen den beiden Arten von Präferenzurteilen zu unterscheiden vermögen und dass sie in den konkreten Fällen, die ihnen in ihrem Leben begegnen, auch bis zu einem gewissen Grad in der Lage sind, von ihren persönlichen Präferenzen zu abstrahieren, wenn ein moralisches Urteil gefordert ist" (HOMANN 1988, S. 221).

Die Annahme, dass nun die Person dieselbe Wahrscheinlichkeit hat, jeder gesellschaftlichen Position zugewiesen zu werden, nennt der Autor das „equiprobability postulate", den gesamten Vorgang hingegen, der auf diesem Postulat aufbaut, nennt er das „equiprobability model of moral value judgments" (HARSANYI 1977a, S. 632). Das equiprobability model bildet in diesem Zusammenhang die „primary definition of rationality".

Kurz gefasst beinhaltet das Modell also zwei zentrale Faktoren: erstens weiß der Entscheidungsträger nicht, welche Position er einnehmen wird, zweitens ist es für ihn gleich wahrscheinlich, in jede Position zu gelangen.

Um nun zur „secondary definition or rationality" zu gelangen und die Implikationen des Modells herauszuarbeiten, wählt HARSANYI eine formale Herangehensweise, da es sich um eine mathematische Problemstellung handelt. Angenommen, die Gesellschaft, die wir betrachten, besteht aus *n* Individuen, die wie folgt durchnummeriert sind: 1, 2, ..., n. Dabei erfolgt die Nummerierung nach der jeweiligen Position, die das Individuum einnehmen soll, so

---

[1] „Der Sache nach ist darin (s. HARSANYI 1953/55) Rawls' Schleier des Nichtwissens antizipiert" (HOMANN 1988 S. 222).

belegt Individuum 1 die erste Position, Individuum 2 die zweite, ... und das n-te Indivuduum die n-te Position in der gegebenen Gesellschaft. Die in der jeweiligen Position 1, 2, ..., n zu genießende Nutzenniveaus sind durch U_1, U_2, ... U_n gegeben. Die Person, die nun eine moralische Entscheidung über das Verdienst von Kapitalismus und Sozialismus treffen will, wird Individuum i genannt.

Durch das „Gleichwahrscheinlichkeitspostulat“ („equiprobablity postulate“) wird Individuum i sich so verhalten, als ob es dieselbe Wahrscheinlichkeit 1/n jede einzelne gesellschaftliche Position und damit verbunden jedes entsprechende Nutzenniveau U_1, U_2, ..., U_n zu erhalten, zuschreibt (vgl. HARSANYI 1977a, S. 632). In der beschriebenen Situation wählt, nach der bayesianischen Entscheidungstheorie, ein rationales Individuum die Gesellschaftsform, die dessen erwarteten Nutzen maximiert, das wäre das arithmetische Mittel aller individuellen Nutzenniveaus in der Gesellschaft:

$$W_i = \frac{1}{n} \sum_{j=1}^{n} U_j$$

Womit die „secondary definition of rationality“ getroffen wäre. Ein Individuum wählt also immer den Durchschnittsnutzen als seine ‚soziale Wohlfahrtsfunktion‘ („social welfare function“) (vgl. Harsyanyi 1977a, S. 633). Anders ausgedrückt, das Individuum nutzt das „principle of expected-utility maximization as his decision rule“ (HARSANYI 1975, S. 598). Es entscheidet sich also für die Gesellschaftsform, die dem einzelnen Gesellschaftsmitglied den höchsten Durchschnittsnutzen bringt. Allgemeiner, das Individuum bewertet jede Gesellschaftsform („every possible social system, institutional framework, social practice, etc.“) hinsichtlich des Durchschnittsnutzens, den sie bietet.

Wichtig ist dabei, dass das Individuum, das nach dieser Entscheidungsregel vorgeht, seinen „individuellen – erwarteten – Nutzen” maximiert, dieser aber identisch ist mit dem „sozialen – erwarteten – Nutzen, [...] weil durch die Gleichwahrscheinlichkeitsannahme in Verbindung mit dem Schleier des Nichtwissens von allen persönlichen Präferenzen abstrahiert ist“ (HOMANN, 1988, S. 222 mit Verweis auf HARSANYI 1976, S. 45).

Dieses Bewertungskriterium wird aus diesem Grund das Durchschnittsnutzenprinzip („principle of average utility“) genannt (vgl. HARSANYI 1975, S. 598). Damit gehört das equiprobability model, das eine ethische Theorie begründet, die auf dem Prinzip des Durchschnittsnutzens fundiert, der utilitaristischen Denktradition an (vgl. HARSANYI 1957, S. 598). Genauer gesagt gehört das Modell der Kategorie des Durchschnittsnutzenutilitarismus an. Ausschlaggebend für die moralische Richtigkeit einer Handlung ist die Maximierung des Durchschnittsnutzens, des Nutzens pro Kopf der Gesellschaft. Dieser Durchschnittsnutzenutilitarismus ist vom Nutzensummen-Utilitarismus abzugrenzen. Diese und weitere Unterscheidungen werden in Kapitel 4 erörtert.

Um zu der ‚sozialen Wohlfahrtsfunktion‘ in dieser mathematischen Form zu gelangen, genügen die oben getätigten logischen Überlegungen, also die Ableitung der „secondary definition of rationality“ (vgl. HARSANYI 1955, S. 316). Darüber hinaus findet sich ein weiterer Ansatz um von den Präferenzen zu der Nutzenfunktion zu gelangen, wobei die Präferenzen und die Nutzenfunktion verschiedene Voraussetzungen (Postulate) erfüllen müssen (vgl. HARSANYI 1955, S. 310 ff. und HARSANYI 1977a, S. 636 f). Dieser Ansatz wurde von HARSANYI entwickelt, da die Postulate es ermöglichten, „to exclude nonethical sujective preferences from social welfare functions“ (HARSANYI 1955, S. 316). Zu diesem Zeitpunkt (1953/1955) hatte HARSANYI seine ethische Theorie, besser gesagt die „primary definition of rationality“, die ja ein ethisch-philosophisch, ein konzeptionelles Problem darstellt, noch nicht explizit aufgestellt, so dass er auf die Postulate zurückgriff: „I shall set out the philosophic case for this ethical theory in a forthcoming publication.“ (HARSANYI 1955, S. 316 Fn.19).

Trotzdem finden sich schon in seinem Aufsatz von 1955 die zentralen Bestandteile des equiprobability model: „I have argued elsewhere that an individual's preferences satisfy this requirement of impersonalitiy if they indicate what social situation he would choose if he did not know what his personal position would be in the new situation choosen [...] but rather had an equal chance of obtainig any of the social positions“ (HARSANYI 1955, S. 316). Allein in diesem einen Satz finden sich schon die zwei Voraussetzungen des Modells: Zum einen kennt das Individuum nicht seine zukünftige Position, zum anderen besteht dieselbe Wahrscheinlichkeit, in jede beliebige Position gesetzt zu werden.

Mit den Worten „I have argued elsewhere“ verweist der Autor auf eine Arbeit aus dem Jahre 1953: „Now, a value judgment on the distribution of income would show the required impersonalitiy to the highest degree if the person who made this judgment had to choose a particular income distribution in complete ignorance of what his own relative position [...] would be within the system chosen. This would be the case if he had exactly the same chance of obtaining the first position [...] or the second or the third, etc., up to the last position [...] available in that scheme.“ (HARSANYI 1953, S. 434 f.) Auch hier finden sich schon beide grundlegenden Gedanken zum equiprobability model. Letztendlich lässt sich sagen, dass HARSANYI sein Modell schon 1953 entworfen hat, obwohl er es erst erheblich später als ein solches explizit bezeichnet (s. HARSANYI 1977a, S. 634; siehe auch HOMANN 1988, S. 221 f.).

Aus diesem Grund wird auch auf die Herleitung der ‚sozialen Wohlfahrtsfunktion‘ anhand der Postulate verzichtet, da diese, um zu ihr zu gelangen, nicht notwendig sind.[2] Das Gedankenxperiment, beziehungsweise die ‚primary definition of rationality‘ reicht nicht nur vollkommen aus, im Gegenteil, es geht noch weit über die alternative, axiomatische Rechfertigung hinaus. Diese basiert auf viel schwächeren Annahmen, zeigt jedoch deutlicher den engen Bezug zur ‚allgemeinen Theorie rationalen Verhaltens‘ (vgl. HARSANYI 1977a, S. 636).

