Kritik des Bayesschen Konsequentialismus

Eine einführende Zusammenschau

1. Einleitung]

Die konsequentialistische Auslegung von Handlungsrationalität hat durch die Entwicklung der Entscheidungs- und Spieltheorie seit Mitte des vergangenen Jh., genauer: mit dem Erscheinen der Monographie von Morgenstern und von Neumann im Jahre 1944, mit welcher sie den Grundstein für die beiden Disziplinen gelegt haben,[1 ein theoretisches Fundament erhalten, mit dem sich jede Kritik des Konsequentialismus auseinandersetzen muss, will sie sich nicht dem Vorwurf aussetzen, nur für ältere und schwächere Formen des Konsequentialismus relevant zu sein (Nida-Rümelin 1995, S. 36).

Mittlerweile ist die Spiel- und Entscheidungstheorie jedoch alt genug, um schon einen beträchtlichen Grad der Wucherung erreicht zu haben, weshalb es das, was wir untersuchen wollen, etwas genauer einzugrenzen gilt. Im Fokus unseres Interesses liegen nicht periphere Verästelungen der erwähnten Teilwissenschaften, sondern vielmehr das Grundprinzip der klassischen Nutzenerwartungswerttheorie, das unter dem Namen Bayes-Kriterium - in Hommage an die bedeutenden Beiträge des britischen Mathematikers zur Wahrscheinlichkeitsrechnung - bekannt ist. Das Bayessche Kriterium besagt:
Diejenige Entscheidung ist rational, die den (subjektiven) Erwartungswert des (subjektiven) Nutzens des Handelnden maximiert. (Nida-Rümelin 1995, S. 36)

In der vorliegenden Abhandlung geht es also um ein Thema, das sowohl dem Bereich der praktischen Philosophie wie auch dem der theoretischen Ökonomie entnommen ist. Genauer gesagt verfolgt diese Arbeit das Ziel, aus philosophischer Perspektive eine Rekonstruktion, eine Einordnung und eine erste kritische Stellungnahme zum Bayesschen Modell zu liefern, das Begrifflichkeiten und Kriterien zur Verfügung stellt, um die Rationalität individueller Entscheidungen zu bewerten. Kurzum: Der Bereich der praktischen Philosophie, dem diese Arbeit zugehört, ist demnach die Theorie der praktischen Rationalität.

Im folgenden Abschnitt (Kapitel 2) wird aufbauend auf dem Bayesschen Kriterium die klassische Nutzenerwartungswerttheorie

dargestellt.[2 Dieses Bayessche Modell kann in diesem Rahmen allerdings nur grob umrissen und nicht ausführlich skizziert werden. Ebenso ist es hier nicht möglich, auf die verschiedenen (teils konkurrierenden) Konzeptionen einzugehen, aus deren Axiomatik sich jeweils eine Erwartungsnutzentheorie herleiten lässt. Stattdessen beschränken wir uns auf eine Formulierung, die sich an den Ausführungen von Nida-Rümelin (1995) orientiert, wobei der formale Teil demgegenüber in etwas veränderter (hoffentlich vereinfachter) Form wiedergegeben ist. Trotz der Einschränkung sollte die aus Kap. 2 gewonnene Einsicht ausreichen, um das für unsere Zwecke Wesentliche entnehmen zu können, nämlich ein starkes Argument für den Konsequentialismus als adäquate Theorie der Entscheidungsrationalität. Gleichzeitig ergibt sich, was in Kap. 3 festzuhalten ist, dass das Prinzip der Nutzenerwartungswertmaximierung nicht als Entscheidungsregel oder -kriterium, sondern vielmehr als Adäquatheitstest für Axiome zu verstehen ist, die individuelle Präferenzen beschränken. Da darum ferner nur die Axiome bei der Beurteilung der Erwartungsnutzentheorie bzgl. ihrer Anwendbarkeit und Normativität interessieren können, schließt sich im letzten Teil (Kap. 4) eine kritische Beleuchtung derselben an. Aus Gründen des Umfangs geschieht dies zum einen in vornehmlich programmatischer Form und zum anderen begnügen wir uns hier mit zwei prinzipiellen Einwänden (im Gegensatz zur Gruppe der empirischen Einwände).

2. Darstellung des Bayesschen Konsequentialismus

Um das Bayessche Prinzip der Maximierung des erwarteten subjektiven Nutzens besser einordnen zu können, betrachten wir ein Grundmodell für die Nutzenerwartungswerttheorie. Diese Reduktion der verschiedenen Modelle auf ihren Kern, der sich in dem Grundmodell widerspiegelt, wird sich im Verlauf der Arbeit noch als strategischer Vorteil entpuppen, weil man somit die Kritik am Bayesschen Konsequentialismus nicht als belanglos oder nebensächlich abtun kann.

In dem Grundmodell gehen wir davon aus, dass der Entscheider - eine bzgl. des Modells ideale rationale Person - (i) in jeder Situation aus einer bestimmten (nicht leeren) Menge von Entscheidungsalternativen zu wählen hat, die jeweils alle ergreifbaren Handlungsoptionen, einschließlich derjenigen, so fortzufahren wie bisher, enthält und wobei sich die jeweiligen Alternativen gegenseitig ausschließen. Formal können wir die in einer Entscheidungssituation zur Verfügung stehenden Alternativen darum in Form der Alternativenmenge F repräsentieren, deren Elemente mit f1, f2, ? bezeichnet werden:
F = {f1, f2, ?}
Ferner sind die Elemente von F im formalen Kontext als Funktionen von der Zustandsmenge Z in die Konsequenzenmenge C zu begreifen, was unserer Intuition gerecht wird, dass Handlungsalternativen unter verschiedenen Umständen verschiedene Konsequenzen haben. Folglich kann jedes f ? F geschrieben werden als

f : Z ? C
z ? f (z)

wobei C die möglichen Konsequenzen von Handlungen erfasst (also: C = {c1, c2, ?}) und Z eine Menge von Weltzuständen (möglichen Welten) ist. Um allerdings eine im intuitiven Sinne realistische und adäquate Modellierung einer Entscheidungssituation zu erzielen, sollten (bspw. in einer Konsequenzenmatrix) statt der Weltzustände die nur jeweils relevanten Umstände angegeben werden, was Savages Differenzierung zwischen einer hypothetischen Menge von so exakt wie möglich formulierten Beschreibungen von Entscheidungssituationen, die er the grand world nennt, und einer für das jeweilige Entscheidungsmodell tauglichen (Zustands-)Menge, die er small world nennt, nahekommt. Wenn ich also etwa an einem Morgen vor der Entscheidung stehe, meinen Regenschirm mitzunehmen oder ihn zuhause zu lassen, so mache ich das in erster Linie davon abhängig, ob vor meiner Haustür Regen fällt oder nicht respektive ob für die Orte und Dauer meines Fußwegs ein Regenguss wahrscheinlich ist oder nicht. Es ist mir dabei gleichgültig, ob es in China regnet oder ob dort ein Sack Reis umfällt. In meiner konkreten Entscheidungssituation kümmern mich lediglich die Umstände, die für mich von Belang sind - nämlich z.B. A: "Regen in Vaduz" und B: "kein Regen in Vaduz" - und nicht die sämtliche Einzelheiten umfassenden Weltzustände, in welchen sowohl berücksichtigt ist, ob es in Vaduz regnet, als auch, ob in China ein Sack Reis umfällt. Denn mir kommt es in diesem Zusammenhang bloß darauf an, ob ich nass werde oder trocken bleibe. Sinnvollerweise lassen sich die verschiedenen Elemente der Entscheidungssituation in einer Konsequenzenmatrix bündeln:

A: Regen in Vaduz
B: Kein Regen in Vaduz

f1: Schirm mitnehmen
bleibe trocken
bleibe trocken und habe mir zu viel Arbeit gemacht

f2: Schirm daheimlassen
werde nass
bleibe trocken

In den Spalten sind die verschiedenen Umstände abgetragen und die Zeilen repräsentieren die Alternativen. Die Einträge in der Tabelle sind die Konsequenzen, die sich bei der jeweiligen Alternative unter dem entsprechenden Umstand ergeben. Genauer handelt es sich um Beschreibungen der Konsequenzen unter Gesichtspunkten, die dem Entscheider wichtig sind. Freilich ist die Situation, die eintritt, wenn es (in Vaduz) regnet und ich einen Schirm bei mir habe, eine andere als die, die eintritt, wenn kein Regen fällt und ich keinen Schirm dabeihabe. Mit Bezug auf das Nasswerden sind die Konsequenzen jedoch gleich: ich bleibe trocken.

