

# Utilix — Verteilungsgerechtigkeit auf Empathiebasis<sup>1</sup>

CHRISTOPH LUMER

Ziel dieses Beitrags ist es, eine moralische Bewertungsfunktion zu entwickeln, die einerseits prudentiell begründet ist, andererseits mit moralischen Intuitionen übereinstimmt.

## 1. Moralische Intuitionen im Experiment — die moralische Bewertungsfunktion Utilix

Im nichtphilosophischen Bekanntenkreis habe ich ein kleines Experiment zu moralischen Intuitionen veranstaltet. Ich habe den Beteiligten diese Tabelle vorgelegt:

*Das Ausgangsbeispiel: Das Wohl unterschiedlicher Personengruppen in verschiedenen Gesellschaftsordnungen*

$u_{ij}$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$
$s_1$	-1	-1	0	0	0	1,5	1,25	-1	-0,5	0	0	0
$s_2$	1	2	1	2	2	1,5	1,25	1	0	0	0,5	0,5
$s_3$	2	2	2	1,5	2	1,5	1,25	2	3,0	3,0	1,5	2,0
$s_4$	4	2	3	1,5	2	1,5	1,25	3,7	3,5	2,0	4,0	3,0
$\Sigma$	6	5	6	5	6	6	5	5,7	6,0	5,0	6,0	5,5

Dazu habe ich folgendes erläutert:  $a_1$  bis  $a_{12}$  seien alternative Gesellschaftsordnungen;  $s_1$  bis  $s_4$  seien gleich große homogene Gruppen innerhalb der Gesellschaft. Die Werte innerhalb der Matrix stellen das je individuelle durchschnittliche Wohlbefinden der Mitglieder dieser Gruppe dar. Der Wert '4' sei dabei das Höchste, was Menschen im Schnitt über längere Zeit an Wohlbefinden erreichen könnten; wer hingegen das Wohlbefindensniveau -1 habe, sei nur gerade so eben noch lebensfähig, weil dauernd am Rande des Selbstmordes. — Die Versuchspersonen sollten nun diese Alternativen in

<sup>1</sup> Ich danke Reinhard Suck für seine geduldige und verständnisvolle Beratung in mathematischen Fragen.

eine Präferenzordnung bringen, allein nach ihren moralischen Intuitionen, und zwar zunächst die ersten acht, dann die letzten vier Alternativen. Die drei wichtigsten intuitiven Präferenzordnungen waren:

*Intuitive Präferenzordnungen über das Ausgangsbeispiel*

Intuition <sub>1t</sub> [=G <sub>e19</sub> =G <sub>e20</sub> ]:	$a_6 > a_7 > a_5 > a_3 > a_4 > a_2 > a_1 > a_8$   $a_{12} > a_{11} > a_{10} > a_9$
Intuition <sub>2t</sub> :	$a_6 > a_7 > a_5 > a_3 > a_4 > a_2 > a_1 > a_8$   $a_{11} > a_{12} > a_{10} > a_9$
Intuition <sub>3t</sub> [=G <sub>e7</sub> ]	$a_6 > a_5 > a_3 > a_7 > a_4 > a_1 > a_8 > a_2$   $a_{11} > a_{12} > a_{10} > a_9$

Zum Vergleich seien schon einige theoretisch bestimmte Präferenzordnungen angefügt, die im folgenden erläutert werden:

*Theoretische moralische Präferenzordnungen*

Utilitarismus [=G <sub>e1</sub> ]:	$a_1 = a_3 = a_5 = a_6 > a_8 > a_2 = a_4 = a_7$   $a_9 = a_{11} > a_{12} > a_{10}$
Leximin [=G <sub>e500</sub> =G <sub>e501</sub> =...]:	$a_6 > a_7 > a_5 > a_4 > a_3 > a_2 > a_1 > a_8$   $a_{12} > a_{11} > a_{10} > a_9$
G <sub>e10</sub> :	$a_6 > a_5 > a_7 > a_3 > a_4 > a_1 > a_2 > a_8$   $a_{12} > a_{11} > a_{10} > a_9$

Die intuitiven Präferenzordnungen sind nicht *utilitaristisch* — wenn ich hier einmal einen am Wohlbefinden orientierten Utilitarismus zugrundelegen darf. Nach den moralischen Intuitionen ist die *Glücksverteilung* nicht gleichgültig; eine bessere Glücksverteilung kann eine geringere Glückssumme aufwiegen:  $a_9$  wird von allen Versuchspersonen einhellig schlechter bewertet als  $a_{10}$ , und  $a_1$  schlechter als  $a_7$ , — trotz erheblich größerer Glückssumme. Anscheinend werden nach den moralischen Intuitionen Verbesserungen im unteren Wohlbefindensbereich höher bewertet als Verbesserungen im oberen Bereich. Ein krasser Fall ist die intuitive Vorordnung von  $a_{10}$  vor  $a_9$ ;  $s_2$  und  $s_3$  verändern sich hier nicht; aber für den Anstieg von  $s_1$  von -0,5 auf 0 wird hier der sehr viel größere Abstieg von  $s_4$  von 3,5 auf 2,0 in Kauf genommen.

