

Die Spieltheorie als Hilfswissenschaft einer modernen Ethik

Magisterarbeit
zur Erlangung eines Grades eines
Magister Artium M. A.

vorgelegt
der
Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

von
Christian Feige
aus
Bonn

Eidesstattliche Erklärung

An Eides statt versichere ich, dass die Arbeit

Die Spieltheorie als Hilfswissenschaft einer modernen Ethik

von mir selbst und ohne jede unerlaubte Hilfe angefertigt wurde, dass sie noch keiner anderen Stelle zur Prüfung vorgelegen hat und dass sie weder ganz noch im Auszug veröffentlicht worden ist. Die Stellen der Arbeit – einschließlich Tabellen, Karten, Abbildungen, usw. –, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall als Entlehnung kenntlich gemacht.

Bonn, 14. August 2006

Inhaltsverzeichnis

1. Einleitung	S. 5
1.1 Motivation des Themas	S. 5
1.2 Spieltheorie und Ethik	S. 6
1.3 Spieltheorie als Hilfswissenschaft	S. 8
1.4 Überblick der Arbeit	S. 9
2. Die utilitaristischen Wurzeln der Spieltheorie	S. 12
2.1 Der Nutzen	S. 13
2.1.1 <i>Der Hedonismus</i>	S. 14
2.1.2 <i>Der Präferenzutilitarismus</i>	S. 16
2.1.3 <i>Der Nutzen in der Spieltheorie</i>	S. 18
2.1.4 <i>Interpersoneller Nutzenvergleich</i>	S. 20
2.2 Akt- vs. Regel-Utilitarismus	S. 22
2.3 Deontologismus vs. Teleologismus	S. 24
3. Die Spieltheorie als strategische Entscheidungstheorie	S. 27
3.1 Die Bayesianische Entscheidungstheorie	S. 28
3.2 Spieltheoretische Gleichgewichtskonzepte	S. 30
3.2.1 <i>Das Gefangenen-Dilemma</i>	S. 30
3.2.2 <i>Dominante Strategien</i>	S. 32
3.2.3 <i>Common Knowledge</i>	S. 32
3.2.4 <i>Das Nash-Gleichgewicht</i>	S. 34
3.2.5 <i>Dynamische Spiele</i>	S. 35
3.2.6 <i>Spiele mit unvollständigen Informationen</i>	S. 37
3.2.7 <i>Probleme der Spieltheorie im Alltag</i>	S. 37
3.3 Die Evolutionäre Spieltheorie	S. 39
3.3.1 <i>Das Taube-Falke-Spiel</i>	S. 39
3.3.2 <i>Nutzen und fitness</i>	S. 40
3.3.3 <i>ESS und Replikatordynamik</i>	S. 41
3.4 Die Bedeutung von strategischem Denken für die Entscheidungsfindung	S. 43

4. Die Vollständigkeit der spieltheoretischen Darstellung	S. 47
4.1 Vermeidung eines moralischen Dilemmas	S. 49
4.2 Erweiterungsmöglichkeiten der Spieltheorie	S. 52
5. Die Unparteilichkeit der Spieltheorie bei der Beschreibung ethischer Ansätze	S. 58
5.1 Akt- vs. Regel-Utilitarismus aus Sicht der Spieltheorie	S. 60
5.1.1 Die Positionen von Harsanyi und von Kutschera	S. 60
5.1.2 Gemischte Strategien als allgemeingültige Regel	S. 62
5.2 Evolutionäre Spieltheorie und deskriptive Ethik	S. 64
5.2.1 Evolutionäre Ethik und Spieltheorie	S. 65
5.2.2 Welches moralische Verhalten bevorzugt die Evolutionäre Spieltheorie?	S. 66
6. Abschließendes Fazit und Ausblick	S. 68
Literaturverzeichnis	S. 70
Lebenslauf	S. 72

Abbildungsverzeichnis

<i>Abb. 3.1 Das Gefangen-Dilemma</i>	S. 31
<i>Abb. 3.2 Beispiel für eine dominante Strategie</i>	S. 32
<i>Abb. 3.3 Einschätzung des Verhaltens eines Mitspielers</i>	S. 33
<i>Abb. 3.4 Spiel mit zwei Nash-Gleichgewichten in reinen Strategien</i>	S. 34
<i>Abb. 3.5 Beispiel für ein dynamisches Spiel</i>	S. 36
<i>Abb. 3.6 Das Taube-Falke-Spiel</i>	S. 40
<i>Abb. 3.7 Stein-Schere-Papier</i>	S. 43
<i>Abb. 3.8 Koordinationsproblem</i>	S. 45
<i>Abb. 4.1 Kernkraftwerk-Beispiel</i>	S. 51
<i>Abb. 4.2 a) Nutzenfunktion in der deontologischen Ethik</i>	S. 55
<i>Abb. 4.2 b) Angenäherte Nutzenfunktion</i>	S. 56
<i>Abb. 4.3 Das Ultimatum-Spiel</i>	S. 56
<i>Abb. 5.1 Gegenbeispiel zum Regel-Utilitarismus</i>	S. 61