Analog zur Menge F sei die Menge der Umstände derart beschaffen, dass nicht zwei Umstände zugleich eintreten können, dass aber das Eintreten von mindestens einem sichergestellt ist. Formal lässt sich dies bewerkstelligen, indem die Menge der in einer konkreten Entscheidungssituation relevanten Umstände als eine Partition von Z notiert wird,[3 sodass jeder Umstand A, B, ? eine Menge von Elementen von Z ist. Es ist zwar mitunter sinnvoll, die Umstände als sich nicht ausschließend anzusehen, doch wollen wir es hier bei einem intuitiven Vorverständnis der Umstände belassen, die formal in der Umstandsmenge Q zusammengefasst werden:
Q = {A, B, ?}

Damit haben wir mit Ausnahme des Begriffs der individuellen Präferenz die wesentlichen Begriffe, die eine Entscheidungssituation charakterisieren, zusammengestellt. Darüber hinaus hat der Entscheider (ii) zu jedem Zeitpunkt eine kohärente subjektive Wahrscheinlichkeitsverteilung, die angibt, für wie wahrscheinlich er das Eintreten eines Umstandes (bspw. von A) hält:
P : Q ? [0, 1] respektive P : 2Z ? [0, 1]
A ? P (A)

Da wir oben die Zustandsmenge Z als endlich angenommen haben, kann eine Wahrscheinlichkeitsfunktion auf der gesamten Potenzmenge von Z, also für alle Ereignisse A ? Z, definiert werden. Daneben wird dem Entscheider (iii) eine (wohldefinierte) Nutzenfunktion Ut zugeschrieben, die bis auf lineare Transformation eindeutig bestimmt ist und die jeder Konsequenz eine reelle Zahl als kardinales Maß der subjektiven Vorlieben und Wertungen des Entscheiders ("Nutzen der Konsequenz") zuordnet.

Ut : C ? ?
c ? Ut (c)

Mit diesem erweiterten Begriffsapparat lässt sich nun das Grundprinzip der Nutzenerwartungswerttheorie reformulieren - eine Formulierung, die sich später als missverständlich herausstellen wird:

? Ut (f(z)) × P {z})
z ? Z

Schließlich (iv) wählt der Entscheider in jeder Situation diejenige Alternative, für die der Erwartungswert des Nutzens maximal ist (Bayes-Kriterium), wobei sich für eine Alternative f ? F der Erwartungswert des Nutzens berechnet zu:

Demnach wähle ich bspw. in der obigen Schirm-mitnehmen-daheimlassen-Entscheidungssituation für eine Regenwahrscheinlichkeit (in Vaduz) von 70 % die Alternative f1: "Schirm mitnehmen", wenn Folgendes gilt:
0,7 × Ut (bleibe trocken) + 0,3 × Ut (bleibe trocken und habe mir zu viel Arbeit gemacht) > 0,7 × Ut (werde nass) + 0,3 × Ut (bleibe trocken).

Halten wir fest: Unser Modell verlangt viel von einem rationalen Entscheider. Es ist ausmachbar, dass die Anwendung des Bayesschen Modells "sowohl als handlungsleitende normative Theorie, wie als analytisches Instrument zur »rationalen Rekonstruktion« menschlichen Entscheidungsverhaltens Schwierigkeiten macht" (Nida-Rümelin 1995, S. 37). Um nur einige wenige Zumutungen zu nennen, sei etwa auf Folgendes verwiesen: (i) Da eine konsistente Bewertung von Konsequenzen die kausalen Verknüpfungen von Ereignissen zu berücksichtigen hat, ist von unserem Entscheider gefordert, ganze Weltverläufe in seiner Bewertung zu bedenken. (ii) Weiterhin muss er imstande sein, den mit einer Handlung verbundenen vorstellbaren Weltverläufen konsistente Wahrscheinlichkeiten zuzuordnen. Und (iii) stellt sich natürlich die Frage, wie man Personen überhaupt solche Nutzen- und Wahrscheinlichkeitsfunktionen zuschreiben kann. Es mag darum bisher nicht einleuchten, dass derart abstrakte Modelle ein starkes Argument für den Konsequentialismus als adäquate Theorie der Entscheidungsrationalität stiften.

Zumindest ist aber zu betonen, dass die konsequentialistischen Theorien vom Typus der Nutzenerwartungswert-Konzeptionen ihre Rechtfertigung in der Gestalt einer Reihe von (plausibel anzunehmenden!) Postulaten erfahren. Und tatsächlich ist das Prinzip der Nutzenerwartungswertmaximierung logisch äquivalent zu der Konjunktion einiger Mindestbedingungen an die Präferenzen einer rationalen Person.[4 Das bedeutet, dass die Zurückweisung unseres Grundmodells bzw. eines der zahlreichen Modelle die Ablehnung mindestens eines der jeweiligen Postulate erfordern würde. Der Kern der verschiedenen Axiomatiken besteht nun darin, eine auf bestimmten Objekten definierte individuelle zweistellige (schwache) Präferenzrelation - hier (und bei Savage) ist sie auf der Menge der Handlungen erklärt - derart durch Rationalitätsanforderungen bzw. Mindestbedingungen zu beschränken, dass die Präferenzen durch den erwarteten Nutzen repräsentiert und somit kardinalisiert werden. Auf diese Weise ergibt sich, mithilfe des soeben angerissenen Repräsentationstheorems (das zentrale Theorem der ökonomischen Nutzentheorie), eine elegante Explikation des Nutzenbegriffs.[5 Wenden wir uns im Lichte dessen erneut dem Bayesschen Prinzip zu:

Es seien eine Nutzenfunktion Ut, eine Wahrscheinlichkeitsfunktion P und individuelle Präferenzen über den Alternativen gegeben, wobei "f1 ? f2" heißt "f1 wird gegenüber f2 schwach vorgezogen" (bzw. "die Person P findet f1 mindestens so gut wie f2"). Sodann

ist das Bayessche Prinzip der Maximierung des Nutzenerwartungswerts erfüllt, gdw. für alle Handlungen f1 und f2 Folgendes gilt: f1 ? f2 ? ?z ? Z Ut (f1(z)) × P({z}) ? ?z ? Z Ut (f2(z)) × P({z})