*Leximin* ist eine ethische Präferenzordnung, die in Anschluß an Rawls' Maximinprinzip entwickelt wurde: An die oberste Stelle der Präferenzordnung gehören diejenigen sozialen Ordnungen, bei denen der jeweils am schlechtesten Gestellte am besten wegkommt. Sind nach diesem Kriterium noch einige Alternativen gleichwertig, so werden innerhalb dieser Teilmenge die Alternativen vorgeordnet, bei denen der jeweils am zweitschlechtesten Gestellte am besten wegkommt; usw. Nach *Leximin* ergeben sich die oben schon angeführten Präferenzordnungen.

*Leximin* bewertet, ähnlich wie die untersuchten Intuitionen, Verbesserungen für die Schlechtestgestellten höher als Verbesserungen für Bessergestellte. Aber *Leximin* tut dies in extremem Ausmaß: Selbst die aller kleinste Verbesserung für den am schlechtesten Gestellten wird riesigen Verbesserungen für Bessergestellte vorgezogen — und sei dies der am zweitschlechtesten Gestellte. *Leximin* ist in einem gewissen Sinn unökonomisch, indem dieses Prinzip nur auf den Schlechtestgestellten schaut und erst, wenn für diesen *überhaupt*

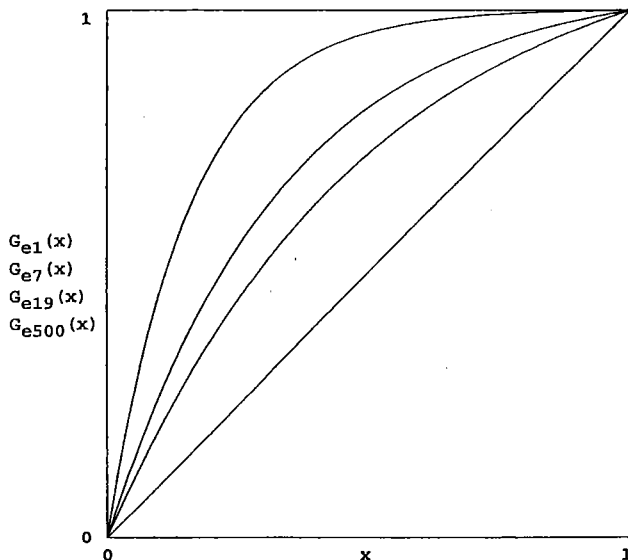
keine Verbesserungen mehr zu erreichen sind, sich um den am zweitschlechtesten Gestellten kümmert; dabei werden Verbesserungen, die für etwas besser Gestellte sehr leicht zu erzielen wären, vernachlässigt. Die getesteten moralischen Intuitionen sind ökonomischer: Sie gewichten Verbesserungen für die Schlechtergestellten zwar höher, vernachlässigen aber auch nicht die Bessergestellten. So liegt — im Gegensatz zu Leximin — in allen intuitiven Präferenzordnungen die Alternative  $a_3$  vor  $a_4$ : Man könnte zwar durch Übergang zu  $a_4$  den am zweitschlechtesten Gestellten von 1 auf 1,5 anheben, aber nur um den Preis, daß das Glücksniveau des am besten Gestellten von 3 auf 1,5 gesenkt wird; dieser Preis ist nach den moralischen Intuitionen zu hoch.

