

DER UTILITARISMUS UND SEINE KRITIKER

Beweise für den Utilitarismus? Präferenzutilitarismus (Stichpunkte zum
7.11.2006)

1 Der Hintergrund

Wir haben bis jetzt viele Argumente gegen den Utilitarismus kennengelernt. Vielleicht fragen sich manche: Warum der Utilitarismus, wenn es so viele Argumente gegen ihn gibt?

Richard M. Hare und J. Harsanyi haben sehr grundlegende Argumente für den Utilitarismus vorgelegt. Diese Beweise sollen hier kurz vorgestellt werden – wenigstens in Grundzügen. Sie führen auch zu einem neuen Begriff des Nutzens.

2 Individuelle Rationalität

Dazu wollen wir zunächst ein paar Überlegungen zur praktischen Rationalität eines Akteurs anstellen.

Wir wollen uns vorstellen, daß Peter vor einer konkreten Entscheidungssituation steht. Er hat mehrere Handlungsalternativen.

Peter zieht einige Alternativen anderen vor. Man spricht in diesem Zusammenhang von Präferenzen.

Oft wird nun angenommen, daß die Präferenzen eines rationalen Akteurs grundlegenden Rationalitätsforderungen genügen. Folgende Forderungen sind „natürlich“:

- Transitivität: Wenn Peter a_1 gegenüber a_2 vorzieht und a_2 gegenüber a_3 vorzieht, dann zieht er auch a_1 gegenüber a_3 vor.
- Vollständigkeit. Für alle Alternativen a_1 und a_2 gilt: Peter zieht entweder a_1 a_2 vor oder a_2 a_1 oder er ist zwischen beiden indifferent – es ist ihm gleichgültig, ob a_1 oder a_2 .

Unter einigen technischen Voraussetzungen gilt nun ein Repräsentationstheorem: Wenn Peter rational ist, dann kann man seine Präferenzen in einer Entscheidungssituation durch eine Funktion u darstellen. Es gilt: Peter zieht a_1 a_2 vor, wenn $u(a_1) > u(a_2)$. u heißt Nutzenfunktion. Wenn es eine Alternative a_j gibt, die Peter allen anderen Optionen vorzieht, dann hat a_j einen maximalen Wert von u . Wenn der Akteur also diejenige Option wählt, die er allen anderen vorzieht, dann maximiert er die Nutzenfunktion.

Dieses Repräsentationstheorem kann man auch für den Fall verallgemeinern, daß der Akteur nicht genau weiß, was geschieht, wenn er eine Option wählt. Um diese Rationalitätsforderungen auszubuchstabieren, müssen wir noch den Begriff einer Lotterie kennenlernen. Eine Lotterie l funktioniert grob wie folgt: Es gibt eine Reihe von möglichen

Ergebnissen (Preisen), wir nennen sie a_i . Jedes Ergebnis tritt mit einer bestimmten Wahrscheinlichkeit $p^l(a_i)$ auf. Natürlich addieren sich die Wahrscheinlichkeiten auf 1.

Es wird nun angenommen, daß Akteure nicht nur Präferenzen über die a_i haben, sondern daß sie auch Präferenzen über Lotterien haben.

Dann lauten die Forderungen der Rationalität bei der Lotterie-Wahl wie folgt:

- Transitivität: Wenn Peter Lotterie l_1 gegenüber l_2 vorzieht und l_2 gegenüber l_3 vorzieht, dann zieht er auch l_1 gegenüber l_3 vor.
- Vollständigkeit. Für alle Alternativen l_1 und l_2 gilt: Peter zieht entweder l_1 l_2 vor oder l_2 l_1 oder er findet beide gleich gut (Indifferenz).
- Unabhängigkeit. Nehmen wir an, Peter zieht l_1 gegenüber l_2 vor. Sei l_3 eine weitere Alternative. Wir bilden nun zwei neue Lotterien. Sei p eine Zahl zwischen 0 und 1. Die erste Lotterie sieht wie folgt aus: Peter erhält mit Wahrscheinlichkeit p l_1 und mit Wahrscheinlichkeit $(1 - p)$ l_3 .¹ Die zweite Lotterie sieht wie folgt aus: Peter erhält mit Wahrscheinlichkeit p l_2 und mit Wahrscheinlichkeit $(1 - p)$ l_3 . Unabhängigkeit heißt: Peter zieht nun die erste Lotterie gegenüber der zweiten vor.
- Archimedisches Axiom. Wir nehmen an, daß Peter l_1 gegenüber l_2 und l_2 gegenüber l_3 vorzieht. Es gibt nun eine Wahrscheinlichkeit p , so daß Peter zwischen l_2 und einer Lotterie indifferent ist, die Peter mit p l_1 und mit $(1 - p)$ l_2 gibt.