Diese Bisubjunktion, der Dreh- und Angelpunkt des Bayesschen Konsequentialismus, besagt demnach, dass einer Person P unter bestimmten zu spezifizierenden Bedingungen eine Wahrscheinlichkeitsfunktion über einer (sinnvoll zu wählenden) Umstandsmenge sowie eine Nutzenfunktion über der Konsequenzenmenge zuzuschreiben sind, sodass P "ihren" erwarteten Nutzen maximiert, gdw. sie im Einklang mit ihren Präferenzen handelt. Anders ausgedrückt: Jeder, der solche wie etwa die im Anhang für unser Grundmodell genannten Postulate als Mindestbedingungen idealen rationalen Verhaltens versteht, ist dazu angehalten, aus logischen Gründen zu akzeptieren, dass eine rationale Person P in ihrem Entscheidungsverhalten die Bewertungsfunktion Ut maximiert (oder zumindest danach strebt). Das heißt aber auch: Jeder, der diese Postulate für vernünftige und plausible Bedingungen rationalen Verhaltens hält, muss die Bayessche Fassung des Konsequentialismus annehmen. Es bleibt daher zu hinterfragen, ob die Axiome der zahlreichen Konzeptionen intuitiv Zuspruch finden respektive sich in empirischen Überprüfungen behaupten. Wir wollen hier hingegen eher prinzipielle Kritik an der axiomatischen Nutzentheorie üben. Doch zunächst seien einige Ergänzungen zur konsequentialistischen Deutung des Bayesschen Modells gegeben.

3. Bemerkungen zur konsequentialistischen Interpretation]
Die übliche Interpretation des Bayesschen Kriteriums lautet: Eine rationale Person P entscheidet sich für eine Handlung f1, weil f1 den Erwartungswert ihres subjektiven Nutzens maximiert. Das Bayessche Prinzip entspricht dahingehend einem Entscheidungskriterium. Der konsequentialistische Charakter dieser Interpretation wird deutlich, wenn wir uns ins Gedächtnis rufen, dass wir von einer Nutzenfunktion ausgegangen sind, die über der Konsequenzenmenge deklariert ist. Insofern macht sich das kardinale Maß P's subjektiver Vorlieben und Wertungen, welches in Form der Nutzenfunktion nach dem Bayesschen Kriterium maximiert werden soll, allein an der Menge der Konsequenzen fest.

Allerdings ist es im Rahmen der axiomatischen Nutzentheorie falsch zu glauben, dass mit dem Nutzenerwartungswertkriterium ein Hilfsmittel zur Entscheidungsfindung an die Hand gegeben ist. Dies ist deswegen ein Trugschluss, weil zur Bildung der Nutzenfunktion bereits die individuellen Präferenzen über Handlungsalternativen vorliegen müssen, diese also nicht aus der Erwartungsnutzentheorie folgen können. Das bedeutet, dass die Präferenzen schon gegeben sind und damit auch der Ausgang der Entscheidungsfindung (Input), weil eben eine rationale Person gemäß ihren (den Axiomen genügenden) Präferenzen handelt; das zentrale Theorem der ökonomischen Nutzentheorie vollbringt dann "nur" noch die Leistung, Zahlenwerte Ut (fi (zi)) anzugeben, sodass der mathematische Erwartungswert dieser Zahlen die Präferenzordnung über Handlungsalternativen wiedergibt/repräsentiert (Output), sofern die Präferenzen einer Person gewisse Bedingungen (siehe etwa die Postulate im Anhang) erfüllen. Mit anderen Worten: Wenn ich bei der Wahl zwischen dem Kauf von Erdbeer- oder Zitroneneis bereits im Vorhinein weiß, dass ich eine Präferenz für Erdbeereis habe, dann brauche ich für die Entscheidung, Erdbeereis zu kaufen, nicht noch reelle Zahlen als kardinales Maß meiner subjektiven Wertung von Erdbeereis bzw. Zitroneneis; denn sie drücken auch bloß meine Vorliebe für Erdbeereis aus.

Der axiomatisch gewonnene Begriff des kardinalen Nutzens ist zunächst einmal nicht mehr als ein theoretisches Werkzeug, das im Hinblick auf ein Entscheidungsmodell eine Rolle spielt. Wer somit davon redet, dass Personen, Unternehmen oder Institutionen ihren erwarteten Nutzen maximieren, trifft eine Wortwahl, die bloß im Kontext einer relativ umfangreichen formalen Theorie zu konkreten Handlungsanweisungen bzw. ?bewertungen führt. Da also das Prinzip von der Maximierung des Nutzenerwartungswerts nur von Theorievokabular Gebrauch macht und keine Rationalitätsanforderungen an beobachtbares Verhalten stellt, scheint die einzig sinnvolle Interpretation, die man geben kann, die folgende zu sein: Das Bayessche Prinzip ist nicht als Entscheidungsregel, sondern vielmehr als Adäquatheitstest für Axiome auszulegen, die Mindestbedingungen an individuelle Präferenzen widerspiegeln.

Wenn dies richtig ist, dann ist die Üblichkeit, das Bayessche Prinzip als das zentrale Entscheidungskriterium klassischer Entscheidungstheorie anzusehen, in vielen Fällen mindestens irreführend, wenn nicht gar falsch! Denn da das Prinzip der Nutzenerwartungswertmaximierung logisch äquivalent zu der Verknüpfung einiger individuelle Präferenzen beschränkender Rationalitätspostulate ist und da ersteres nicht, letztere schon Beobachtungsvokabular enthalten, können nur letztere Postulate zur Beantwortung etwa der Frage, ob Individuen Nutzenerwartungswertmaximierer sind, herangezogen werden. Im Übrigen sollte man

dann auch bei der entscheidungstheoretischen Beschreibung von rationalen Individuen nicht mehr von Nutzenerwartungswertmaximierern, sondern besser von Axiomenerfüllenden sprechen.

Schließlich gilt es noch zu konstatieren: Keines der Rationalitätspostulate und keine Kombination dieser Postulate impliziert, dass eine rationale Person P Handlungen unter dem Gesichtspunkt ihrer Folgen beurteilt. Ob die jeweiligen Postulate erfüllt sind, hängt ausschließlich von der Gestalt der Präferenzrelation ab. Eine konsequentialistische Theorie der Handlungsrationalität lässt sich infolgedessen entgegen erstem Augenschein mit dem Repräsentationstheorem nicht rechtfertigen. Die konsequentialistische Standardinterpretation des Bayesschen Modells wird durch die zugehörige Axiomatik, durch die es wie wir angedeutet haben überhaupt erst seine Legitimation erfährt, nicht gestützt.

Summa summarum: Da wir nicht das Nutzenerwartungswertprinzip, sondern vielmehr die Rationalitätsanforderungen als für die Anwendungsfrage relevant ausgewiesen haben, werden diese und daher auch der Begriff der individuellen Präferenz im Zentrum unseres Interesses stehen müssen. Weiterhin wurde darauf hingewiesen, dass die konsequentialistische Auslegung des Bayesschen Modells auf der bloßen Grundlage der Mindestbedingungen an die Präferenzen einer rationalen Person P nicht logisch zwingend ist, doch dies allein wäre kein hinreichender Grund, den Bayesschen Konsequentialismus zu verwerfen. Es bleibt darum aufzuzeigen, dass er als Rationalitätskonzeption inadäquat ist.