1. Einleitung

1.1 Motivation des Themas

Moralische Entscheidungen sollten nicht strategisch motiviert sein. So sagt sinngemäß beispielsweise Kant in seiner *Grundlegung zur Metaphysik der Sitten*¹. Es bedürfe keines Grundes, um etwas zu tun, das *moralisch richtig* ist, außer der Tatsache, *dass* es moralisch richtig sei, so zu handeln. Viele Menschen haben allerdings Schwierigkeiten zu verstehen, warum eine Handlung allein aus Pflichtgefühl erfolgen soll. Selbst wenn sie sich an den Gedanken gewöhnen können, dass moralische Handlungen offenbar keinen Zweck verfolgen, der sich in greifbaren Folgen einer Tat bemerkbar macht, bleibt immer noch die Frage, ob es nicht eigentlich *doch* ein höheres Ziel gibt, für dessen Erreichen man die Ethik überhaupt erst entwickelt hatte: Man soll eine Handlung wählen, die moralisch richtig ist, aber warum ist es *wichtig*, moralisch richtig zu handeln? Wofür das Ganze?

Der Utilitarismus gibt eine Antwort auf diese Frage: moralisches Handeln geschieht zum Wohl der Gesellschaft; also besteht der Zweck der Ethik darin, ein Verhalten vorzugeben, welches das Wohlergehen bzw. den Nutzen aller Beteiligten maximiert. In der heutigen Zeit streben ohnehin viele Menschen nach einer Verbesserung ihrer Lebensumstände oder nach sinnlichen Genüssen. Dieser *Hedonismus*, also das Streben nach Glück, ist zwar nur eine Variante der utilitaristischen Ethik, die noch dazu inzwischen sogar von den meisten Utilitaristen selbst in Frage gestellt wird. Grundsätzlich ist der Schritt von der Maximierung des eigenen körperlichen Wohlbefindens zur Anwendung des Prinzips der Nutzenmaximierung aber nicht groß, so dass das Ziel einer Verbesserung des gesellschaftlichen Wohlergehens hinsichtlich eines weiter gefassten Nutzenbegriffs recht einfach vermittelt werden kann.

Daher hat eine Ethik, die zur Bewertung von Handlungen auf einen Nutzenwert zurückgreift, bessere Chancen auch von großen Teilen der Gesellschaft akzeptiert zu werden. Das ist natürlich noch kein Grund um eine derartige Ethik tatsächlich zu verfolgen, aber allein der Umstand, dass zweckorientierte Normen weit verbreitet sind, legt vielleicht schon nahe, diese Auffassung von Moral etwas genauer zu untersuchen. Wer versucht, einem Kind zu erklären, warum eine Handlung moralisch richtig ist, wird sich wahrscheinlich auch nicht gleich auf das Sittengesetz berufen, sondern dem Kind zunächst die Folgen der zur Wahl stehenden Handlungen schildern.

Sobald aber die Folgen einer Handlung für andere Menschen abgewogen werden, bekommen die Entscheidungen zwangsläufig strategischen Charakter: nicht nur meine eigene

¹ Vgl. Immanuel Kant (1785), insbesondere Zweiter Abschnitt.

Handlung allein ist von Bedeutung, sondern die Handlungen aller an einer Situation beteiligten Personen erzeugen gemeinsam ein bestimmtes Ergebnis, das in seiner Tragweite (oder bzgl. seines Nutzens) außerdem von jeder Person unterschiedlich eingeschätzt wird. Wenn in so einer Situation aber mehrere Menschen gleichzeitig Entscheidungen zwischen möglichen Handlungen treffen, kann nicht mehr so leicht vorhergesagt werden, was am Ende dabei herauskommt. Denn mit jedem weiteren Akteur potenziert sich die Anzahl der möglichen Handlungskombination, so dass eine entsprechende Analyse schnell unübersichtlich wird. Die Spieltheorie ist jedoch auf genau dieses Problem spezialisiert und deshalb unverzichtbar als Hilfswissenschaft einer Ethik, die das moralisch Richtige mit den Folgen von Handlungen verbindet.

In komplexeren Situationen kann eine Hilfswissenschaft die Wahl der moralisch richtigen Handlung erleichtern, weil das mathematische Handwerkszeug in jedem Fall einen Lösungsmechanismus bietet. Die eigentliche Aufgabe einer Hilfswissenschaft in der Ethik besteht aber darin, die ethischen Fragestellungen sauber darzustellen, so dass mögliche Probleme leichter erkannt werden. Damit wird sichergestellt, dass Vertreter unterschiedlicher ethischer Positionen in Diskussionen dieselbe Sprache sprechen und die Argumente ihrer Kritiker nachvollziehen können.