Mathematisch kann man die moralischen Intuitionen wie folgt erfassen: Statt wie im Utilitarismus sämtliche Wohlbefindensniveaus der Gesellschaftsmitglieder zu addieren, werden die Wohlbefindenswerte zunächst gewichtet und dann erst addiert; so gehen Verbesserungen im Wohlbefinden *aller* Beteiligten in die Summe ein; aber die Wohlbefindenswerte werden so gewichtet, daß sich Verbesserungen im unteren Wohlbefindensbereich in der Gesamtsumme stärker niederschlagen als Verbesserungen im oberen Bereich. Statt der Wohlbefindenswerte setzt man in die Summe den dem Wohlbefindenswert zugehörigen Wert aus einer Gewichtungsfunktion der in der folgenden Grafik dargestellten Art ein:

Die exponentielle Gewichtungsfunktion  $G_{ea}$

$$G_{ea}(x) = \frac{a}{a-1} \cdot (1 - a^{-x}), \text{ für } a > 1; G_{e1}(x) = x$$

Graphen der Funktion  $G_{ea}$  für  $a=1; 7; 19; 500$



Die Wohlbefindenswerte von -1 bis +4 sind hier positiv-linear in das Intervall  $[0;1]$  transformiert und auf der x-Achse eingetragen. Die Funktionswerte  $G_e(x)$  der Kurven ergeben dann den zugehörigen gewichteten Wert. An der untersten konkaven Kurve sieht man z.B., daß eine Verbesserung von 0 auf 0,1 der normierten Wohlbefindenswerte sich in einer Erhöhung des gewichteten Wertes um gut 0,2 niederschlägt, während eine Verbesserung von 0,9 auf 1,0 der normierten Wohlbefindenswerte sich in einer Erhöhung des gewichteten Wertes um 0,036 niederschlägt; sie wird also ungefähr 5,5 mal schwächer bewertet. Als solche Gewichtungsfunktionen kommen monoton steigende Funktionen in Frage, die vom Punkt  $\langle 0;0 \rangle$  zum Punkt  $\langle 1;1 \rangle$  verlaufen und oberhalb der Diagonale liegen. Die Diagonale selbst stellt den Utilitarismus dar. Je mehr die Kurve in die obere linke Ecke ausgebeult ist, desto stärker werden die Verbesserungen im unteren Bereich gegenüber denen im oberen Bereich gewichtet. Das andere Extrem ist Leximin; deren Gewichtungsfunktion ist im Extremfall der rechte Winkel oben links, also die Kurve, die von  $\langle 0;0 \rangle$  nach  $\langle 0;1 \rangle$  und von dort nach  $\langle 1;1 \rangle$  verläuft.

Als mathematische Beschreibung solcher Funktionen habe ich eine exponentielle Funktion gewählt:  $G_{ea}(x) = (a/a-1) \cdot (1-a^{-x})$ , für  $a > 1$ ;  $G_{e1}(x) = x$ . Diese Funktion erzeugt eine Kurvenschar in Abhängigkeit davon, was man für den Koeffizienten  $a$  einsetzt. Bei  $a=1$  entsteht die utilitaristische Gewichtungsfunktion; bei  $a=7$  bzw.  $a=19$  und  $a=500$  entstehen die in der obigen Grafik dargestellten Gewichtungsfunktionen. Der Koeffizient  $a$  gibt also das Maß des Engagements für die Unglücklichen an. Bei  $a=7$  wird die gleiche Präferenzordnung zwischen unseren Beispielalternativen erzeugt wie nach der Intuition<sub>3'</sub>; bei  $a=19$  und  $a=20$  wird die gleiche Präferenzordnung erzeugt wie nach der Intuition<sub>1'</sub>; und bei  $a=500$  entsteht die Leximinordnung. Ich habe bewiesen, daß es bei beliebigen Alternativenmengen immer ein entsprechend großes  $a$  gibt, ab dem die Gewichtungsfunktion  $G_{ea}(x)$  immer Leximinordnungen erzeugt. Die intuitiven moralischen Bewertungsfunktionen ( $G_{e7}$  oder  $G_{e19}$ ) liegen also in mathematisch präzisierbarer Weise zwischen Utilitarismus und Leximin. Ich habe sie deshalb „*Utilex*“ genannt. *Utilex* liegt nicht nur mathematisch zwischen Utilitarismus und Leximin, sondern ist auch eine *Synthese* aus beiden, die die Vorteile der beiden anderen Systeme vereint, ohne ihre Nachteile zu teilen: *Utilex* berücksichtigt Verbesserungen für alle Beteiligten und ist ökonomisch; *Utilex* bevorzugt aber die Schlechtergestellten. *Utilex* ist nicht unökonomisch und ignorant gegenüber dem Wohl der am besten Gestellten wie Leximin; und *Utilex* ist nicht ignorant gegenüber der *Verteilung* des Wohlbefindens wie der Utilitarismus.