4. Kritik des Bayesschen Konsequentialismus

Conceptually, the SEU [= subjective expected utility, C.H.] model is a beautiful object deserving a prominent place in Plato's heaven of ideas. But vast difficulties make it impossible to employ it in any literal way in making actual human decisions.
(Simon 1983, S. 13)

Viele kritische Stimmen pflichten Simon bei. Um diesen Kritiken auszugs- und ansatzweise Rechnung zu tragen, widmet sich dieses vierte Kapitel zwei grundsätzlichen Einwänden an der Konzeption der Nutzenerwartungswertmaximierung. Allgemein lassen sich die in der Debatte vorgebrachten Einwände in zwei Gruppen aufteilen: Die empirischen Einwände laufen darauf hinaus, dass sich aufgrund von Beobachtungen und Experimenten erkennen lässt, dass Menschen wesentliche und für die Erwartungsnutzenmaximierung notwendige Rationalitätspostulate systematisch verletzten.[6 Daneben gibt es eine Gruppe von eher prinzipiellen Einwänden. Es wird hier daraus ein Interpretationsproblem dargestellt, aus dem zu folgen droht, dass Rationalitätsprinzipien "empirisch leer" sind, dass sie also - und mit ihnen auch die jeweiligen Entscheidungsmodelle - den normativen Anspruch nicht einzulösen vermögen (4.2.). Aber zuvor bringen wir die Bemerkungen von Amartya Sen als Illustration für das, was in dieser Arbeit als "Reduktionismus-Vorwurf" bezeichnet wurde (4.1.).

4.1. Der Reduktionismus-Vorwurf

Wir lernen in diesem Abschnitt in Anlehnung an Sen ein Argument dafür kennen, dass es sinnvoll ist, die Grundbegrifflichkeiten der formalen Theorie individueller Entscheidungen gegenüber dem Begriffsapparat, der den verschiedenen axiomatischen Modellen vom Theorientypus der Nutzenerwartungswert-Konzeptionen zugrundeliegt, zu erweitern.

Machen wir uns klar, dass die Nutzenerwartungswerttheorie entwickelt wurde, um das Verhalten von Menschen zu beschreiben (ausgelegt als deskriptive Theorie) und/oder zu bewerten (ausgelegt als normative Theorie). Auch wenn die Theorie in der Form, in welcher sie in dieser Arbeit präsentiert wurde - nämlich in der Form einer axiomatischen Theorie - keine motivationale Komponente hat, sondern ausschließlich den individuellen Präferenzen Kohärenzbedingungen auferlegt, so ist doch der historische und ebenso der intuitive Ursprung der Axiomatisierung des subjektiven Nutzens in der These zu sehen, dass Handlungsakteure ihr Eigeninteresse verfolgen. Sofern man dann die moderne Nutzentheorie derartig interpretiert, dass ein axiomatisch fundierter Nutzenbegriff mit dem Begriff der individuellen Wohlfahrt gleichzusetzen ist - eine Interpretation, die angesichts der in Kap. 3 gemachten Bemerkungen zu weit geht -, dann kann man der Auffassung sein, dass altruistische Erwägungen im Zusammenhang mit der Erwartungsnutzentheorie verfehlt sind. Sen fasst diesen Gedanken in dem Ausdruck "definitional egoism" (Sen 1977, S. 323) ein. Dies scheint jedoch eine Auffassung zu sein, die nicht mit einer plausiblen Interpretation der formalen Entscheidungstheorie

kompatibel ist: Es ist problemlos möglich, die Objekte (z.B. Handlungsalternativen), auf denen die Präferenzen einer Person erklärt sind, in ihrer Beschreibung so zu modifizieren bzw. so zu verfeinern, dass das Wohlergehen anderer Personen berücksichtigt ist.[7 Folglich ist rational choice etwa mit altruistischen Erwägungen verträglich.[8

Und mehr noch: Selbst wenn man eine Vielzahl an Aspekten hat, die bei der Entscheidungsfindung relevant sind - Sen nennt v.a. "sympathy" und "commitment" -, kann man diese offenbar innerhalb einer auf individuelle Präferenzen (über geeignet individuierte Alternativen) bezugnehmenden formalen Entscheidungstheorie rekonstruieren. Es ist allerdings fraglich, ob dies methodologisch sinnvoll ist: Die Reduktion einer Pluralität von Aspekten, Überlegungstypen usw. auf eine einzige Präferenzordnung scheint eine tour de force zu sein, der man sich nicht anschließen muss. Sen schreibt polemisch:
A person is given one preference ordering, and as and when the need arises this is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choice and behavior. Can one preference ordering do all these things?
(Sen 1977, S. 335f.)

Der Vorwurf zielt demzufolge darauf ab, dass die "Abbildung" aller relevanten Aspekte in einer Entscheidungssituation auf eine einzige Präferenzordnung ein (zu) reduktionistisches Vorgehen darstellt. Denn wenn sich die unterschiedlichen Aspekte und Kriterien, welche die Bewertung einer rationalen Person P prägen, im Rahmen des Bayesschen Modells in einer einzigen Präferenzordnung niederschlagen müssen, verlangt dies im Vorhinein oftmals eine Abwägung schwer vergleichbarer Größen - wie z.B. die zwischen eigenem Wohlergehen und dem Maß der Gerechtigkeit einer Verteilung -, was darum eher Anlass zur Kritik als zu einer Befürwortung der Reduzierung gibt. An diesem Punkt seien in einem kleinen Exkurs auch auf die kläglichen Explikationsversuche des Präferenzbegriffs hingedeutet, weil vor diesem Hintergrund Sens Einwand eine noch durchschlagendere Kraft entwickelt.

Exkurs: Die beiden klassischen Möglichkeiten, den Begriff der individuellen Präferenz einzuführen, finden sich schon bei Savage (1954, S. 27f.): Einerseits kann man versuchen, die Wahrheitsbedingungen einer Aussage der Form: "P findet f1 mindestens so gut wie f2" unter Rekurs auf Äußerungen von P anzugeben. Dies erfordert jedoch, recht starke theoretische Annahmen zu treffen, die nicht alle Theoretiker teilen. Denn (i) muss man eine gemeinsame Sprache oder doch wenigstens eine weitgehende Übersetzbarkeit voraussetzen und (ii) muss der Person ein großes Wissen, das sprachlich transportierbar ist, über ihre eigenen Präferenzen zugestanden werden.

Den zweiten Explikationsversuch kann man schlagwortartig als revealed preference-Modell umschreiben. Die gesuchten Wahrheitsbedingungen für Aussagen über individuelle Präferenzen werden danach unter Rückgriff auf prinzipiell beobachtbares Entscheidungsverhalten angegeben. Dieser Ansatz ist theoretisch sehr gut ausgearbeitet (vgl. z.B. Kern/Nida-Rümelin 1994, Kap. 1), er scheint aber zu Problemen zu führen, die nahelegen, dass der erste an Äußerungen anknüpfende Explikationsversuch zu früh aufgegeben wurde, ohne auf Fruchtbarkeit hin untersucht zu werden.