Wie zu Beginn des Kapitels schon angedeutet, kann das Modell durch zwei unterschiedliche Arten von Präferenzen beschrieben werden. Zum einen die persönlichen Präferenzen, die ihren Niederschlag in der Nutzenfunktion U_i finden und zum anderen moralische Präferenzen, die ihren Niederschlag in der sozialen Wohlfahrtsfunktion W_i finden (vgl. im folgenden HARSANYI 1977a, S. 635). Verfolgt eine Person ihre persönlichen Präferenzen, so gibt sie den eigenen Interessen sowie den Interessen von Freunden und Familie eine viel höhere Gewichtung als den Interessen von irgendwelchen Fremden. Sollte eine Person aber danach streben, eine moralische Entscheidung zu fällen, so wird sie versuchen eine „special impersonal and impartial attitude, that is, a moral attitude“ (HARSANYI 1977a, S. 635) einzunehmen und damit den Interessen aller Individuen dasselbe Gewicht zugestehen.

Damit sieht HARSANYI sein Modell als „a modern restatement of ADAM SMITH’s theory of an impartially sympathetic observer. His impartiality requirement corresponds to my equiprobability postulate, whereas his sympathy requirement corresponds to my assumtion

[2] Eine axiomatische Herleitung findet sich in HARSANYI (1955), S. 310 ff. und (1977a), S. 636 f.

that individual i will make his choice in terms of interpersonal utility comparisons based on empathy [...]" (HARSANYI 1977a, S. 635).

3. Zentrale Voraussetzungen/Annahmen des Modells

Wie gezeigt wurde, wird der soziale Nutzen als der durchschnittliche individuelle Nutzen definiert. (Vgl. im folgenden HARSANYI, 1977a S. 633) Damit diese Schlussfolgerung überhaupt Sinn macht, müssen zwei Voraussetzungen erfüllt sein. Zum einen muss es zulässig sein, die verschiedenen Nutzenniveaus zu addieren und den Durchschnitt zu bilden, zum anderen muss es möglich sein, taugliche interpersonelle Nutzenvergleiche anzustellen. Diese beiden Voraussetzungen sollen im Folgenden näher betrachtet werden.

3.1 Kardinale Nutzenfunktionen

Damit die Rechenschritte, die die soziale Wohlfahrtsfunktion enthält, zulässig sind, muss eine kardinal skalierte Nutzenfunktion vorliegen. Ausgehend von der Wohlfahrtsökonomik sowie der Theorie von Entscheidungen unter Risiko, versuchte HARSANYI (1953) die in der jeweiligen Disziplin unterschiedlichen Konzepte kardinaler Nutzenfunktionen auf einen Nenner zu bringen (jedoch ohne dem Ziel, die soziale Wohlfahrtsfunktion zu rechtfertigen, wie man sehen wird).

In der Wohlfahrtsökonomik wird davon ausgegangen, dass der Grenznutzen mit steigendem Einkommen abnimmt (GOSSENsche Gesetze), bei Entscheidungen unter Risiko wird im Gegensatz (zumindest über einen weiten Verlauf) ein zunehmender Grenznutzen angenommen (vgl. HARSANYI 1953, S. 434). Dies ist nach HARSANYI jedoch nicht entscheidend (vgl. im Folgenden HARSANYI 1953, S. 434 f). Betrachtet man „judgments concerning social welfare" also „nonegoistic impersonal judgments of preference", so zeigt dieses Werturteil „the required impersonalitiy to the highest degree if the person who made this judgment had to choose a particular income distribution in complete ignorance of what his own relative position [...] would be within the system chosen." Was genau dann der Fall ist, wenn die Person dieselbe Wahrscheinlichkeit hat, jede mögliche Position einzunehmen.

„This choice in that hypothetical case would be a clear instance of a ‚choice involving risk‘.“ Daraus schlussfolgert HARSANYI, „that the cardinal utility ‚maximized‘ in value judgments concerning social welfare and the cardinal utility maximized in choices involving risk may be regarded as being fundamentally based upon the same principle.“ Den jeweils unterschiedlichen Grenznutzen erklärt HARSANYI damit, dass in einem Spiel faire Voraussetzungen gelten, dagegen aber keine Chancengleichheit in der Gesellschaft („social game“) besteht (vgl. HARSANYI 1953, S. 435).

Wie in Kapitel 2 beschrieben, hatte HARSANYI 1953 noch nicht das equiprobability model explizit aufgestellt. Offensichtlich war es Ziel seines Aufsatzes (von 1953), kein Modell für ethische Entscheidungen aufzustellen, sondern die Ähnlichkeit bzw. Gleichheit kardinaler Nutzenfunktionen in der Wohlfahrtsökonomik und der Theorie von Entscheidungen unter Risiko zu beweisen. Denn genau dies macht er, und als Beweis dafür entwirft er die Vorform des equiprobability models.

Zwei Jahre später legt HARSANYI dazu einen formalen Beweis anhand mehrerer Postulate vor und gelangt zu einer sozialen Wohlfahrtsfunktion, die der oben beschriebenen, endgültigen sehr ähnelt (vgl. HARSANYI 1955, S. 314). Nach der expliziten Aufstellung des equiprobability model (1975 bzw. 1977a), sah sich HARSANYI dann natürlich erneut gezwungen, den Gebrauch von kardinalen Nutzenfunktionen zu rechtfertigen. Dies versuchte er anhand der VON-NEUMANN-MORGENSTERN-(vNM)-Nutzenfunktionen. vNM-Nutzenfunktionen drücken die Einstellungen von Personen gegenüber der Übernahme von Risiko aus, aber sie versuchen auch, diese Einstellung, durch die Wichtigkeit, die Personen einem möglichen Gewinn zuschreiben, zu erklären (vgl. HARSANYI 1977a ,S. 642). „(..) the primary task of a vNM utility function is *not* to express a given individual's attitudes toward risk taking; rather, it is to indicate how much utility, i.e., how much subjective *importance*, he assigns to various goals“ (HARSANYI 1955, S. 600).

Angenommen ein Individuum i kauft sich einen Lotterieschein für 5 Dollar, das ihm einen Gewinn von 1.000 Dollar bei einer Gewinnwahrscheinlichkeit von 1/1000 garantiert (vgl. im Folgenden HARSANYI 1979, S. 299 f.). Wie lässt sich erklären, dass es bei so schlechten Gewinnchancen trotzdem spielt? Die Begründung sieht HARSANYI darin, dass das Individuum einem möglichen Gewinn von 1.000 Dollar eine hohe relative Wichtigkeit zuschreibt, gegenüber den 5 Dollar, die es zu investieren hat. Folglich misst eine vNM-Nutzenfunktion die

relative Wichtigkeit, die eine Person einem möglichen Gewinn beimisst anhand der Präferenzen für unterschiedliche Lotterien.[3]

„Anschließend kann die kardinale Messung auch auf Entscheidungen unter Sicherheit übertragen werden“ (HOMANN 1988, S. 223). Allgemein gesprochen heißt das, dass vNM-Nutzenfunktionen es ermöglichen, eine soziale Wohlfahrtsfunktion aufzustellen, die die relative Wichtigkeit, die die Individuen ihren verschiedenen Wünschen zuschreiben berücksichtigt, sowie die relative Intensität dieser Präferenzen. „Consequently, vNM utility functions have a completely legitimate place in ethics because they express the subjective importance people attach to their various needs and interests“ (HARSANYI 1975, S. 600). „The use of vNM utility functions implies that, [...], our social welfare function will assign higher social priorities to those objectives to which the individual members of society themselves assign high personal priorities“ (HARSANYI 1979, S. 300).[4]

Die Annahme kardinaler Nutzenfunktionen bzw. die Verwendung der vNM-Nutzenfunktion wurden vor allem von JOHN RAWLS kritisiert. So argumentierte RAWLS (1975), wie auch Arrow (1963), dass vNM-Nutzenfunktionen „auf Entscheidungen zwischen risikobehafteten Möglichkeiten beruhen” (RAWLS 1975, S. 191) und keine ethische Aussagekraft haben, was HARSANYI dazu bewog, seine Beweisführung immer wieder neu zu überarbeiten (siehe oben).

Seit einigen Jahren setzt sich, entgegen dem ordinalen Diktum der letzten Jahrzehnte, die Auffassung, Nutzen kardinal messen zu können, auch in wirschaftswissenschaftlichen Teilgebieten wie z.B. der Glücksökonomie durch (vgl. Ruckriegel 2007). Sie begreift Glück als Glücksgefühl oder Lebenszufriedenheit und richtet ihre Bemühungen danach aus, dieses zu steigern. Die Glücksforschung findet damit ihre philosophischen Wurzeln im Utilitarismus und gründet sich als Kritik an der Mainstreamökonomie, die nur auf materielle Güter und Größen wie Wirtschaftswachstum oder Pro-Kopf-Einkommen abstellt.

---

[3] Für einen formalen Beweis siehe HARSANYI (1979), S. 296 f. Dort zeigt sich, dass sich anhand der bayesianischen Entscheidungstheorie (durch das Durchführen einer Elementarlotterie) eine kardinale Nutzenfunktion (vNM-Nutzenfunktion) definieren lässt.