## 2. Zur Begründung moralischer Prinzipien

Der erste Teil meines Beitrages hat vielleicht den Eindruck erweckt, hinsichtlich des methodischen Ansatzes der Moralbegründung sei ich Intuitionist. Dieser Eindruck wäre falsch. Mein Haupteinwand gegen einen solchen Intuitionismus ist, daß es keinen natürlichen moralischen Sinn oder ein natürliches Gewissen gibt. Vielmehr entstehen unsere moralischen Intuitionen sozial und sind deswegen historisch und intersubjektiv verschieden und schwanken intrasubjektiv. Moralische Intuitionen können deshalb nicht die Grundlage für eine Moralbegründung bilden.

Ich habe einen alternativen, formalen Ansatz zur Moralbegründung entwickelt, der in eine prudentialistische Richtung weist.<sup>2</sup> Der Ansatz beginnt mit der formalen Idee, daß eine triftige Begründung moralischer Prinzipien dasselbe ist wie eine gültige Argumentation für eine spezielle These über diese Prinzipien, für die *Moralbegründungsthese* nämlich. Wie man gültig und adäquat für eine bestimmte These argumentiert, sagt uns schon die Argumentationstheorie; dies ist keine spezifisch moralische Frage. Das eigentliche Problem ist: Welche der vielen möglichen Thesen über Moralprinzipien ist eine Moralbegründungsthese derart, daß man sagen kann: Wenn diese These bewiesen ist, dann sind die Prinzipien begründet?

Zur Beantwortung dieser Frage habe ich vier Adäquatheitsbedingungen für Moralbegründungsthese entwickelt:

*Adäquatheitsbedingung 1:* Der Gegenstand der Moralbegründung (das moralische Prinzip etwa) ist auch der Gegenstand der Begründungsthese. Andernfalls wäre die Begründung keine Begründung dieses Prinzips mehr.

*Adäquatheitsbedingung 2:* Die Begründungsthese für moralische Normen und moralische Handlungen sind in dem Sinne *motivierend*, daß, wenn ein Subjekt von diesen Thesen überzeugt ist, das Subjekt wenigstens ein Stück weit motiviert ist, die Normen bzw. Handlungen zu verwirklichen. Diese Forderung ist die *praktische* Komponente des Moralbegründungskonzepts. Sie soll die praktische Relevanz der Moralbegründung sichern, nämlich daß es bei der Moralbegründung nicht um *irgendeine* Einsicht über die Moral geht, sondern um eine praktisch verändernde.

*Adäquatheitsbedingung 3:* Die motivierende Wirkung der begründeten Überzeugung von einer Moralbegründungsthese geht durch zusätzliche wahre Informationen nicht verloren; d.h.: Wenn jemand von der Begründungsthese begründet überzeugt ist, dann gibt es keine wahre Information, für die gilt, wenn der Betroffene diese Information hätte, würde er entweder nicht mehr an die Begründungsthese glauben, oder er würde weiterhin an sie

<sup>2</sup> Christoph Lumer: Was ist eine triftige Moralbegründung? In: Georg Meggle; Ulla Wessels (Hg.): *Analyomen 1*. Berlin; New York 1994. S. 785-796.

glauben, aber dieser Glaube hätte keine motivierende Wirkung mehr. Diese *Aufklärungsstabilität der motivierenden Wirkung* ist die *epistemisch rationale* Komponente des Moralbegründungskonzepts. Sie sichert so etwas wie die *Weisheit* des Moralbegründungskonzepts, d.h., daß man einer bestimmten Moral nicht lediglich aus Naivität oder Uninformiertheit oder Fehlinformation anhängt.

*Adäquatheitsbedingung 4:* 1. Die nach diesem Konzept als 'begründet' ausgezeichneten Handlungen und Normen haben diejenigen formalen Eigenschaften, die von allen (entwickelteren) Ethiken übereinstimmend als moralisch relevant angesehen werden; 2. und es gibt von irgendwelchen Menschen schon angenommene, entwickeltere moralische Intuitionen, die ungefähr die gleichen Handlungen und Normen auszeichnen wie das Begründungskonzept. Diese *Intuitionskoinzidenz* ist die *moralische* Komponente des Moralbegründungskonzepts. Wenn sie nicht erfüllt ist, kann man von den durch das Begründungskonzept ausgezeichneten Handlungen und Normen nicht mehr sagen, sie seien *moralisch* begründet.