Eines der Probleme, die zu diesem Urteil berechtigen, manifestiert sich eben darin, dass der oben geschilderte Reduktionismus-Vorwurf an Brisanz gewinnt, wenn man als Explikation des Präferenzbegriffs die revealed preference-Konzeption zugrundelegt. Denn es scheint sehr viel mehr Informationen über eine Person zu geben, die man innerhalb einer an Präferenzen orientierten Konzeption miteinbeziehen möchte, als diejenigen Daten, die sich über Wahlverhalten ermitteln lassen (etwa wann eine Person Kompromisse eingeht oder wie sie in kontrafaktischen Situationen entscheiden würde). Darüber hinaus ist man an eine gewisse Konstanz individueller Präferenzen gebunden, wenn man eine individuelle Präferenzrelation gleichsam durch eine Datenerhebung (die Zeit benötigt) eruieren und sie auf Kohärenz mit den Axiomen prüfen möchte. Sofern man nicht von einer Konstanz ausgeht, ist eine Nutzenfunktion jeweils bloß für einen Moment gültig. Aber insbesondere dann, wenn die Alternativen derart fein individuiert sind, dass etwa die Wohlfahrt (vieler) anderer als Individuations- respektive Unterscheidungskriterium fungiert, mutet die Konstanz von Präferenzen an, eine allzu stark idealisierende und darum illusorische Annahme zu sein. So mag - in Anspielung auf den Schwesterartikel von mir im Forum Wirtschaftsethik - zwar der Unternehmer P zum Zeitpunkt t1 f1': "Installiere Filter und Katalysatoren in einer schadstoffemittierenden Fabrik, wodurch eine bessere Luftqualität einen Rückgang der Asthmafälle und damit von Leid im nahegelegenen Dorf bewirkt" gegenüber f2' (= ¬f1') vorziehen, weil die Regierung zu t1 und in

absehbarer Zeit hohe Subventionen für den Umstieg auf umweltfreundliche Anlagen und Technologien zahlt; doch verkehrt sich diese Präferenz wohl schnell ins Gegenteil, wenn P plötzlich zu t2 keinerlei staatliche Förderung mehr erfährt, um sich teure Filter und Katalysatoren anzuschaffen. Sollten dennoch korrekte Argumente für die Konstanz-Annahme angeführt werden können, so bleibt mutmaßlich ein letzter Punkt weiterhin unberührt:

Es mag zwar durchaus überzeugen, dass gewisse konstante Elemente existieren, die in den praktischen Überlegungen von Personen zu verschiedenen Zeiten - auch im Kontext der rationalen Entscheidungsfindung - eine Rolle spielen, weil sich etwa der Charakter einer Person (womöglich) entscheidungstheoretisch durch das Befolgen gewisser konstanter Regeln ausdrücken lässt (in diesem Sinne wählen bspw. aufrichtige Menschen solche Handlungsalternativen, die sich nicht durch einen Akt des Lügens auszeichnen). Nichtsdestotrotz bleibt dieser Aspekt der Beständigkeit der Person außen vor, wenn es um die Beständigkeit von Präferenzen über äußerst fein individuierten und sehr speziellen Alternativen oder Konsequenzen geht (d.h. etwa wenn ein aufrichtiger Entscheider zwischen Alternativen zu wählen hat, die sich allesamt nicht durch einen Akt des Lügens auszeichnen). Zugespitzt formuliert, findet also der Aspekt der Konstanz der Person im klassischen Modell rationaler Entscheidungstheorie keinen Platz.[9

Mit Sen ist nun aus dem skizzierten "Reduktionismus-Vorwurf" zu folgern, dass die Begrifflichkeiten, der sich die klassischen Konzeptionen in der formalen Theorie rationaler Entscheidungen bedienen, in mindestens einer Hinsicht (im Fokus stand der Begriff der individuellen Präferenz) nicht unseren Anforderungen an eine angemessene Beschreibung von rationalem Entscheidungsverhalten gerecht werden. Es bleibt also eine Erweiterung des Begriffsapparats gefordert; oder um es mit Amartya Sen zu sagen: "To make room for the different concepts related to [rational, C.H.] behavior we need a more elaborate structure" (Sen 1977, S. 336).

4.2. Das Interpretationsproblem und abschließende Bemerkungen

Abschließend beschäftigen wir uns in diesem Abschnitt mit einem Problem, welches weitaus fundamentaler und letztlich gravierender zu sein scheint als die Frage nach der empirischen Adäquatheit der (bspw. im Anhang dargestellten) Rationalitätsprinzipien, weshalb wir ihm hier auch gegenüber den empirischen Einwänden den Vorzug einräumen. Auf den ersten Blick mag es grundsätzlich zwei mögliche Reaktionen auf eine beobachtete Verletzung eines als normativ aufgefassten Rationalitätsaxioms geben, nämlich (i) das Axiom beizubehalten und demgemäß alle Verletzungen respektive Zuwiderhandlungen als Instanzen von Irrationalität zu deklarieren oder (ii) das Axiom aufzugeben und damit von der Rationalität der (Test-)Personen auszugehen, an der sich eine normative Theorie zu messen hat.

Darüber hinaus gibt es aber noch eine dritte Alternative, die man zwar schon sehr früh erkannt, aber die erst in der jüngeren Debatte an Brisanz gewonnen hat. Sie besteht in ihrer allgemeinsten Form darin, schlichtweg zu leugnen, dass Rationalitätsprinzipien empirisch testbar sind:

If today [and probably today too, C.H.] you were to poll economists of different schools, you would almost certainly find the coexistence of beliefs (i) that the rational behavior theory is unfalsifiable, (ii) that it is falsifiable and so far unfalsified, and (iii) that it is falsifiable and indeed patently false.
(Sen 1977, S. 325)

Genauer: Es ist nicht feststellbar, ob in einer bestimmten Entscheidungssituation gegen ein Rationalitätsprinzip verstoßen worden ist oder nicht; oder anders formuliert: Jede Entscheidungssituation, in welcher ein Rationalitätsaxiom verletzt worden ist, kann man so uminterpretieren, dass keine Verletzung vorliegt. Insofern gibt es also verschiedene mögliche Interpretationen einer Entscheidungssituation, von denen immer eine das Axiom verletzt und eine andere dies nicht tut. Die eigentliche Krux besteht nun darin, dass man keine Kriterien zur Verfügung hat, anhand derer sich entscheiden ließe, welches "die richtige" und welches "die falsche" Interpretation ist.

Die Uminterpretation einer bestimmten Entscheidungssituation läuft im Grunde auf die Einführung neuer Individuierungskriterien hinaus, sodass Handlungsalternativen feiner individuiert werden als man es tun müsste, damit ein bestimmtes Rationalitätsprinzip empirisch "greift". Um den Gedankengang besser nachvollziehen zu können, illustrieren wir im Folgenden das

Interpretationsproblem am Beispiel eines simplen Postulats (aus der Klasse der im Anhang gelisteten Postulate), welches besagt, dass die auf bestimmten Objekten definierte individuelle zweistellige schwache Präferenzrelation transitiv ist (siehe Postulat 1).

Unter der Annahme, dass die Präferenzrelation über der Menge der Handlungsalternativen F erklärt ist, dürfen wir ihr - der Relation ? - die Eigenschaft der Transitivität zusprechen, gdw. gilt:
f1 ? f2 ? f2 ? f3 ? f1 ? f3 für alle f1, f2, f3 ? F

In einer informellen Schreibweise stellt also das Postulat 1 (genauer: 1c) an eine (bzgl. des Bayesschen Modells) rationale Person P die Mindestbedingung, dass wenn sie die Handlung f1 gegenüber der Handlung f2 schwach vorzieht und gleichzeitig f2 der Handlung f3 schwach vorzieht, dann sollte P auch f1 Vorrang gegenüber f3 einräumen oder zwischen f1 und f3 indifferent sein.

Im Falle der Transitivitätsforderung sieht nun das entsprechende Interpretationsproblem im Einzelnen wie folgt aus: In seinem Aufsatz "Rationality and the Sure Thing-Principle" konstruiert John Broome ein Beispiel und ein Argument, die beide die Möglichkeit praktischer Signifikanz des Transitivitätsprinzips bedrohen. Hier ist sein Beispiel:
George prefers visiting Rome to mountaineering in the Alps, and he prefers staying at home to visiting Rome. However, if he were to have a choice between staying at home and mountaineering in the Alps, he would choose to go mountaineering.
(Broome 1991a, S. 76)

Die Präferenzen von George scheinen die Anforderung der Transitivität nicht zu erfüllen. Darauf hingewiesen, behauptet George jedoch, dass dies nicht der Fall sei. Er argumentiert, dass er sich feige fühlen würde, wenn er statt bergzusteigen daheim bliebe und deshalb bei der Entscheidung zwischen Bergsteigen und Daheimbleiben ersteres wählen würde. Wenn er aber gefragt würde, ob er lieber Bergsteigen oder nach Rom ginge, wäre Feigheit nicht im Spiel. Folglich individuiert George die Handlungsalternativen feiner als der Entscheidungstheoretiker, welcher ihm einen Verstoß gegen das Transitivitätsprinzip und damit Irrationalität unterstellen möchte.