[4] Den Umstand, dass es unterschiedlich geartete vNM-Nutzenfunktionen gibt, sprich mit abnehmendem und zunehmendem Grenznutzen, erklärt HARSANYI nun anhand von Komplementen und Substituten. Aus Platzgründen wird hier nur darauf verwiesen. Siehe HARSANYI (1979), S. 298 f..

Wie bereits erwähnt, ist die kardinale Nutzenmessung für die Haltbarkeit der sozialen Wohlfahrtsfunktion unverzichtbar, so dass man sich von der Glücksforschung eine weitere Fundierung oder Bestätigung der Richtigkeit des equiprobability-models erwarten darf.

3.2 Interpersonelle Nutzenvergleiche („interpersonal utility comparisons“)

Die Verrechnung von Nutzen, wie sie die soziale Wohlfahrtsfunktion tätigt, setzt darüber hinaus die Möglichkeit interpersoneller Nutzenvergleiche („interpersonal utility comparisons“) voraus. Nach HARSANYI (vgl. im Folgenden HARSANYI 1979, S. 300 und 1977a, S. 638) tätigen wir alle tagtäglich interpersonelle Nutzenvergleiche, oder versuchen zumindest, welche zu tätigen. Immer und immer wieder steht die Entscheidung an, welchem Familienmitglied, Freund oder welcher wohltätigen Sache wir unsere Zeit widmen, wem wir Geld leihen, oder andere knappe Ressourcen investieren. Folgende Überlegung wird immer eine Rolle spielen: Wer braucht unsere Hilfe mehr, oder wer kann aus ihr den größeren Nutzen ziehen? Im Endeffekt werden also interpersonelle Nutzenvergleiche angestellt. „I do not think it is the task of a philosopher or a social scientist to deny the obvious fact that people often feel quite capable of making such comparisons“ (HARSANYI 1977a, S. 638).

Die Frage ist jedoch, auf welcher Grundlage solche Vergleiche basieren. Nach HARSANYI basieren sie auf Empathie: „[...] the basic intellectual operation in such interpersonal comparisons is imaginative empathy“ (HARSANYI, 1977a S.638). „[...] an ability to imagine ourselves to be in the shoes of other people“ (HARSANYI 1979, S. 300). Es liegt auf der Hand (vgl. im Folgenden HARSANYI 1979, S. 300 f.), dass man nicht einfach die Nutzenniveaus anderer Menschen mit dem eigenen vergleichen kann, wenn man nur die offensichtlichen Unterschiede („differences between their objective conditions“) wie Wohlstand, soziales Ansehen oder Gesundheitszustand in Betracht zieht. Vielmehr müssen auch mögliche Unterschiede bezüglich Vorlieben und Einstellungen berücksichtigt werden. „[R]ather, we have to make proper allowances also for possible differences between these people’s tastes or personal attitudes and our own“ (HARSANYI 1979, S. 300).

Auch in diesem Zusammenhang entpuppt sich ADAM SMITH als Ausgangspunkt der Überlegungen von HARSANYI. SMITH schreibt Gefühlen und der Fähigkeit Sympathie zu empfin-

den eine zentrale Rolle zu: Er sieht sie als Bestandteil der Natur des Menschen, welche ihm die Möglichkeit geben, spezifisch menschlich zu handeln (vgl. WILSON, DIXON 2008, S. 72 f.)[5]. Die Fähigkeit, sich in die Lage des anderen hineinzuversetzen „fellow-feeling" (SMITH 1790, S. 5), zu fühlen was der andere fühlt, ist die Grundvoraussetzung für Empathie.

SMITH (und damit auch HARSANYI) steht damit in überraschend großer Übereinstimmung mit den Erkenntnissen der heutigen Psychologie. Das unwillkürliche Erwidern eines charmanten Lächelns oder die unbewusste Einnahme ähnlicher Sitzhaltungen zweier Gesprächspartner sind Alltagsphänomene, in denen Menschen auf Stimmungen und Situationen anderer emotional einsteigen und die zu einem Gefühl gehörende Verhaltensweise unbewusst imitieren (vgl. BAUER 2008, S. 7 f). „When we see a stroke aimed and just ready to fall upon the leg or arm of another person, we naturally shrink and draw back our own leg or our own arm; and when it does fall, we feel it in some measure, and are hurt by it as well as the sufferer" (Smith 1790, S.5). Verantwortlich für diese sogenannten Resonanzphänomene sind die in der Neurobiologie entdeckten Spiegelnervenzellen (vgl. GALLESE, GOLDMAN 1998). Über die Resonanzphänomene hinaus verhelfen sie dem Menschen dazu, ein intuitives Wissen über den Verlauf von Situationen zu gewinnen. „Ohne intuitive Gewissheiten darüber, was eine gegebene Situation unmittelbar nach sich ziehen wird, wäre das Zusammenleben von Menschen kaum denkbar. „[...] Intuitiv zu spüren, was zu erwarten ist, kann [...] überlebenswichtig sein" (BAUER 2008, S. 13).

Da HARSANYI diese Erkenntnisse noch nicht zur Verfügung standen und die Aufgabe, sich in die Position des anderen hineinzuversetzen und sich zu fragen, wie viel Nutzen die Situation einem stiften würde, wenn man die Vorlieben und die persönlichen Einstellungen des anderen hätte, ist offensichtlich nicht leicht zu bewältigen und die Antwort darauf höchst unbestimmt und vage. Daher entwarf HARSANYI das Ähnlichkeitspostulat („similarity postulate").[6] Maßgeblich war in diesem Zusammenhang auch die Kritik von RAWLS an interpersonellen Nutzenvergleichen. So beklagt dieser, dass „sich diese Vergleichsregeln [...] beispielsweise aus bestimmten psychologischen Gesetzen herleiten, die für den Nutzen von Menschen maß-

[5] Hier wurde Smith stark vom Denken David Humes geprägt. Wobei sie sich auch hier in manchen Punkten widersprechen (vgl. WILSON, DIXON 2008, S. 71-72).

[6] „Of course, if any interpersonal utility comparisons are to be used in any formal theory then they must satisfy certain *consistency requirements* [...] From a logical point of view, interpersonal utility comparisons are based on what I have called the *similarity postulate*" (HARSANYI 1979 S. 301).

gebend sind, [...] diese Überlegung [jedoch] eine bestimmte Vorstellung vom Menschen voraussetzt. Die Beteiligten werden so vorgestellt, daß sie keine bestimmten Interessen höchster Stufe oder grundlegenden Ziele haben, anhand derer sie entscheiden würden, was für Menschen sie sein möchten. Sie haben gewissermaßen keinen bestimmten Charakter oder Willen" (RAWLS 1975, S. 200).

Es zeigt sich, dass auch hier die Kritik nicht ohne Wirkung geblieben ist, denn so hat HARSANYI seine ausführliche Begründung von 1955 komplett überarbeitet und erst 1977a das „similarity postulate" aufgestellt, um Nutzenvergleiche auf einen sicheren Boden zu stellen. 1979 ergänzte er diese Überlegung noch durch einen weiteren Punkt.[7]

Das „similarity postulate" besagt, dass unterschiedliche Menschen dieselben psychologischen Reaktionen in gegebenen Situationen zeigen und denselben Nutzen daraus ziehen werden, wenn den Unterschieden zwischen ihnen Rechnung getragen wird. „[I]t is reasonable to assume that different people will show very *similar* psychological reactions to any given objective situation, and will derive much the *same* utility or disutility from it – *once proper allowances have been made for any empirically observed differences*" (HARSANYI 1979, S. 301).

Dem Problem, dass interpersonelle Nutzenvergleiche als empirische Vorgänge nun auf einem nicht empirischen a priori Prinzip, nämlich dem Ähnlichkeitspostulat beruhen, tritt HARSANYI wie folgt entgegen: „The similarity postulate is the logical basis, not only for interpersonal comparisons of utility , but also *for assigning other people concious experiences at all*" (HARSANYI 1979, S. 302). Diejenigen die das Ähnlichkeitspostulat und damit interpersonelle Nutzenvergleiche, die darauf basieren, ablehnen, so argumentiert HARSANYI, müssten konsequenterweise daraus schlussfolgern, dass sie die einzigen Menschen mit Bewusstsein sind, die anderen hingegen „mindless automata". Wenn wir also die Annahme tätigen, in einer Welt mit Millionen anderen Menschen zu leben, die genau so real und bewusst leben wie wir, verlassen wir uns auf dasselbe Ähnlichkeitspostulat (vgl. HARSANYI 1977a, S. 640 f).