Diese Adäquatheitsbedingungen sind zum einen nicht trivial; die meisten Ethiken scheitern an der einen oder anderen Bedingung; dies gilt z.B. für die Diskursethik, eine Kantische Ethik oder den Intuitionismus. Zum anderen weisen diese Adäquatheitsbedingungen in eine prudentielle Richtung: Denn die prudentielle Handlungsbegründung kann man ähnlich konzipieren wie die Moralbegründung; als Argumentation für eine Begründungsthese über die Handlung; und die Adäquatheitsbedingungen für diese Begründungsthese decken sich z.T. sogar mit denen für die Moralbegründungsthese: Es muß eine These über die Handlung sein; die Überzeugung von dieser These muß zum entsprechenden Handeln motivieren; und diese Motivation muß auch aufklärungsstabil sein. Die Forderung nach moralischer Intuitionskoinzidenz entfällt natürlich bei einer prudentiellen Handlungsbegründung; ein weiterer wichtiger Unterschied ist, daß die Motivationsforderung bei der prudentiellen Handlungsbegründung viel stärker ist: Man muß nicht nur *ein Stück weit motiviert* sein, so zu handeln, sondern im allgemeinen dann *tatsächlich so handeln*.

Rein theoretisch könnte es aufklärungsstabile *intrinsische* Motive zu moralischem Handeln geben; dann wäre die Moralbegründung besonders einfach. Wie meine kurze Intuitionismuskritik schon andeutete, halte ich diese Möglichkeit für empirisch nicht gegeben. Sodann gibt es *extrinsische* Motive zu moralischem Handeln, wie etwa die Furcht vor Sanktionen. Solche Motive sind zwar wichtig für die moralische Motivation, sie helfen aber nicht bei der *inhaltlichen Bestimmung* der Moral, weil sie nur eine jeweils akzeptierte Moral *verstärken*. Dann bleiben als Grundlage der praktischen Moralbegründung nur amoralische, aber inhaltlich moralnahe Motive wie z.B. Mitgefühl, Soli-

darität, Achtung vor der Würde anderer oder das Interesse an bestimmten Arten der Kooperation.

Bei einem dieser Motive, dem Mitgefühl, möchte ich im folgenden untersuchen, welche Bewertungen von alternativen Gesellschaftsordnungen sich allein aus der Perspektive dieses Motivs ergeben. Anders gesagt: Ich suche im folgenden nach der *empathischen altruistischen Wünschbarkeitsfunktion*.

### 3. Die empathische altruistische Wünschbarkeitsfunktion

Ich habe ein vereinfachendes mathematisches Modell entwickelt, wie sich das Befinden anderer Personen, denen gegenüber wir keine speziellen Sym- oder Antipathien haben, in unserer erwarteten Empathie niederschlägt, also in der erwarteten Menge an Mitfreude und Mitleid. Kurz: Das Modell gibt Auskunft über den eigenen Empathieumfang in Abhängigkeit vom Wohlbefinden der anderen. Zu den vereinfachenden Annahmen des Modells, die ich hier nur anführen, aber nicht begründen kann, gehören u.a.: 1. Gegenstand des Mitgefühls ist das angenommene Wohlbefinden der Empathieobjekte. 2. Irrtümer in den Annahmen über das Wohlbefinden gleichen sich statistisch aus. 3. Untersucht wird hier nur ein *universelles* Mitgefühl, das sich auf mehr oder weniger Fremde richtet, für die man keine speziellen Sym- oder Antipathien empfindet und die man auch nicht moralisch bewertet. 4. In einer stark flexiblen Gesellschaft sind die Chancen, mit dem Schicksal anderer Gesellschaftsmitglieder konfrontiert zu werden, für alle Empathieobjekte gleich; die Aufdringlichkeit des Schicksals ist statistisch gleich verteilt. 5. Die Stärke des Mitgefühls hängt sicherlich von der Intensität und der Dauer der Beschäftigung mit dem einzelnen Schicksal ab; die Erwartungswerte dieser beiden Größen sind wieder intersubjektiv gleich. 6. Kluge Subjekte *haben* Mitgefühle und entziehen sich ihnen nicht.