In allgemeinerer Form kann die Idee, der dieses Beispiel entspringt, folgendermaßen formuliert werden: Angenommen wir beabsichtigen, jemanden der Intransitivität zu bezichtigen. Wir müssten dann etwas sagen der Art: "Du hast f1 gegenüber f2 (schwach) vorgezogen und f2 gegenüber f3. Außerdem hast Du aber auch f3 gegenüber f1 (schwach) vorgezogen." Mit Broome kann aber der Beschuldigte immer erwidern: "Ich habe f1 der Alternative f2, wenn die andere Alternative f1 war, (schwach) vorgezogen und f2, wenn die andere Alternative f3 war, habe ich gegenüber f3 (schwach) vorgezogen. Demnach ist die Struktur meiner Präferenzen: f1 ? f2' , f2" ? f3 sowie f3 ? f1, wobei f2' ? f2". Ergo sind meine Präferenzen nicht intransitiv."

Halten wir fest: Wir können einerseits dem Entscheider prinzipiell Intransitivität zum Vorwurf machen, gdw. seine (schwachen) Präferenzen dem Antezedenz der Transitivitätsbedingung (f1 ? f2 ? f2 ? f3) gerecht werden, aber nicht dem Sukzedens (f1 ? f3). Andererseits können wir eine entsprechende Situation stets so (um-)interpretieren, dass das Antezedenz falsch und somit das Postulat - aufgrund des (in der klassischen Logik gültigen) Prinzips: Ex falso sequitur quodlibet - trivial erfüllt ist: "[?] among practical preferences the requirement of transitivity is vacuously satisfied" (Broome 1991a, S. 77). Die Frage, die offen bleibt, ist nun, welche der gegensätzlichen Interpretationen zutreffen - die Gruppe derjenigen (oder eine daraus), welche das Axiom verletzen oder die Gruppe derjenigen (oder eine daraus), welche dem Axiom genügen. Doch wie wir eingangs bereits erwähnt haben, gibt es keine passende Antwort auf diese Frage, weil wir keine Kriterien zur Hand haben, die eine der Interpretationen (oder gar mehrere) als "die richtige(n)" und die andere(n) als "die falsche(n)" ausweisen würden. Nachträglich scheint ferner aus dem genannten Interpretationsargument noch ein weiterer Kritikpunkt am klassischen Präferenzbegriff - dem revealed preference-Modell - hervorzugehen, da wir am Exempel der Transitivitätsbedingung gesehen haben, dass man möglicherweise jedes Entscheidungsverhalten als rational interpretieren kann. Obendrein ist es leicht zu erkennen, dass sich das gleiche Argument auch auf andere auf Präferenzen bezogene Rationalitätsprinzipien übertragen lässt.

Im Lichte dieser Beobachtung ist das Prinzip der Nutzenerwartungswertmaximierung nicht falsifizierbar, falls man es als empirisches Prinzip versteht respektive nicht verletzbar, falls man es normativ versteht. Dieser letztlich auf Wittgensteins

Regel-folgen-Argument zurückgehende Punkt wurde im entscheidungstheoretischen Kontext vermutlich erstmals von Samuelson (1952, S. 677) formuliert:
In what dimensional space are we "really" operating? If every time you find my axiom falsified, I tell you to go to a space of still higher dimensions, you can legitimately regard my theories as irrefutable and meaningless.

Es ist darum zu bezweifeln, ob zwischen einem Rationalitäts- und einem Interpretationsprinzip überhaupt ein Unterschied besteht. Im Besonderen bleibt daher unklar, inwiefern die auf Rationalitätsanforderungen an Präferenzen basierenden, d.h. axiomatischen Theorien angesichts dessen ihren normativen Status aufrechterhalten können. Offensichtlich werden bei der Interpretation von Entscheidungssituationen substantielle Interpretationsconstraints benötigt, die entscheiden helfen, was als gute und was als schlechte Interpretation gelten kann. Eine solche Einschränkung ließe sich prinzipiell auf zwei Weisen vornehmen: (i) Man könnte herauszufinden versuchen, welche constraints dem Entscheider wichtig sind, um so Interpretationskriterien zu gewinnen (subjektivistischer Ausweg). (ii) Man könnte auch geneigt sein, intersubjektiv gültige Standards dafür festzulegen, was als gute Interpretation einer Entscheidungssituation anzusehen ist (objektivistischer Ausweg, z.B. von Broome eingeschlagen).

In beiden Fällen gilt jedoch, dass die Frage, was für Rationalitätsbetrachtungen relevante Beschreibungsaspekte sind, wichtiger und problematischer ist als die Frage nach der Adäquatheit von Axiomen oder der Nutzenerwartungswerttheorie. Wenn sie nicht gar verkehrt ist, dann ist diese Gewichtung oder Priorisierung zumindest einmal zu hinterfragen; denn immerhin wollte die vorliegende Arbeit zu einer kritischen Beleuchtung des Bayesschen Konsequentialismus beitragen und anregen. Sollte man aber aus den prinzipiellen Bedenken, denen die Theorien vom Typus der Nutzenerwartungswert-Konzeptionen ausgesetzt sind, folgern, dass der Grad der Kohärenz unserer Intuitionen mit Bezug auf praktische Rationalität so gering ist, dass das Projekt einer klaren und insbesondere formalen Theorie praktischer Rationalität von vornherein zum Scheitern verurteilt ist? Bernard Williams bekundet seine Zustimmung, wenn er schreibt:
[?] it is unclear what the limits are to what an agent might arrive at by rational deliberation [?]. I regard it as a basically desirable feature of a theory of practical reasoning that it should preserve and account for that unclarity.
(Williams 1980, S. 110)

Die Frage, inwieweit sich eine angemessene formale Theorie entwickeln lässt, kann man nicht einfachhin in dieser Weise a priori entscheiden. Trotz Schwierigkeiten verschiedener Art, welche auszugsweise in dieser Arbeit berücksichtigt wurden, haben formale Theorien praktischer Rationalität durchaus viele Vorzüge, die zu anziehend wirken, als dass es unversucht bleiben sollte, etwaige Defizite durch Veränderung und Erweiterung - ganz im Sinne von Sens (1977) Plädoyer - zu bereinigen. Obgleich der Tenor dieser Arbeit vielleicht einen anderen Abschluss hätte vermuten lassen, so sollte doch in Anbetracht von vielversprechenden Lösungsansätzen zu zeigen versucht werden, dass das Forschungsprogramm "Formale Theorie praktischer Rationalität" nach einem Sturm von Kontroversen und einer Welle von Einwänden nicht als überholt und gescheitert anzusehen ist.

5. Literatur

Allais, Maurice: Fondements d' une Théorie Positive des Choix comportant un Risque et Critique des Postulats et Axiomes de l' École Américaine, in: Économetrie 40 (1953), S. 257-332.

Broome, John: Rationality and the Sure Thing-Principle, in: Thoughtful Economic Man: essays on rationality, moral rules and benevolence, hg. v. J. G. Meeks, Cambridge 1991, S. 74-102 (= Broome 1991a).