HARSANYI schlussfolgert daraus: „It is simply illogical to admit that people do have feelings and, therefore, do derive *some* satisfaction from a good meal in the same way we do; yet to resist the quantitative hypothesis that the *amount* of satisfaction they acually obtain from a

[7] So ist bei Abwesenheit eines klaren Beweises, die Annahme, dass das Verhalten der Menschen und deren psychologische Reaktion in gleichen Situationen diesselben sind, immer gültig (vgl. HARSANYI 1979, S. 301).

good dinner – that is, the personal importance they attach to a good dinner – must be much the same as it is in our own case, after proper allowances have been made [...]“ (HARSANYI 1977a, S. 641).[8]

Aufgrund der von HARSANYI angeführten Argumente und der neuen Erkenntnisse aus den Verhaltenswissenschaften bietet es sich an, die Frage der Ähnlichkeit zu entwirren und zweigeteilt zu beantworten. Zum einen ist das Ähnlichkeitspostulat insofern kein Postulat, sondern vielmehr ein Faktum, wie es die oben beschriebenen Resonanzphänomene nahe legen. Somit kann es als empirisch gesichert betrachtet werden, dass wir zumindest bis zu einem gewissen Grad ähnlich reagieren und ähnlich sind. „Die allererste Erfahrung, die Individuen unmittelbar miteinander teilen, ist die von Angst, Schmerz, Trauer, Not. Aber auch Freude, Liebe, Glück, Lust teilen sich direkt mit, unabhängig von den unterschiedlichen Anlässen, die solche Gefühle hervorrufen. Was wir am und mit dem eigenen Leibe spüren, gibt die Folie ab, auf welcher wir fremdes Leid und Glück mitsamt den jeweiligen Bedürfnissen ohne weiteres nachvollziehen und verstehen” (PIEPER 2007, S. 138).[9] So begründet sich diese Ähnlichkeit in unserem (empirischen) Sinneswesen: „Die Vertrautheit mit dem eigenen Leib, mit Freude und Leid als sinnlich erlebten Qualitäten verbindet die menschlichen Lebewesen über alle trennenden Schranken hinweg“ (PIEPER 2007, S. 138).

Zum anderen ist das Ähnlichkeitspostulat bzw. Ähnlichkeit notwendige Voraussetzung dafür, dass Ethik überhaupt möglich ist. Wenn es nichts gibt, was viele verschiedene Individuen gemein haben, so ist es unmöglich, je ein gemeinsames Verständnis von etwas zu erzeugen, geschweige denn eine Regel zu etablieren. Man befände sich in einem Naturzustand, in dem alle denkbaren Gleichgewichte oder Ergebnisse möglich sind. Denkt man sich eine Weltbevölkerung, die nur aus einem Kannibalen und einem heutigen Europäer besteht, so zeigt sich, dass zwar mehrere Resultate möglich sind wie z.B., dass der Europäer gegessen wird oder dass der Kannibale in Notwehr getötet wird, aber aufgrund der zu großen Verschiedenheit der zwei Akteure kann es nichts geben, worauf sie sich einigen könnten. Es ist dieser Unmöglichkeit geschuldet, dass sich eine Moral entwickeln kann, dass die Ethik (in obigem Falle) überflüssig wird, da ihr wissenschaftlicher Untersuchungsgegenstand, nämlich die Mo-

[8] Hier zeigt sich die Verknüpfung zur kardinalen Nutzenfunktion. Anhand interpersoneller Nutzenvergleiche wird versucht, den Nutzen des anderen zu messen. Beide Annahmen sind eng miteinander verbunden.

[9] Dabei sind viele unterschiedliche Anlässe denkbar. Wichtig ist, dass ähnliche Anlässe bei unterschiedlichen Personen ähnliche Gefühle hervorrufen.

ral, inexistent ist. „Die Ethik mag keine Moral zu erzeugen, sie ist ganz im Gegenteil auf ein Moralbewußtsein angewiesen“ (PIEPER 2007, S. 113).

Dies soll keineswegs dem klassischen Relativismusvorwurf Bahn brechen, sondern nur auf die hinreichende Ähnlichkeit verweisen, die notwendigerweise gegeben sein muss, damit sich Institutionen etablieren können. In einem Satz: Wir müssen *hinreichend ähnlich* sein, damit es Moral und damit auch eine Wissenschaft von der Moral, sprich Ethik geben kann. Die Rolle der Ähnlichkeit als „logische Basis” seiner Überlegungen hatte HARSANYI (s.o.) bereits angeführt, nur nicht in dieser Konsequenz. Die Kritik von RAWLS: „Der Utilitarismus nimmt die Verschiedenheit der einzelnen Menschen nicht ernst“ (RAWLS 1975, S. 45, siehe auch S. 214) erscheint damit in einem ganz neuen Licht, denn auch sein Theorieentwurf von Gerechtigkeit muss unweigerlich Spielerei sein, sollten wir nicht hinreichend ähnlich sein.

In diesem Zusammenhang stellt sich unweigerlich die Frage, inwiefern, auf welche Art und Weise oder auch ab welchem Maß man von hinreichender Ähnlichkeit sprechen kann. Ob es ausreicht, ein vernunftfähiges Wesen im kantianischen Sinne zu sein oder ob es notwendigerweise einer Schnittmenge an Moralvorstellungen oder Präferenzen bedarf. Vor dem Hintergrund der Definition des similarity postulate von HARSANYI scheint es sinnvoll, an genau diesem festzuhalten. Die psychischen grundlegenden Reaktionen bei der Nutzenmaximierung, sprich der Bedürfnisbefriedigung, müssen sehr ähnlich sein, in unseren persönlichen Präferenzen hingegen mag sich die evidente Verschiedenheit der Individuen und die Pluralität der Lebensentwürfe widerspiegeln.[10]

4. Abgrenzung zu anderen utilitaristischen Ansätzen

Das equiprobability model lässt sich verschiedenen utilitaristischen Denktraditionen zuordnen. Diese Zuordnung sowie eine Abgrenzung der jeweiligen Strömungen stehen hier im Mittelpunkt. So soll in diesem Kapitel zuerst der Präferenzutilitarismus gegenüber dem Hedonismus und dem Ideal Utilitarismus abgegrenzt werden, danach der Regelutilitarismus vom

[10] Eine eingehendere Untersuchung dieser *hinreichenden Ähnlichkeit* wird in einer kommenden Publikation seinen Platz finden.

Handlungsutilitarismus und zuletzt soll noch kurz ein Blick auf die Unterschiede vom Durchschnittsnutzenutilitarismus zum Nutzensummen-Utilitarismus geworfen werden.

4.1 Präferenzutilitarismus versus Hedonismus/Idealutilitarismus

In jeder utilitaristischen Theorie spielt die Maximierung von ‚sozialem Nutzen' eine tragende Rolle (vgl. im Folgenden HARSANYI 1977b, S. 27). Die klassischen Utilitaristen, wie vor allem BENTHAM, definierten ‚sozialen Nutzen' über die Kategorien Vergnügen und Schmerz („pleasure" und „pain") und gehören somit dem hedonistischen Ansatz an. Im Gegensatz dazu entwarf MOORE eine ideal-utilitaristische Definition von sozialem Nutzen, nämlich durch „mentale Zustände intrinsischen Wertes"[11] („mental states of intrinsic worth"). HARSANYI enthält sich dieser inhaltlichen Debatte mit dem equiprobability model, da hier sozialer Nutzen *formal* über Nutzen oder Präferenzen definiert wird: „The utilitarian theory I have proposed defines social utility in terms of individual utilities, and defines each person's utility function in terms of his personal preferences. Thus, in the end, social utility is defined in terms of people's personal preferences" (HARSANYI, 1977a, S. 644). Diesen Ansatz nennt er Präferenzutilitarismus („preference utilitarianism"). Damit gelingt es HARSANYI, eine *formale* Definition von Gerechtigkeit zu finden, die sich vollständig inhaltlichen Debatten enthält. Auch und gerade darin liegt die Stärke dieser Konzeption.

Wichtig ist, im Gegensatz zur goldenen Regel, dass andere Menschen so behandelt werden, wie sie es wollen, nach ihren Präferenzen (vgl. im Folgenden HARSANYI 1977b, S. 27 f). Dabei ist es gleichgültig, ob sie Vergnügen anstreben oder Geisteszustände intrinsischen Wertes, entscheidend ist ihre Präferenz, ihr Wille, etwas zu erreichen, unabhängig davon, was es ist. HARSANYI sieht sich damit im Einklang mit dem zentralen Prinzip der Präferenzautonomie (vgl. HARSANYI 1977a, S. 645). Zu entscheiden, was schlecht oder gut für einen ist, ist alleinige Aufgabe des betroffenen Individuums. Anhand der eigenen Wünsche und Präferenzen hat nur es allein dies zu beurteilen.