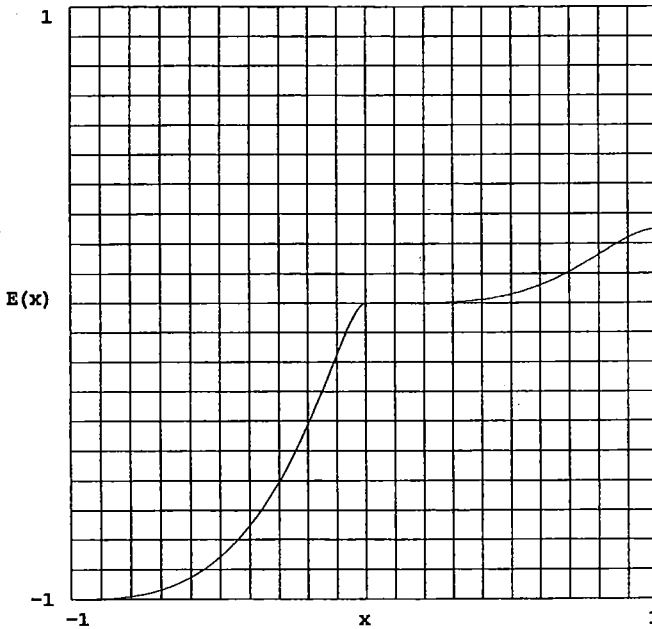
Der erste Schritt ist, die Intensität des Mitgefühls in Abhängigkeit vom angenommenen Befinden des Objekts zu bestimmen. Formal normiere ich das Wohlbefinden im Intervall  $[-1;1]$  mit 0 als Indifferenzpunkt, ebenso das Mitgefühl; negative Werte bedeuten also Mitleid, positive Werte Mitfreude. Inhaltlich mache ich folgende Annahmen: Die Empathiefunktion wächst monoton. Auf ein neutrales Befinden reagieren wir auch empathisch indifferent; die Funktion geht also durch den Punkt  $\langle 0;0 \rangle$ . Mitleid ist erheblich intensiver als Mitfreude; wie ich wieder bei einem kleinen Befragungsexperiment zur Tauschbereitschaft bei solchen Gefühlen ermitteln konnte, war das Mitleid gegenüber extremsten Leidenssituationen 4 bis 10 mal intensiver als bei extremstem Wohlbefinden; ich setze im folgenden den Faktor 4 an. Man erhält als Extrempunkte also  $\langle -1;-1 \rangle$  und  $\langle 1; 0,25 \rangle$ . Das normale Wohlbefinden liegt bei 0 bis 0,4; die empathische Reaktion auf dieses Normalbefinden ist minimal. Jenseits dieses Normalbereichs steigt die Empathie stark an,

beim Mitleid stärker als bei der Mitfreude. Bei der Annäherung an extreme Befindlichkeiten sättigt sich der Empathiezuwachs. — Aus diesen Annahmen erhält man folgende Empathiefunktion:

*Empathiefunktion  $E(x)$  in Abhängigkeit vom angenommenen Wohlbefinden des Empathieobjekts*

$$Es_s, (WBS_\mu, t=x) = E(x) =$$

1. für  $-1 \leq x \leq 0$  :  $-0,15 \cdot (x+1)^{23} + 1,15 \cdot (x+1)^3 - 1$ ;
2. für  $0 \leq x \leq 1$  :  $-0,25 \cdot x^8 + 0,5 \cdot x^4$ .



Dabei ist auf der  $x$ -Achse das Wohlbefinden des Objekts eingetragen — positive Werte geben Wohlbefinden wieder, negative Werte Unwohlsein -, auf der  $y$ -Achse ist die zugehörige Empathie verzeichnet — positive Werte bedeuten hier Mitfreude, negative: Mitleid. Wichtig ist an dieser Funktion vor allem: Das Mitleid ist erheblich intensiver als die Mitfreude; und auf normales Wohlbefinden, das zwischen den Wohlbefindenswerten 0 und 0,4 liegt, reagieren wir kaum mit Mitgefühl.

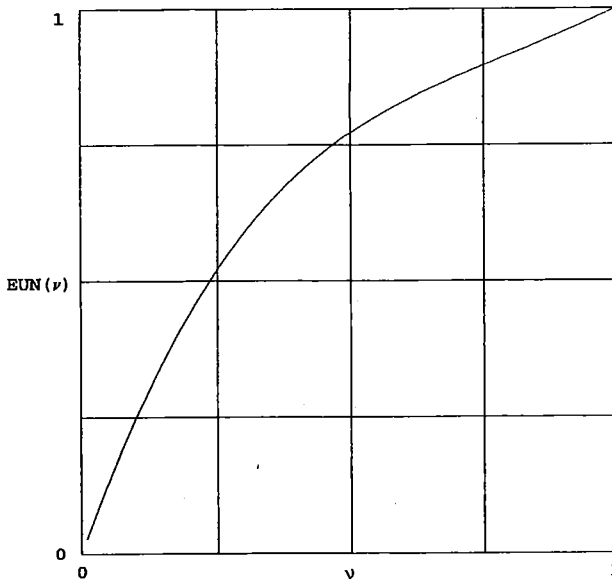
Der zweite Schritt ist, die *Verteilung* des Wohlbefindens bei verschiedenen Empathieobjekten zu ermitteln. Um nachher den zugehörigen Empathieumfang zu bestimmen, kommt es auf den genauen Verlauf der Wohlbefindenskurve nicht an, sondern nur auf die über das ganze Leben verteilte Dauer der einzelnen Wohlbefindenswerte. Wieder vereinfachend nehme ich an,