Wenn Peter diese Rationalitätsforderungen erfüllt, dann gilt das sogenannte v. Neumann-Morgenstern-Repräsentationstheorem. Es besagt ungefähr folgendes: Wenn Peter wie oben beschrieben rational ist, dann können wir den einzelnen Preisen a_i , vor denen er steht, Zahlen $u(a_i)$ zuordnen, so daß gilt: Peter zieht l_1 gegenüber l_2 vor, wenn und nur wenn $\sum_i p^{l_1}(a_i)u(a_i) > \sum_i p^{l_2}(a_i)u(a_i)$.

Das heißt, insofern Peter rational ist, maximiert er den Erwartungsnutzen.

Wenn das stimmt, dann gilt: Rationalität hat etwas mit Maximierung zu tun.

3 R. M. Hare über die Sprache der Moral

In seinem Buch „The Language of Morals“ (1952) hat R. M. Hare versucht, grundlegende Eigenschaften unserer moralischen Sprache zu beschreiben. Nach Hare zeichnet sich unsere moralische Sprache durch zwei Eigenschaften aus:

- Universalisierbarkeit (auch Supervenienz genannt): Wer dem Urteil „Peter sollte φ tun“ zustimmt, der stimmt auch dem Urteil zu, daß jeder, der sich einer Situation befindet, die der von Peter in allen relevanten Hinsichten ähnelt, φ tun sollte.
- Präskriptivität: Wer dem Urteil „Ich sollte φ tun“ zustimmt, der stimmt auch dem Imperativ „Laß mich φ tun“ zu und will demnach φ tun.

4 Ein moralisches Begründungsmodell auf der Basis von Hare

1. Die grundlegende Idee: Siehe den Text aus „Freedom and Reason“ von Hare als Basis.
2. Verallgemeinerung im Trompeter-Beispiel (FR, 7.1):

¹ Man spricht hier von einer zusammengesetzten Lotterie („compound lottery“).

„In dem von uns erörterten Beispiel wählte ich absichtlich einen Fall, wo B und A die *gleiche* Neigung hatten – nämlich die, nicht ins Gefängnis gesteckt zu werden; und obwohl es von mir nicht ausdrücklich betont wurde, so hatte dabei die Neigung natürlich nicht nur hinsichtlich ihres Gegenstands, sondern auch in ihrer Intensität die gleiche zu sein. Wir wollen uns jetzt einen komplizierteren Fall ansehen, wo die Neigungen der zwei Parteien nicht gleich sind. Um ein etwa triviales Beispiel aufzugreifen (das im täglichen Leben aber zu bösen Folgen führen könnte), nehmen wir an, A möchte auf seinem Plattenspieler gerne klassische Kammermusik anhören, während der im Zimmer nebenan lebende B sich eben überlegt, ob er auf seiner Trompete Jazz üben soll. Nun ist es offensichtlich für B witzlos, sich zu fragen, ob er bereit ist, die universelle Vorschrift aufzustellen, daß man Trompete spielen sollte, wenn man neben jemandem wohnt, der sich gerade Platten mit klassischer Musik anhört. Denn wenn B selbst gerade klassische Platten anhörte (die ihn unerträglich langweilen), dann würde er es nur zu gerne sehen, wenn jemand im Zimmer nebenan anfangen würde, Trompete zu spielen.

Der Ausweg aus dieser Schwierigkeit scheint sehr naheliegend zu sein. B hat sich nicht nur vorzustellen, daß er in A's Situationen mit seinen eigenen (B's) Wünschen und Abneigungen ist, sondern, daß er eben mit A's Wünschen und Abneigungen in A's Situation ist. Das Moralurteil, das er über diese Situation zu fällen hat, muß dabei jedoch sein eigenes (B's) bleiben, wie auch jedes andere präskriptive Urteil, das er fällt, wenn es für die Begründung eine Rolle spielen soll (6.9). Der natürliche Begründungsgang besteht für B also darin, zuzugeben, daß er nicht bereit ist, die universelle Vorschrift aufzustellen, jemandes Wünsche oder Abneigungen sollten von den anderen mißachtet werden – weil daraus logisch folgen würde, daß auch seine eigenen Wünsche und Abneigungen von anderen mißachtet werden sollten. [...]"