Mit der Definition sozialen Nutzens über individuelle Präferenzen „verschiebt sich die Wertbasis des Utilitarismus grundlegend. [...] Das Ziel der Moral ist nicht mehr die Herstel-

[11] Übersetzung des Autors.

lung bestimmter subjektiver Zustände, sondern die Herstellung bestimmter Weltzustände" (BIRNBACHER 2002, S. 101).[12] Durch die Berücksichtigung der Präferenzautonomie handelt sich HARSANYI jedoch ein schwieriges Problem ein. Was ist mit irrationalen Präferenzen? Nach seiner Meinung muss aus diesem Grund jede ethische Theorie zwischen rationalen und irrationalen Präferenzen unterscheiden. „It would be absurd to assert that we have the same moral obligation to help other people in satisfying their utterly unreasonable wants as we have to help them in satisfying their very reasonable desires" (HARSANYI 1977a, S. 646). „It would be absurd for any ethical theory to disregard this distinction" (HARSANYI 1977b, S. 29).

Um diese Unterscheidung sinnvoll vornehmen zu können, unterscheidet HARSANYI zwischen manifesten Präferenzen und wahren Präferenzen („manifest/explicit preferences" bzw. „true preferences") (vgl. im Folgenden HARSANYI 1977a, S. 646). Manifeste Präferenzen sind die Präferenzen, die wirklich vorliegen, falsche Vorstellungen, mangelhaftes Nachdenken oder starke Gefühle, die rationale Entscheidungen verhindern, eingeschlossen. Wahre Präferenzen sind die Präferenzen, die vorliegen würden, wenn all die relevanten Informationen vorliegen, mit besonderer Nachsicht überlegt wird und sich der Verstand in einem Zustand befindet, der zugänglich für rationale Entscheidungen ist (eben genau die obengenannten Faktoren ausgeschlossen sind). „Given this distinction, a person's rational wants are those consistent with his true preferences and, therefore, consistent with all the relevant factual information and with the best possible logical analysis of this information, whereas irrational wants are those that fail that test" (HARSANYI 1977a, S. 646). Daraus schlussfolgert HARSANYI, dass der soziale Nutzen anhand wahrer (rationaler) Präferenzen und nicht anhand manifester Präferenzen zu definieren ist.

Außerdem fügt HARSANYI dieser Unterscheidung hinzu, dass antisoziale Präferenzen ebenfalls bei der Bestimmung des sozialen Nutzens auszuschließen sind: „We have to disregard, not only preferences distorted by factual or logical errors, but also preferences based on clearly anitsocial attitudes, such as sadism, resentment, or malice" (HARSANYI 1977b, S. 30). Dass

[12] Das Argument ist nicht völlig einleuchtend. Ziel der Moral ist es Zustände herzustellen, nach denen die Individuen streben, sie definieren sie anhand *ihrer* Präferenzen. Das mögen subjektive Zustände sein wie auch Weltzustände. Genau aus dieser inhaltlichen Debatte, die die Präferenzautonomie illegitimerweise einschränkt, entging HARSANYI durch die *formale* Definition des sozialen Nutzens anhand persönlicher Präferenzen. Die Inhalte geben die Individuen anhand ihrer Präferenzen vor!

dazu auch ein klares theoretisches Kriterium notwendig ist, ist HARSANYI bewusst, jedoch stellt er keines dazu vor.

HARSANYI gelingt es folglich, die bisherigen Einengungen von Hedonismus und Idealutilitarismus aufzulösen, jedoch steht er damit vor dem Problem irrationaler und antisozialer Präferenzen. So schreibt BIRNBACHER (2002), dass der Präferenzutilitarismus das „Verifikationsproblem", das er umgehen wollte, wiederherstellt. „Die wahren Präferenzen eines Menschen herauszufinden scheint nicht weniger schwierig als herauszufinden, was ihn glücklich macht" (BIRNBACHER 2002, S. 102). Dem ist sicherlich zuzustimmen, was den entscheidenden Unterschied jedoch ausmacht ist, dass der Präferenzutilitarismus als einziger Ansatz dem Prinzip der Präferenzautonomie gerecht wird.

Insoweit ist der Ausschluss von sogenannten asozialen Präferenzen („antisocial preferences") weit stärker zu kritisieren, da damit dieses Prinzip sofort wieder verletzt wird, also eine *inhaltliche* Einschränkung vorliegt. HARSANYI zählt zu diesen Präferenzen z.B. Sadismus, Neid, Verbitterung und Bösartigkeit (vgl. HARSANYI 1977a, S. 647 bzw. s.o.). Ein Kriterium, anhand dessen diese Präferenzen bestimmt werden können, liefert er nicht. Diese unbegründete inhaltliche Einengung der wahren Präferenzen ist aber überhaupt nicht notwendig. Das equiprobability model dient ja gerade dazu, das richtige Maß an Regeln und Regelbindung, das soziale Optimum, zu finden. Will nun ein bestimmter Bevölkerungsteil seine sadistischen Präferenzen ausleben, so hat die Gesamtheit *n* anhand des equiprobability model darüber zu bestimmen, welche Regel sie sich dazu geben will. Diese ist dann von den Gesellschaftsmitgliedern (mit eventuellen Ausnahmen) einzuhalten. Die Gesellschaft definiert selbst anhand dieses Entscheidungsinstruments das soziale Optimum. Folglich kann mit dem equiprobaility model festgestellt werden, welche Präferenzen asozial sind, denn dafür werden Regeln erlassen werden, die keine oder nur sehr wenige Ausnahmen dulden (im Sinne eines Verbots, dasselbe wird aber auch bei wichtigen Geboten oder Pflichten der Fall sein, womit dies doch nicht eindeutig festgestellt werden kann aufgrund der formalen Definition).

4.2 Regelutilitarismus versus Handlungsutilitarismus

Wie bereits angesprochen, basiert die Grundidee des Utilitarismus darauf, moralische Probleme anhand des Prinzips der Nutzenmaximierung, genauer, der Maximierung des sozialen Nutzens, zu lösen (vgl. im Folgenden HARSANYI 1983, S. 234). Für einen Handlungsutilitaristen besteht das moralische Grundproblem nun darin, unter den Handlungsalternativen diejenige auszusuchen, die den sozialen Nutzen maximiert. Im Gegenteil dazu besteht das moralische Grundproblem für einen Regelutilitaristen nicht darin, die richtige Handlung unter den möglichen Alternativen zu wählen, sondern darin, unter verschiedenen möglichen moralischen Regeln („moral rules“ oder „moral codes“) die richtige Alternative zu wählen (eben die, die den sozialen Nutzen maximiert). „Indeed, at a more fundamental level, it is a choice among alternative general *moral codes*“ (HARSANYI 1983, S. 234). Zu dieser Kategorie zählt HARSANYI sein equiprobability model.

Im Regelutilitarismus geht es also in erster Linie darum, das utilitaristische Entscheidungskriterium (Maximierung sozialen Nutzens) auf die Wahl zwischen verschiedenen möglichen moralischen Regeln anzuwenden und nicht direkt auf die Wahl zwischen verschiedenen möglichen Handlungen (vgl. im Folgenden HARSANYI 1980, S. 116). Folglich ist die moralisch richtige Handlung in einer gegebenen Situation die, die der richtigen moralischen Regel für eine solche Situation entspricht. Die richtige moralische Regel ist als die Regel definiert, die, wenn sie von allen in dieser Art von Situation eingehalten wird, den sozialen Nutzen maximiert. „Accordingly, for a rule utilitarian the basic moral problem will be to choose that particular *moral code* that will yield the highest expected social utility if it is followed by all rational and morally motivated people“ (HARSANYI 1983, S. 234).

Was natürlich voraussetzt, dass die Regel jedem bekannt ist. Es wird also davon ausgegangen, dass - wenn eine Regel angenommen wird - dies automatisch allgemeines Wissen wird. So lautet das Entscheidungsproblem im Endeffekt wie folgt: „It is the problem of choosing that particular moral code that would yield the highest expected social utility if all rational and morally motivated people followed this moral code and if everybody knew the contents of this moral code and also knew that it would be followed by all rational and morally motivated people“ (HARSANYI 1989, S. 235). Das bedeutet: wenn eine moralische Regel nach dem sozialem Nutzen den sie stiftet, beurteilt werden soll, auch Erwartungs- und Anreizeffekte („ex-

pectation effect“ und „incentive effect“)[13] zu berücksichtigen sind, die sie auf Personen ausübt, sobald diese sie kennen.