daß sich diese Wohlbefindenswerte nach einer Gaußschen Normalverteilung (Glockenkurve) verteilen. Diese Glockenkurve bedeutet, daß die Wohlbefindenswerte unter dem Gipfelpunkt der Kurve prozentual am häufigsten vorkommen und bessere und schlechtere Wohlbefindenswerte entsprechend seltener. Die offenen Parameter einer solchen Gaußschen Verteilung sind: 1. Welcher Wohlbefindenswert kommt am häufigsten vor?; d. h., über welchem Wohlbefindenswert liegt das obere Extrem der Verteilungskurve, was ist der Mittelwert ( $\mu$ )? 2. Wie steil oder flach ist die Kurve?; d. h., wie ist die Streuung ( $\sigma$ ) der Wohlbefindenswerte? Es gibt nun empirische Untersuchungen über das längerfristige subjektive Wohlbefinden; wenn man deren Werte in unsere Skala [-1; 1] umrechnet, dann liegen die Mittelwerte für die verschiedenen glücklichen Personen zwischen 0 und 0,4 ( $0 \leq \mu \leq 0,4$ ); die Glücklichen bringen es also über längere Zeit auf einen Mittelwert von 0,4. Ich nehme nun wieder vereinfachend an, daß sich zwar die Mittelwerte des Wohlbefindens bei Glücklichen und Unglücklichen intersubjektiv unterscheiden, nicht aber die Streuungen. Aus plausiblen Annahmen über die absolute Dauer einzelner extremer Glücksgefühle kann man die Streuung  $\sigma=0,16$  errechnen. Man erhält also eine Schar von Gaußschen Kurven, die jeweils die Wahrscheinlichkeiten aller möglichen Wohlbefindenswerte für typische Individuen angeben; alle diese Kurven haben die gleiche Form, aber die Mittelwerte liegen — je nach individuellem Glück oder Unglück — zwischen 0 und 0,4 (die Kurven sind entsprechend nach rechts oder links verschoben).

Der dritte Schritt ist, die Wahrscheinlichkeiten aus dieser Gaußschen Normalverteilung ( $P(WBs, t=x)$ ) mit der Empathiefunktion  $E(x)$  zu multiplizieren und über die Produktfunktion das Integral von -1 bis +1 zu bilden. Auf diese Weise erhält man dann den *Empathieumfang* (EU), also die Summe aller erwarteten Mitgefühle, die man für eine bestimmte Person empfindet, in Abhängigkeit vom mittleren Wohlbefinden ( $\mu$ ) dieser Person. ( $EU(\mu) = \int_{-1}^{+1} [P(WBs, t=x) \cdot E(x)] dx$ .) Bei einer Normierung der in einer Gesellschaft vorhandenen *mittleren* Wohlbefindenswerte und der resultierenden Empathieumfangswerte jeweils in das Intervall [0;1] ergeben sich folgende *normierte Empathieumfänge* (EUN) in Abhängigkeit vom normierten mittleren Wohlbefinden ( $v$ ):

### Normierte Empathieumfangsfunktion



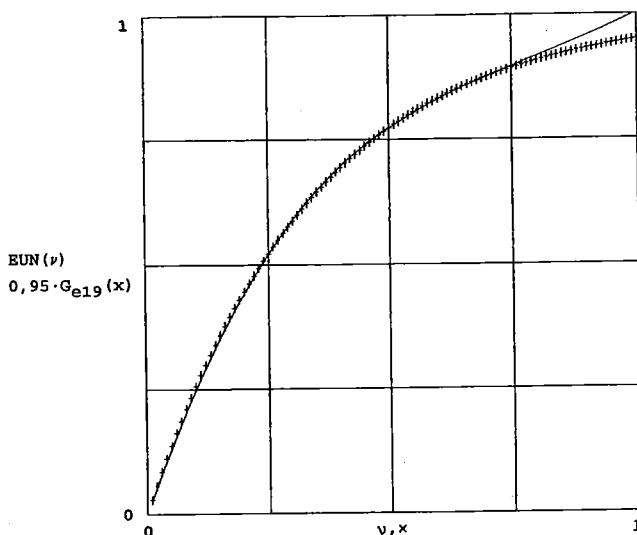
Wenn man nun eine Gesellschaftsordnung allein unter Empathiegesichtspunkten bewerten will, dann muß man zunächst das mittlere Wohlbefinden der einzelnen Mitgliedergruppen dieser Gesellschaft schätzen, anschließend aus der Empathieumfangsfunktion den zugehörigen Empathieumfang ermitteln und schließlich alle diese einzelnen Empathieumfänge addieren.