3. Weitere Verallgemeinerung in „Moralisches Denken“, Kapitel 6 (zitiert nach Glover, *Utilitarianism and Its Critics*):

„It follows from universability that if I now say that I ought to do a certain thing to a certain person, I am committed to the view that the very same thing ought to be done to me, were I in exactly his situation, including having the same personal characteristics and in particular the same motivational states. [...]"

A problem arises here, however, about this conflict between my own and my victim's preferences. There is first of all the difficulty, which we shall be dealing with in the next chapter, of comparing his preferences with mine in respect of intensity. How am I to say which is the greater, and by how much? But even if we assume that this difficulty can be overcome, the problem remains of why my preferences, even if they are less intense, should be subordinated to his. And if mine are more intense than his, ought they to be subordinated at all? Suppose, for example, that all I think I ought to do to him is move his bicycle so that I can park my car, and he has a mild aversion to my doing this [...]. This problem seems even more pressing in multilateral cases in which the preferences of many people are affected; but it will do no harm to deal with it in this simple bilateral conflict.

I can see no reason for not adopting the same solution here as we do in cases where our own preferences conflict with one another. For example, let us change the case and suppose that it is my own bicycle, and that it is moderately inconvenient to move it, but highly inconvenient not to be able to park my car; I shall then naturally move the bicycle, thinking that that is what, prudentially speaking, I ought to do, or what I most want, all in all, to do. Reverting now to the bilateral case: we have established, that, if I have full knowledge of the other person's preferences, I shall myself have acquired preferences equal to his regarding what should be done to me were I in his situation; and these are the preferences which are now conflicting with my original prescription. So we have in effect not an interpersonal conflict of preferences or prescriptions, but an intrapersonal one; both the conflicting preferences are mine. I shall therefore deal with the conflict in exactly the same way as with that between two original preferences of my own.

4. Harsanyi (1977, *Social Research* 44, 4, nachgedruckt in Sen und Williams): Diskutiere interpersonelle Situation intrapersonal und maximiere den Nutzen (Erwartungsnutzen): Annahme: Es ist gleich wahrscheinlich, daß ich jede der beteiligten Personen bin. Ergebnis: Maximiere den Nutzen.

5 Das Harsanyi-Theorem

Ein alternativer Beweis für den Utilitarismus wurde von J. Harsanyi gefunden (siehe dazu wiederum Harsanyi 1977).

Annahmen: Alle Menschen haben persönliche und moralische Präferenzen.

- Die persönlichen Präferenzen jeder Person unterliegen den Rationalitätsforderungen oben.
- Die moralischen Präferenzen einer Person i unterliegen den Rationalitätsforderungen oben.
- Für die moralischen Präferenzen dieser Person gilt: Wenn eine Person persönlich l_1 gegenüber l_2 vorzieht, und niemand l_2 gegenüber l_1 vorzieht, dann zieht i l_1 gegenüber l_2 moralisch vor.
- Die moralischen Präferenzen von i machen keinen Unterschied zwischen unterschiedlichen Personen.

Es folgt: Man kann die moralischen Präferenzen von i durch eine Nutzenfunktion darstellen, die einfach die Nutzenfunktionen der einzelnen Menschen aufaddiert. i ist daher ein Utilitarist.

6 Der Präferenzutilitarismus

Wir haben eine neue Begründung für den Utilitarismus auf der Basis „vernünftiger“ Axiome. Allerdings ändert sich die Interpretation des Nutzens. Der Nutzen wird jetzt durch die Präferenzen von Personen bestimmt, und zwar unabhängig davon, ob diese Präferenzen der Person Freude etc. machen.

7 Literatur

<http://catalogue.polytechnique.fr/Files/gamelecture1.pdf>

Berger, James O., *Statistical Decision Theory and Bayesian Analysis*. Zweite Auflage
New York 1985.

Harsanyi, in Williams and Sen.

Hare, R. M., *The Language of Morals* (1952), *Freedom and Reason* (1963, dt. 1983
Frankfurt am Main), , *Moral Thinking* (1981)