Ein Beispiel über das Halten von Versprechen veranschaulicht dies (vgl. im Folgenden HARSANYI 1983, S. 237 f.).[14] Wenn jemand ein Versprechen gibt und die Einhaltung als sehr schwer empfindet, ist es dann moralisch in Ordnung, wenn er es bricht? Ein Handlungsutilitarist schaut sich dazu drei Faktoren an: Erstens den Nutzenzugewinn, den der Versprechensbrecher verzeichnet, zweitens der Nutzenverlust, den derjenige verzeichnet, dem das Versprechen gegeben wurde und drittens der Effekt auf die Gesellschaft, die eine minimale Nutzeneinbuße dahingehend erleidet, dass ihr Vertrauen in gegebene Versprechen sinkt. Natürlich ist letzterer Effekt vernachlässigbar, da ja ein gebrochenes Versprechen keine große Auswirkung auf das generelle Vertrauen in Versprechen haben kann. Sobald also der Nettonutzen größer Null ist, ist es für einen Handlungsutilitaristen moralisch zulässig (wenn nicht geboten) das Versprechen zu brechen, was z.B. unserem Moralverständnis völlig widerspricht.

Der Regelutilitarist hingegen stellt sich die Frage, wie die richtige moralische Regel in diesem Zusammenhang auszusehen hat: „[...] would you like to prefer to live in a society in which even highly moral people regularly broke their promises on similar occasions or in a society in which they felt obliged to keep their promises on such occasions?“ (HARSANYI 1983, S. 238). Zur Beantwortung werden die eben genannten drei Stichpunkte herangezogen, wobei die ersten beiden gleich behandelt werden, nicht jedoch der dritte. Hier ist folgendes zu berücksichtigen: „The utility *loss* the entire society would suffer by a diminshed confidence in promises if it were *commonly known* that even very moral people would *regularly break* their promises on similar occasions“ (HARSANYI 1983, S. 238). Natürlich ist die Auswirkung, wenn *erwartet* wird, dass regelmäßig Versprechen gebrochen werden, um vieles schwerwiegender als ein einmaliger Vorgang auf das generelle Vertrauen in Versprechen. Genau aus diesem

[13] Des weiteren werden noch der „coordination effect“ und der „division-of-labor effect“ von HARSANYI angeführt. In dem zitierten Paper von 1980 findet sich zudem eine ausführliche spieltheoretisch begründete Unterscheidung von Regel- zu Handlungsutilitarismus, bei der sich der Handlungsutilitarismus als unkooperativ gespieltes Spiel in extensiver Form und der Regelutilitarismus als kooperativ gespieltes Spiel in Normalform charakterisieren lässt. Hieraus wird dann der „coordination effect“ abgeleitet.

[14] In diesem Paper finden sich drei weitere Beispiele, anhand derer sich die beiden Utilitarismusarten vortrefflich voneinander abgrenzen lassen. Aus Platzgründen kann an dieser Stelle nur darauf verwiesen werden.

Grund wird es bei einem Regelutilitaristen viel weniger Situationen geben, in denen das Nichteinhalten von Versprechen erlaubt ist.[15]

Geht man an diesem Punkt konsequent weiter, so zeigt sich, dass der Handlungs-Utilitarismus nicht dazu in der Lage ist, soziale Institutionen wie beispielsweise individuelle Rechte, aber auch Pflichten zu berücksichtigen. Alles wird aufgrund von sozialen Nutzenerwägungen überrollt („overridden merely on the basis of social-expediency considerations"). „[A]ct utilitarianism essentially destroys all individual rights and all personal obligations" (HARSANYI 1980, S. 125). Der Regelutilitarismus hat im Gegenteil dazu keinerlei Schwierigkeiten, diese Rechte und Pflichten mitzuberücksichtigen und kommt damit erst als ernsthafte ethische Theorie in Frage.

Darüber hinaus bietet er eine rationale Rechtfertigung derselben anhand des sozialen Nutzens. „It not only recognizes the *existence* of such morally binding rights and obligations, but also provides a *rational justification* for them in terms of their social utility. Moreover, it also provides a rational criterion for discovering the legitimate *exceptions* to actual observance of these rights and obligations, which common-sense morality cannot do" (HARSANYI 1980, S. 126).

So ist der Regelutilitarismus zudem in der Lage, legitime Ausnahmen zu entdecken und kann damit das 'soziale Optimum' („social optimum") (an Regelbindung und Ausnahmen) bestimmen. HARSANYI sieht den Regelutilitarismus damit im Mittelfeld zwischen dem Handlungsutilitarismus, der nicht dazu in der Lage ist, individuelle Rechte und Pflichten zu berücksichtigen und vielen anderen ethischen Theorien, wie deontologische, vertragstheoretische und anspruchsbasierte Theorien,[16] die diese zwar berücksichtigen, aber daran scheitern, eine rationale Rechtfertigung dafür zu finden (vgl. HARSANYI 1980, S. 127).

Es kommt mitunter vor, dass der Regelutilitarismus als „ethischer Zwitter" (vgl. BIRNBACHER 2002, S. 100) kritisiert wird, da er in der Lage ist, deontologische Regeln konsequenzialistisch anhand extrinsischer Werte (sozialer Nutzen) herzuleiten und zu begründen: „Wenn der Regelutilitarismus die *Geltung* der Sekundärprinzipien durch die Nützlichkeit ihrer gesellschaftlichen Akzeptanz rechtfertigt, warum rechtfertigt er nicht dann auch ihre *Befolgung* utilitaristisch und lässt Situationen, in denen deren Befolgung katastrophal schlechte Folgen

[15] Natürlich kommt es im Endeffekt immer auf die Präferenzen aller Betroffenen an. Sie entscheiden, ob sie eine eher laxe oder eher strikte Regelhandhabung wünschen (vgl. HARSANYI 1983, S. 239).

erwarten lässt, Ausnahmen zu?“ (BIRNBACHER 2002, S. 100). Oben wurde gezeigt, dass der Regelutilitarismus in der Lage ist, die Geltung *und* die Befolgung von Regeln utilitaristisch zu rechtfertigen. Zurück zum Beispiel: „[A]ny reasonable moral code must permit some exceptions to the general obligation of keeping our promises, at least in cases of special hardship. On the other hand, it must not permit too many easy exceptions [...] the social-utility maximization principle, at least in theory, does provide a conceptually clear criterion for finding this social optimum“ (HARSANYI 1980, S. 126).

Das Stichwort liegt im Wort ‚*finden*‘. Anhand des equiprobability model lässt sich zumindest theoretisch dieses Optimum finden, definiert und festgelegt wird es durch die Präferenzen der Individuen. Genauer: „rule utilitarian theory will in principle always provide a conceptually clear *theoretical criterion* [...] for deciding precisely *what* particular situations should be regarded as *morally permitted exceptions*“ (HARSANYI 1983, S. 239). Der Regelutilitarismus bietet also ein klares theoretisches Kriterium, anhand dessen die Ausnahmen bestimmt werden können. Es kommt jedoch auf die Präferenzen der Individuen an, wo sie, für sich, dieses soziale Optimum sehen und welche Ausnahmen sie haben möchten, damit ist die Befolgung der Regel utilitaristisch begründet. Das equiprobability model hilft ihnen nur dabei, dieses Optimum zu finden. Diese Kritik ist folglich zurückzuweisen, zeigt gerade darin doch der Regelutilitarismus seine Potenz als ethische Theorie.

4.3 Durchschnittsnutzenutilitarismus versus Nutzensummenutilitarismus

Im Gegensatz zu obigen Gegenüberstellungen ist die folgende sehr einfach. Der Unterschied zwischen diesen beiden Richtungen wurde in Kürze schon in Kapitel 2 besprochen. Das equiprobability model definiert eine soziale Wohlfahrtsfunktion, die den Durchschnittsnutzen maximiert und gehört damit zum Durchschnittsnutzen-Utilitarismus. Wählt man als Entscheidungskriterium die Nutzensummenmaximierung, so bewegt man sich natürlich im Bereich des Nutzensummenutilitarismus. Nimmt man nun an, dass n konstant ist, gibt es zwischen beiden Arten keine Unterschiede. Da n jedoch, zumindest auf längere Sicht variabel ist, ergeben sich bei beiden Arten unterschiedliche Entscheidungskriterien. HARSANYI ist davon über-

[16] Übersetzung des Autors.

zeugt, dass das Durchschnittsnutzenkriterium in diesem Fall zu eindeutig besseren Ergebnissen führt (vgl. HARSANYI 1977a, S. 633 Fn. 12).

5. Das equiprobability model als Theorie der Gerechtigkeit

Bekanntermaßen ist JOHN RAWLS ein vehementer Kritiker HARSANYIs und umgekehrt. So schrieb HARSANYI dazu: „I am the last person to defend John Rawls's ethical theory" (HARSANYI 1978, S. 1). Im Gegensatz zur herrschenden Lehrmeinung soll im Folgenden gezeigt werden, dass die utilitaristische Konzeption von Gerechtigkeit nach der Spielart von HARSANYI derjenigen von RAWLS überlegen ist. Dabei erfolgt ein weiterer Blick auf den Kern des Modells, das „equiprobability postulate", das grundlegend für die Herleitung der sozialen Wohlfahrtsfunktion ist. Zusätzlich soll an dieser Stelle auch die Kritik von HARSANYI an RAWLS dargestellt werden, da sich hierhin zum Teil eine Rechtfertigung für die eigene Vorgehensweise finden lässt.