Broome, John: The Structure of Good: Decision Theory and Ethics, in: Foundations of Decision Theory: Issues and Advances, hg. v. M. Bacharach u. S. Hurley, Oxford 1991, S. 123-46 (= Broome 1991b).

Davidson, Donald: Psychology as Philosophy, erstm. veröff. 1974, wiederabg. in u. zit. nach Davidson: Essays on Actions and Events, Oxford 1980, S. 229-39.

Ellsberg, Daniel: Risk, Ambiguity, and the Savage Axioms, in: Quarterly Journal of Economics, Vol. 75, No. 4 (1961), S. 643-69.

Fehige, Christoph: Instrumentalism, in: Varieties of Practical Reasoning, hg. v. E. Millgram, Cambridge/Mass. 2001, S. 49-76.

Gesang, Bernward: Eine Verteidigung des Utilitarismus, Stuttgart 2003.

Güth, Werner: Spieltheorie und ökonomische (Bei)Spiele, Berlin 1992.

Hahn, Frank: Beneconfusion, in: Thoughtful Economic Man, hg. v. J. G. Meeks, Cambridge 1991, S. 7-11.

Hausner, Melvin: Multidimensional Utilities, in: Decision Processes, hg. v. R. M. Thrall u. C. H. Coombs u. R. L. Davis, New York 1954, S. 167-180.

Jeffrey, Richard C.: The Logic of Decision, Chicago 19832.

Kahnemann, Daniel/Tversky, Amos: Prospect Theory: an Analysis of Decision Under Risk, in: Econometrica, Vol. 47, No. 2 (1979), S. 263-91.

Kern, Lucian/Nida-Rümelin, Julian: Logik kollektiver Entscheidungen, München 1994.

Kripke, Saul A.: Wittgenstein on Rules and Private Language. An Elementary Exposition, Oxford 1982.

Luce, R. Duncan/Raiffa, Howard: Games and Decisions, New York 1957.

Marschak, Jacob: Rational Behavior, Uncertain Prospects, and Measurable Utility, in: Econometrica, Vol. 18, No. 3 (1950), S. 111-41.

Neumann, John von/Morgenstern, Oskar: The Theory of Games and Economic Behavior, Princeton 1944.

Neumann, John von/Morgenstern, Oskar: The Theory of Games and Economic Behavior, dt. Übers. v. M. Leppig: Spieltheorie und wirtschaftliches Verhalten, hg. v. F. Sommer, Würzburg 1961.

Nida-Rümelin, Julian: Kritik des Konsequentialismus, München 1995.

Samuelson, Paul A.: Probability, Utility, and the Independence Axiom, in: Econometrica, Vol. 20, No. 4 (1952), S. 670-78.

Savage, Leonard J.: The Foundations of Statistics, New York 1954.

Sen, Amartya K.: Choice, Orderings and Morality, in: Practical Reason, hg. v. S. Körner, Oxford 1974, S. 54-67.

Sen, Amartya K.: Rational Fools: A Critique of the Behavioral Foundations of Economic Theory, in: Philosophy and Public Affairs, Vol. 6, No. 4 (1977), S. 317-44.

Sen, Amartya K.: Beneconfusion, in: Thoughtful Economic Man, hg. v. J. G. Meeks, Cambridge 1991, S. 12-16.

Simon, Herbert: Reason in Human Affairs, Stanford 1983.

Spohn, Wolfgang: Grundlagen der Entscheidungstheorie, Kronberg/Ts. 1978.

Williams, Bernard: Internal and External Reasons, erstm. veröff. 1980, wiederabg. in u. zit. nach Williams: Moral Luck,

Cambridge 1981, S. 101-13.

6. Anhang: Eine Axiomatik mit fünf Postulaten für das Grundmodell
Nehmen wir an, die dem Entscheider vorliegenden Alternativen seien Lotterien, die man wie folgt definieren kann:
Eine einfache Lotterie l ist ein n-Tupel, das Paare von je einer Konsequenz aus der
(endlichen) Konsequenzenmenge C und einer reellen Zahl enthält, also:

l = {(p1, c1), ?, (pn, cn)}

wobei gilt, dass pi ? 0 (für i = 0, ?, n) und ?ni=1 pi = 1.

Die Menge aller einfachen Lotterien werde mit L* bezeichnet.

Der Ausdruck l = {(p1, c1), ?, (pn, cn)} sei wie folgt interpretiert:
"one and only one prize will be won and the probability that it will be [ci] is pi "
(Luce/Raiffa 1957, S. 24).

Demgegenüber sind wir noch auf den Begriff der zusammengesetzten Lotterie angewiesen, der den der einfachen Lotterie insoweit verallgemeinert, als in zusammengesetzten Lotterien auch Lotterien (zusammengesetzte und einfache) als Ausgänge (Preise) auftreten können.
Die Menge der Lotterien L ist die kleinste Menge mit:

- L* ? L und

Für alle n-Tupel positiver reeller Zahlen (p1, ?, pn) mit ?ni=1 pi = 1 und für alle n-Tupel von Lotterien (l1, ?, ln) ? Ln gilt: {(p1, l1), ?, (pn, ln)} ? L.

Nun sei die Präferenzrelation auf der Menge der Lotterien erklärt, was keine größeren Abweichungen bedeutet (Luce/Raiffa 1957, S. 25). Unter der Voraussetzung des Wahrscheinlichkeitsbegriffs genügen dann nachstehende Postulate, um das nutzenerwartungswerttheoretische Theorem, das sogenannte Repräsentationstheorem, herzuleiten.

Postulat 1: Die schwache Präferenzrelation über Lotterien ist eine Ordnung.
Wir fassen damit die ersten drei Bedingungen bei Nida-Rümelin (1995, S. 37f.) zusammen, die er an die Präferenzen einer rationalen Person P stellt. Im Einzelnen handelt es sich dabei um:

Reflexivität (ist trivial erfüllt): ? l1 ? L: l1 ? l1.
Konnexität: Es wird angenommen, dass P für beliebige Lotterien l1, l2 aus L eine Präferenz hat: Entweder zieht P l1 gegenüber l2 vor oder sie zieht l2 gegenüber l1 vor oder sie ist zwischen beiden indifferent:? l1, l2 ? L: l1 ? l2 ? l2 ? l1.
Transitivität: Die dritte Bedingung verlangt von P, dass sie eine schwache Präferenz für l1 gegenüber l3 hat, wenn sie l1 gegenüber l2 schwach vorzieht und l2 gegenüber l3 schwach vor-zieht:? l1, l2, l3 ? L: l1 ? l2 ? l2 ? l3 ? l1 ? l3.

Postulat 2: Eine zusammengesetzte Lotterie ist gleich zu bewerten wie eine einfache Lotterie über Konsequenzen, wenn die Konsequenzen in beiden Fällen die gleichen sind und mit den gleichen Wahrscheinlichkeiten auftreten.
Dies formal wiederzugeben, ist etwas aufwendiger (da wir als Zwischenschritt eine Funktion einzuführen hätten, die zusammengesetzte Lotterien in einfache abbildet). Deshalb begnügen wir uns an dieser Stelle mit einer mathematisch etwas unsauberen Schreibweise,[10 was hoffentlich dem Verständnis zugute kommt:
? c1, c2 ? C ? p, p' ? [0, 1] : [p, (p', c1; (1 ? p'), c2); (1 ? p), c2] ~ [p' × p, c1; (1 ? p' × p), c2].[11

Mit anderen Worten: Wir akzeptieren, dass Umformungen von Lotterien nach den Regeln der Wahrscheinlichkeitsrechnung die Präferenz nicht verändern, was natürlich stochastische Unabhängigkeit der Ereignisse voraussetzt. Dieses Postulat wird häufig auch als Reduktionsaxiom bezeichnet.