Dies ist aber genau dasselbe Verfahren wie bei der moralischen Bewertung nach dem Utilitarismus, Leximin oder Utilex. Der Unterschied ist nur, daß als Gewichtungsfunktion nun die Empathieumfangsfunktion EUN verwendet wird und nicht  $G_{e1}$ ,  $G_{e19}$  oder z.B.  $G_{e500}$ . Beim Vergleich dieser Funktionen zeigt sich aber, daß die Empathieumfangsfunktion EUN in frappierender Weise auf weite Strecken mit der Utilex-Gewichtungsfunktion  $G_{e19}$  übereinstimmt. (Um die Übereinstimmung visuell deutlicher zu machen, habe ich die Utilex-Gewichtungsfunktion um den Faktor 0,95 gestaucht.)

Vergleich Empathieumfangsfunktion und moralische Gewichtungsfunktion

Normierte Empathieumfangsfunktion  $EUN(v)$ : — ;

Moralische Gewichtungsfunktion  $0,95 \cdot G_{e19}(x)$ : +++.



Der Hauptunterschied zwischen beiden Kurven ist, daß die Empathieumfangsfunktion am äußersten rechten Ende wieder leicht nach oben gebogen ist. Diese Abweichung kann man als die Auswirkung eines empathischen Perfektheitsideals ansehen: Wir freuen uns besonders darüber, daß es einmal jemanden gibt, dem es nach dem, was Menschen erreichen können, nahezu perfekt geht. Dieses Perfektheitsideal ist allerdings ein recht schwacher Effekt; es führt nur zu einer Bevorzugung der am besten Gestellten vor ohnehin schon sehr, sehr gut Gestellten.

Abgesehen von diesem Effekt kann man also sagen: Die *prudentielle* Bewertung von alternativen Gesellschaftsordnungen unter rein empathischen Gesichtspunkten führt zum selben Resultat wie die intuitive moralische Utilex-Bewertung; jene ist eine prudentielle Rekonstruktion von dieser. Dies war auch für mich ein völlig überraschendes Resultat.

#### 4. Ethische Konsequenzen aus Utilex

Empathie ist nicht das *einzige* moralnahe Motiv; andere sind z.B. Solidarität, Achtung vor der Würde anderer oder Interesse an bestimmten Formen der Kooperation. Eines meiner Ziele ist, auch solche anderen Motive zu un-

tersuchen und zu ermitteln, welche Bewertungsfunktionen sich aus ihnen ergeben. Die Summe all dieser Bewertungsfunktionen könnte dann *die* prudentiell begründete moralische Bewertungsfunktion darstellen.

Nach dem im zweiten Teil meines Beitrags skizzierten Ansatz wäre solch eine summierte Bewertungsfunktion moralisch begründet; moralisch begründet wären auch Gesellschaftsordnungen und Handlungen, die nach dieser Bewertungsfunktion optimal wären. Neben den moralnahen haben wir aber auch noch ganz andere Motive; und es wäre aus der Perspektive des Handelnden sicher häufig *nicht klug*, das moralisch Optimale zu tun. Zudem ist es (u.a. weil es unter heutigen Bedingungen oft eine Zumutung wäre) auch nicht immer moralisch geboten, das moralisch Optimale zu tun — entgegen der Ansicht mancher deontischer Logiker. Was hat man dann mit der moralischen Bewertungsfunktion gewonnen? Moralische Normen sind nichts Überhistorisches, sondern werden erst historisch langfristig durchgesetzt; dabei werden die moralischen Standards im Laufe der Jahrtausende höhergeschraubt. Die soziale Durchsetzung moralischer Normen schließt auch ihre Bewehrung mit Sanktionen ein, die dann wieder zusätzliche prudentielle Gründe für die Normbefolgung liefern. Moralische Bewertungsfunktionen, wie ich sie eben rekonstruiert habe, hätten in diesem historischen Prozeß zwei Funktionen: Zum einen hätten sie eine *orientierende* Funktion: Sie könnten diejenigen Normen bestimmen, deren Durchsetzung ein moralischer Fortschritt wäre. Zum anderen hätten sie die *motivationale* Funktion, bei den vielen gegensätzlichen Interessen der Individuen, die zur Durchsetzung von Normen beitragen, einen Kern von gemeinsamen Interessen zu bilden, der die Grundlage für einen Kompromiß über durchzusetzende Normen bilden könnte.