HARSANYI selbst bezeichnet sein Modell, oder besser gesagt sein Gedankenexperiment, wie folgt mit Anspielung auf JOHN RAWLS: „This is my own version of the concept of the ‚original position'." (HARSANYI 1975, S. 598). „HARSANYI besteht (allerdings) auf der historischen Priorität dieser Idee, wenn er auch [...] RAWLS' Terminus „original position" in diesem Zusammenhang nicht benutzt hat" (HOMANN 1988, S. 222 mit Verweis auf HARSANYI 1975, S. 595). Konzeptionell unterscheiden sich die Ansätze von RAWLS und HARSANYI kaum, doch HARSANYI argumentiert, dass die Nützlichkeit des Ansatzes entscheidend davon abhängt, ob dazu eine sinnvolle Entscheidungsregel gefunden wird oder nicht, was er bei RAWLS bestreitet (vgl. HARSANYI 1975, S. 595 f.).

In der Terminologie von HARSANYI bedeutet dies, dass die „primary definition of rationality" bei beiden sehr ähnlich ist, sich die „secondary definition of morality" jedoch grundlegend unterscheidet. RAWLS lehnt den Gebrauch von Wahrscheinlichkeiten im Urzustand strikt ab und wählt aus diesem Grund das Maximin Prinzip (Waldregel) als Entscheidungskriterium (vgl. RAWLS 1975, S. 201). Wobei RAWLS nur von einer Beziehung zur Maximin-Regel

spricht (RAWLS 1975, S. 177).[17] HARSANYI argumentiert für die Zulassung von Wahrscheinlichkeiten anhand dem 'Prinzip des mangelnden Grundes' („principle of indifference“), also dem Umstand, dass, sofern im Zustand von Unsicherheit, Wahrscheinlichkeiten benutzt werden, jedem möglichen Ereignis dieselbe Wahrscheinlichkeit zuzuweisen ist und kommt zu der bekannten Erwartungswertregel. Die von RAWLS gewählte Maximin Regel kritisiert er scharf als höchst irrationale Entscheidungsregel: „I have argued that the maximin principle would often lead to highly irrational decisions in everyday life“ (HARSANYI 1975, S. 596).

RAWLS weist die LAPLACEsche Regel nicht zurück, er weiß, dass „der Schleier des Nichtwissens [...] unmittelbar auf das Problem der Entscheidung unter völliger Unsicherheit“ (RAWLS 1975, S. 196) führt und somit von diesem Prinzip des unzureichenden Grundes auszugehen ist. Er hat trotzdem Vorbehalte gegenüber diesem Vorgehen, da es „[w]esentlich ist, daß die gewählten Grundsätze nicht von der Risikofreudigkeit abhängen dürfen.“[18] Zweitens nimmt RAWLS an, dass „Wahrscheinlichkeitsaussagen eine objektive Grundlage haben müssen“ (RAWLS 1975, S. 197). Dem entgegnet HARSANYI, dass es unmöglich ist, subjektive Wahrscheinlichkeiten auszublenden: „any decision maker whose behavior is consistent with a few – very compelling – rationality postulates simply *cannot help* acting *as if* he used subjective probabilities. (More precisely, he cannot help acting *as if* he tried to maximize his expected utility, computed on the basis of some set of subjective probabilites.)“ (HARSANYI 1975, S. 599).

Über diese Rechtfertigung anhand des Prinzips des unzureichenden Grundes hinaus rechfertigt HARSANYI die „equiprobability assumption“ wie folgt: „[I]t may be regarded as an expression of the purely *moral* principle, that, in making basic moral value judgments, we must give the same *a priori* weight to the interests of all members of the society“ (HARSANYI 1975, S. 598 Fn. 10).

„Die Annahme gleicher Wahrscheinlichkeiten im Urzustand soll ausdrücken, daß den Nutzenfunktionen aller Individuen ‚the same equal weight‘[19] zukommt“ (HOMANN 1988, S. 225). Dies drückt sich dann in der Entscheidungsregel konkret (mathematisch) aus. Jede Nutzen-

---

[17] Dazu vertritt HOMANN folgende Meinung: „Es ist also schon philologisch unmöglich, Rawls auf die ‚Begründung‘ seiner Gerechtigkeitstheorie auf der Maximin-Regel festzulegen und diese Begründung dann zu kritisieren: Dies führt zu nichts [...]“ (HOMANN 1988, S. 231).

[18] Bei der Erwartungswertregel von HARSANYI liegt Risikoneutralität vor, worauf RAWLS leider nicht eingeht.

funktion, beziehungsweise jedes Nutzenniveau U_1, U_2, ..., U_n, das ein Individuum 1, 2, ..., n unter dem jeweiligen System genießen könnte, erhält dasselbe Gewicht. „Hinter HARSANYIs Gleichwahrscheinlichkeitsannahme steht also eine erklärtermaßen moralische Position: Allen Individuen kommt in der Gesellschaft das ‚gleiche Gewicht' zu" (HOMANN 1988, S. 226).

Diese Frage des Gewichts hat HOMANN als zentralen Punkt herausgearbeitet „an dem sich die Wege von RAWLS und HARSANYI trennen [...]: Die konzeptionelle Differenz zwischen der vertragstheoretischen Gerechtigkeitstheorie von RAWLS und dem Utilitarismus von HARSANYI ist darauf zurückzuführen, daß HARSANYI und RAWLS dem ‚gleichen Gewicht' implizit verschiedene Interpretationen zugrunde legen und daß es daher, weil ‚Gleichheit' eine unverzichtbare Forderung von Demokratie ist, unterschiedliche Vorstellungen von Demokratie sind, die die konzeptionelle Differenz begründen" (HOMANN 1988, S. 225 f.). Denn „[*u*]*nbeschadet der Gleichheit kann dieses Gewicht sehr unterschiedlich sein*" (HOMANN 1988, S. 226).

„Der entscheidende Unterschied", so argumentiert HOMANN, „liegt zwischen – mathematisch gesprochen – einem *endlichen* und einem *unendlichen* Gewicht". Liegen Stimmen mit endlichem Gewicht vor, so ist es mathematisch möglich, diese zu addieren und zu vergleichen, haben sie aber unendliches Gewicht, so ist dies zwar nach wie vor möglich, jedoch wenig sinnvoll (vgl. HOMANN 1988, S. 226). „Im ersten Fall kann es Mehrheitsentscheidungen geben, im letzten Fall nicht: Gibt es im letzten Fall auch nur eine einzige Gegenstimme, dann ist die Entscheidung infolge des unendlichen Gewichts dieser Stimme nicht gültig. Bei einem unendlichen Gewicht hat jeder einzelne ein Vetorecht gegen kollektive Entscheidungen" (HOMANN 1988, S. 226).

HOMANN schlussfolgert daraus, dass HARSANYI die Gleichheit des Gewichts im „Sinne eines endlichen Gewichts" und RAWLS im „Sinne eines unendlichen Gewichts" verstanden haben: HARSANYI lässt sich von dem „Ein-Mann-eine-Stimme-Prinzip leiten, das für die normative Grundlegung der Demokratie jedoch völlig ungeeignet ist. RAWLS hingegen geht implizit von dem unendlichen Gewicht der Interessen eines jeden Individuums aus, was sich

[19] HARSANYI (1976), S. 35.

daran zeigt, daß er die ‚moralische Person' als Grundbegriff der Theorie etabliert und ihr das klassische Merkmal der ‚Würde' zuspricht" (HOMANN 1988, S. 227).[20]

HOMANN kommt folglich zu dem Schluss, dass „für eine Gerechtigkeitstheorie in einer Demokratie ein RAWLSscher Theorietyp den Vorzug gegenüber einem utilitaristischen Typ verdient" (HOMANN 1988, S. 230). Dieser Kritik ist eindeutig entgegenzutreten, scheint es doch, als ob HOMANN ein wichtiges Detail übersehen hat: Die soziale Wohlfahrtsfunktion ist für *jeden* Entscheidungsträger im Urzustand *dieselbe* – wenn die entsprechenden Voraussetzungen (siehe Kapitel 3) gelten: „Typically, different individuals will have very different utility functions U_i but, as can be seen from Equation (1)[21] above, in theory they will tend to have identical social-welfare functions – but only if they agree in their factual assumptions on the nature of the individual utility functions U_i and on the conversion ratios between different individuals' utilities (as decided by interpersonal utility comparisons) – which, of course, may not be the case" (HARSANYI 1977a, S. 635).