Postulat 3: Sei ci ? C. Ferner gelte: (1, c1) ? (1, ci) ? (1, cn). Dann existiert ein p ? [0, 1] mit
(1, ci) ~ (p, c1; 0, c2; ?; 0, cn ? 1; (1 ? p), cn);
wobei ci Sicherheitsäquivalent zu der Lotterie (mit c1 und cn) heißt.

Anders ausgedrückt: Bei oben gegebener Präferenzordnung gibt es immer eine Wahrscheinlichkeit p, für welche die sichere Konsequenz ci als indifferent zu der Lotterie l': (p, c1; (1 ? p), cn) beurteilt wird. Dieses Axiom ist eine Stetigkeitsannahme, deren Berechtigung sich leicht einsehen lässt. Denn wenn die Person P cn gegenüber ci vorzieht, dann muss es eine, womöglich sehr hohe Wahrscheinlichkeit für cn in l' geben, bei der P l' gegenüber ci vorzieht. Ebenso müsste es umgekehrt eine eventuell sehr hohe Wahrscheinlichkeit für c1 geben, bei der P ci gegenüber l' vorzieht. In diesem Kontinuum sich ändernder Wahrscheinlichkeiten für c1 respektive cn muss es eine Verteilung geben, bei der P zwischen l' und ci indifferent ist.

Postulat 4: Seien l1, l2 ? L zwei Lotterien, zwischen denen Indifferenz besteht. Dann sind l1 und l2 in beliebigen Kontexten austauschbar, ohne dass sich die Präferenz ändert.
Diese Forderung bezeichnet man oftmals als Unabhängigkeit von irrelevanten Alternativen: Wenn P also l1 genauso gut findet wie l2, dann kann P l1 überall, d.h. bspw. wenn l1 Bestandteil einer Lotterie ist, durch l2 substituieren, ohne dass sich an ihren jeweiligen Präferenzen etwas ändert.

Postulat 5: Es gelte (1, c1) ? (1, c2) für alle c1, c2 ? C. Dann gilt für alle p, p' ? [0, 1]:
(p, c1; (1 ? p), c2) ? (p', c1; (1 ? p'), c2) ? p ? p'

Die hinter dieser Monotonieforderung stehende Intuition ist erneut äußerst plausibel: Wenn eine Lotterie die bevorzugte Konsequenz c1 mit größerer Wahrscheinlichkeit hervorbringt als eine andere Lotterie (und ansonsten alle Komponenten der beiden Lotterien gleich sind), dann ist sie zu bevorzugen.

[1 Im Besonderen ist an dieser Stelle hervorzuheben, dass von Neumann / Morgenstern erstmalig ein Metrisierungstheorem für einen kardinalen Begriff des Nutzens bewiesen haben. Vgl. von Neumann / Morgenstern 1961, Kap. I.3 (dt. Übers.).

[2 Streng genommen gibt es nicht eine einzige Nutzenerwartungswerttheorie, sondern einen ganzen Theorientypus, der viele voneinander abweichende Nutzenerwartungswert-Konzeptionen subsumiert. Diese Differenzierung spielt für die folgende

Untersuchung jedoch keine bedeutende Rolle.

[3 Eine Partition der Menge M ist eine Familie von paarweise disjunkten Teilmengen von M, deren Vereinigung ganz M ist.

[4 Ein ausführlicher Beweis findet sich bei Güth 1992 für ein geringfügig anderes Axiomensystem.

[5 Wie sich bei einem Vergleich der Konzeptionen zeigen würde, geht eine Reihe von Modellen von der Existenz subjektiver Wahrscheinlichkeiten aus und versucht (lediglich), eine Explikation des Nutzenbegriffs zu geben. Diesen Ansatz verfolgen bspw. die Modelle von Marschak (1950) und Luce/Raiffa (1957). Eine andere Art mit dem Problem der individuellen Wahrscheinlichkeit umzugehen, schlagen Savage (1954) und Jeffrey (1983) vor, die einen solchen Wahrscheinlichkeitsbegriff aus der Beschränkung der individuellen Präferenzen durch gewisse Axiome gewinnen.

[6 Vgl. dazu insbesondere: Allais (1953), Ellsberg (1961) und Kahnemann/Tversky (1979).

[7 Statt also bspw. P zwischen f1: "Installiere Filter und Katalysatoren in einer schadstoffemittierenden Fabrik" und f2 (= ¬f1) wählen zu lassen, kann die Entscheidungssituation auch so beschrieben werden, dass P's Wahl zwischen f1': "Installiere Filter und Katalysatoren in einer schadstoffemittierenden Fabrik, wodurch eine bessere Luftqualität einen Rückgang der Asthmafälle und damit von Leid im nahegelegenen Dorf bewirkt" und f2' (= ¬f1') besteht. Usw.

[8 Es handelt sich hierbei aber nicht um eine Reduktion von Altruismus auf Egoismus (im wertenden Sinne der Umgangssprache). Die abwertende Verwendung des Begriffs "Egoismus" bezieht sich auf direkte egoistische Motive, nicht auf indirekte: "Egoisten sind nicht Leute, die ihrem Herzen folgen, sondern Leute, deren Herzen kalt sind. Sie können nicht als Menschen definiert werden, die ihren Wünschen folgen, sondern als Menschen, denen Wünsche einer bestimmten Art fehlen, nämlich die Wünsche, dass es anderen gut gehen soll" (Fehige 2001, S. 61; übers. nach Gesang 2003, S. 46f.).

[9 Vgl. hierzu auch Davidson (1974), S. 235: "[?] a theory like Ramsey's has no predictive power at all unless it is assumed that beliefs and values do not change over time. The theory merely puts restrictions on a temporal cross-section of an agent´s dispositions to choose".

[10 Es wird etwa einfachhin eine Schachtelungstiefe von 2 unterstellt, d.h. die zusammengesetzte Lotterie enthält eine einfache und nicht eine weitere zusammengesetzte Lotterie (beliebiger Schachtelungstiefe).

[11 Um den Einsatz von Klammern zu reduzieren, wird eine vereinfachte Schreibweise gewählt: "pi, ci" stehe kurz für "(pi, ci)". Darüber hinaus symbolisiert die Tilde Indifferenz (welche wie folgt definiert ist: l1 ? l2 ? l2 ? l1 ? l1 ~ l2).

Der Autor

Dr. Christian Hugo Hoffmann

Dr. Christian Hugo Hoffmann ist Assistenzprofessor am Lehrstuhl für Finance der Universität Liechtenstein, wo er sich vor allem den Themen Fintech (API Economy, Future of Banking, Insurtech sowie Cryptocurrencies), Entrepreneurial Risks und Austrian

Economics verschreibt. Seinen Postdoc hat er an der ETH Zürich, speziell am ETH Risk Center absolviert. Promotion erfolgte an der Universität St. Gallen (HSG) mit längerem Aufenthalt an der Yale University im Programm Financial Stability and Systemic Risks. Daneben ist Christian leidenschaftlicher Unternehmer mit zwei Fintech-Gründungen in Deutschland und der Schweiz, als Vizedirektor des Swiss Fintech Innovation Lab, als Direktor von Startup Grind Genf, als Startup-Coach/-Judge und -Mentor in verschiedenen Programmen (Mass Challenge, Vroom, Kickstart Accelerator), als früherer Head of Finance der erfolgreichen Robotikfirma Verity Studios sowie mit Beteiligungen in verschiedenen Startups.