Gelten also die von HARSANYI getroffenen Annahmen, so kommt es, dass jeder *dieselbe* Wohlfahrtsfunktion besitzt und somit *alle* nach derselben Entscheidungsregel vorgehen. Eine Stimmenverrechnung findet *nicht* statt, vielmehr garantiert die soziale Wohlfahrtsfunktion eine *einstimmige* Entscheidung.[22] Dies trifft jedoch nur zu, das soll nochmals betont werden, wenn kardinale Nutzenfunktionen vorliegen und interpersonelle Nutzenvergleiche sinnvoll möglich sind. An diesen Voraussetzungen übt HOMANN jedoch keine schwerwiegende Kritik (vgl. HOMANN 1988, S. 223 f.).

Da dies (zumindest theoretisch) zutrifft, ist letztendlich HARSANYIs Moral des Durchschnittsnutzens eindeutig der Vorzug gegenüber der Theorie der Gerechtigkeit von RAWLS zu geben. Stimmen die Annahmen, so findet eine Stimmenverrechnung nicht statt, da die soziale Wohlfahrtsfunktion eine einstimmige Entscheidung garantiert, die mathematische Gewich-

---

[20] Dieses Zusprechen von ‚Würde' ist ein unzulässiger *inhaltlicher* Vorgriff. Diesem Problem entgeht HARSANYI in seiner Konzeption. Es soll ja gerade im Urzustand beschlossen werden ob die Menschen sich eine solche Regel geben wollen oder nicht.

[21] Damit verweist HARSANYI auf die soziale Wohlfahrtsfunktion.

[22] Die Kritik die HOMANN an RAWLS äußert, „daß er hinter seinem Schleier des Nichtwissens den kollektiven Entscheidungen und den Abstimmungsregeln für kollektive Entscheidungen ebensowenig Aufmerksamkeit widmet wie HARSANYI" (HOMANN 1988, S. 227), trifft damit auch auf ihn selbst zu.

tung von Stimmen spielt somit keine Rolle und mit HARSANYIs equiprobability model gelangt man zu eindeutig besseren Ergebnissen wie mit der Maximin Heuristik von RAWLS.[23]

6. Fazit

Nach einer kurzen Darstellung der Wurzeln des „equiprobability model of moral value judgments“ lag der Fokus auf dem Modell selbst. Zu diesem Modell gelangt HARSANYI anhand der Gleichwahrscheinlichkeitsannahme. Als Entscheidungsregel definiert er daraufhin die Maximierung des Erwartungswerts, die soziale Wohlfahrtsfunktion. Um seine Überlegungen auf ein festes Fundament zu setzten trifft HARSANYI zwei Annahmen: Zum einen geht er von kardinalen Nutzenfunktionen aus, zum anderen begründet er das Ähnlichkeitspostulat um interpersonelle Nutzenvergleiche für die Theorie nutzen zu können. In diesem Zusammenhang wurde der Standpunkt vertreten, dass es sich bei dem equiprobability postulate zum einen um ein Faktum handelt und zum anderen um eine theoretische Notwendigkeit. Das Merkmal der Ähnlichkeit, so wurde vorgeschlagen, hat sich dabei auf grundlegende psychische Reaktionen zu beziehen.

Das equiprobability model kann verschiedenen utilitaristischen Denkschulen zugeordnet werden. HARSANYIs Modell gehört erstens zum Präferenzutilitarismus, womit es, im Gegensatz zu anderen Ansätzen, die Präferenzautonomie achtet. Allerdings bringt dies das Problem irrationaler Präferenzen mit sich, das HARSANYI durch eine Unterscheidung von wahren und manifesten Präferenzen zu lösen versucht. Der Ausschluss von ‚asozialen Präferenzen‘, wie ihn HARSANYI fordert, zerstört die zuvor berücksichtigte Präferenzautonomie und ist meiner Meinung nach zu kritisieren und darüber hinaus auch nicht notwendig. Zentrales Merkmal dieser Konzeption von Gerechtigkeit ist auch ihre rein formale Definition, die dem ökonomischen Ansatz zu verdanken ist, inhaltliche Vorausnahmen oder Setzungen entfallen vollständig.

Des Weiteren ist das Modell dem Regelutilitarismus zuzuordnen. Dieser ist dazu in der Lage, persönliche Rechte und Pflichten zu berücksichtigen und rational zu begründen, damit etabliert sich das Konzept erst zu einer ernsthaften ethischen Theorie. Im Gegensatz zum

[23] Siehe dazu z.B. HARSANYI (1975), S. 594 ff..

Handlungsutilitarismus stimmen die Ergebnisse, zu denen er gelangt, stark mit denen unserer gängigen Moralvorstellungen überein. Als letztes ist das Modell dem Durchschnittsnutzenutilitarismus zuzuordnen. Aus diesem Grund lässt sich auch von einer Ethik des Durchschnittsnutzens sprechen.

Die Auseinandersetzung mit JOHN RAWLS zeigt, dass sich HARSANYI immer wieder erneut um die Rechfertigung seines Modells bemüht hat. Im Gegenzug ist er selbst schärfster Kritiker von RAWLS. Obwohl beide von einem fast identischen ‚Urzustand' ausgehen, leiten beide höchst verschiedene Entscheidungsregeln daraus ab. HOMANN hat dabei die Stimmengewichtung als den zentralen Punkt herausgearbeitet, an denen sich die Wege der Zwei trennen und kommt zu dem Schluss, der Theorie von RAWLS den Vorzug zu geben. Ich bin gegensätzlicher Auffassung, da es bei gültigen Annahmen im Modell HARSANYIs zu einstimmigen Entscheidungen und besseren Ergebnissen kommt.

Daher ist die ethische Theorie von HARSANYI zu Unrecht als inferiore Alternative angesehen, die in der Lehre auch so behandelt wird. Es lässt sich hoffen, dass diese moderne Form des Utilitarismus wieder verstärkt seinen Niederschlag in der Ethik wie auch z.B. in der politischen Philosophie oder der Rechtsphilosophie findet.

Literaturverzeichnis

ARROW, KENNETH J. (1963): Social Choice and Individual Values. Wiley: New York.

BAUER, J. (2008): Warum ich fühle, was du fühlst – Intuitive Kommunikation und das Geheimnis der Spiegelneurone. Heyne: München.

BECKER, M., SCHMIDT, J., ZINTL, R. (2006): Politische Philosophie. Schöningh, Paderborn.

BIRNBACHER, DIETER (2002): Utilitarismus/Ethischer Egoismus. In: Düwell, Marcus; Hübenthal, Christoph; Werner, Micha H. (Hrsg.): Handbuch Ethik. Metzler: Stuttgart/Weimar.

GALLESE, V., GOLDMAN, A. (1998): Mirror neurons and the simulation theory of mind-reading, *Trends in Cognitive Sciences* 2(12), 493-501.

HARSANYI, J. C. (1953): Cardinal Utility in Welfare Economics and in the Theory of Risk-taking, *The Journal of Political Economy* 51(5), 434-435.

HARSANYI, J. C. (1955): Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility, *The Journal of Political Economy* 63(4), 309-321.

HARSANYI, J. C. (1975): Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory, *The American Political Science Review* 69(2), 594-606.

HARSANYI, J. C. (1976): Essays on Ehtics, Social Behavior, and Scientific Explanation. D. Reidel Publishing Company: Dordrecht .

HARSANYI, J. C. (1977a): Morality and the Theory of Rational Behavior, *Social Research* 44, 623-656.

HARSANYI, J. C. (1977b): Rule Utilitarianism and Decision Theory, *Erkenntnis* 11, 25-53.

HARSANYI, J. C. (1978): Sneed on RAWLS's Theory of Social Institutions: Some Comments, *Erkenntnis* 13, 225-230.

HARSANYI, J. C. (1979): Bayesian Decision Theory, Rule Utilitarianism, and Arrow's Impossibility Theorem, *Theory and Decision* 11, 289-317.

HARSANYI, J. C. (1980): Rule Utilitarianism, Rights, Obligations, and the Theory of Rational Behavior, *Theory and Decision* 12, 115-133.

HARSANYI, J. C. (1983): Basic Moral Decisions and Alternative Concepts of Rationality, *Social Theory and Practice* 9, 231- 244.

PIEPER, A. (2007): Einführung in die Ethik. Francke: Tübingen/Basel.

RAWLS, J. (1975): Eine Theorie der Gerechtigkeit. Suhrkamp: Frankfurt am Main.

RUCKRIEGEL, K. (2007): Glücksforschung, Das *Wirtschaftsstudium* 36(10), 515-530.

SMITH, ADAM (1790): The Theory of Moral Sentiments. 6. Auflage, London. http://www.ibiblio.org/ml/libri/s/SmithA_MoralSentiments_p.pdf, zugegriffen am 13.05.2007.

WILSON, D., DIXON, W. (2008): Reciprocity and economics in historical perspective, *International Review of Economics* 55(1-2), 65-76.