1.2 Spieltheorie und Ethik

Die Spieltheorie wurde ursprünglich entwickelt um ökonomische Probleme zu lösen und ist deshalb vor allen Dingen als Teildisziplin der Wirtschaftswissenschaften bekannt. Das Besondere in dieser Theorie liegt in der Berücksichtigung der Handlungen anderer *Spieler*, d.h. wirtschaftlichen Konkurrenten oder auch einfach nur irgendwie beteiligten Personen, bei der Wahl der eigenen Handlungen bzw. der *Strategie*. Die Spieltheorie ist eine *Entscheidungstheorie*, da sie Entscheidungen in eine mathematisch-logische Sprache übersetzt und Strategien hinsichtlich einer Zielvorgabe bewertet. Da die Spieltheorie über die wirtschaftswissenschaftliche Anwendung auch durch den Utilitarismus geprägt wurde, verwendet sie für diese Bewertung ein Nutzenkalkül, das die persönlichen Ziele eines Entscheiders in eine Nutzenfunktion übersetzt, die maximiert wird, wenn sich der Entscheider seinen Zielen annähert. Im weiteren Verlauf der Arbeit wird erläutert, was genau unter „Nutzen“ zu verstehen ist, wo dabei möglicherweise Probleme liegen und welcher Zusammenhang mit dem Nutzenbegriff des Utilitarismus besteht. Allerdings sollte schon jetzt betont werden, dass der „Nutzen“ in der Spieltheorie nur eine mathematische Größe ist, die erst durch die Anwendung in den Wirtschaftswissenschaften mit einer Bedeutung versehen

wird. Auch eine Ethik muss deshalb zunächst diesen Nutzenbegriff mit Inhalt füllen, wenn sie spieltheoretische Konzepte verwenden will.

Bereits David Hume und Jean-Jacques Rousseau verwenden Beispiele, in denen das strategische Verhalten der beteiligten Personen von Bedeutung ist, um zu zeigen, dass gelegentlich durch Zusammenarbeit ein besseres Ergebnis erzielt werden kann.² Aber erst mit dem Buch *Theory of Games and Economic Behavior* von John von Neumann und Oskar Morgenstern³ wurden in der Mitte des 20. Jahrhunderts die mathematischen Grundlagen für die Spieltheorie geschaffen und John Nash⁴ konnte sein Gleichgewichtskonzept entwickeln. Nur wenige Jahre später bemühte sich Richard B. Braithwaite⁵ um eine Umsetzung dieser neuen Erkenntnisse für die Ethik. An dem Beispiel zweier Musiker – ein Pianist und ein Trompetenspieler, die jeweils auf ihren Instrumenten spielen wollen, ohne dabei von der Musik des anderen gestört zu werden – zeigt er, wie die Spieltheorie eine Lösung finden kann, mit der alle beteiligten Personen zufrieden gestellt werden können.

In dem Beispiel von Braithwaite ist es offensichtlich vernünftig, wenn sich beide Musiker auf eine Vereinbarung einigen, die regelt, wer wann musizieren darf. In der Spieltheorie kann so eine Regelung als *gemischte Strategie* dargestellt werden. Dabei wird eine neue Handlungsmöglichkeit oder Strategie aus einer Kombination von zwei bereits bekannten Handlungsmöglichkeiten gebildet. Jede der zwei (oder auch mehr) Handlungen wird mit einer gewissen Wahrscheinlichkeit belegt, so dass eine gemischte Strategie im Musikerbeispiel lauten könnte: 30% der Zeit wird Trompete gespielt, die restlichen 70% Klavier. Dadurch vervielfältigt sich natürlich die Zahl der Handlungsmöglichkeiten und damit auch die Anzahl der Kandidaten für eine moralisch richtige Handlung in dieser Situation. Auch wenn das Konzept der gemischten Strategie nicht unproblematisch ist (was im Verlauf der Arbeit noch gezeigt werden soll), wird dadurch vielleicht schon deutlich, dass die Spieltheorie die Mittel einer Ethik, insbesondere des klassischen Utilitarismus, erweitern könnte.

Ein halbes Jahrhundert nach Braithwaite hat die Spieltheorie in verschiedenen Bereichen der Ethik Fuß gefasst. Bruno Verbeek und Christopher Morris⁶ nennen drei unterschiedliche Anwendungsgebiete für spieltheoretische Konzepte in der Ethik: den Funktionalismus, die Theorie der Gesellschaftsverträge und die Evolutionäre Ethik.

² Vgl. Hume (1739), Buch 3, Teil 2, §2 bzw. Rousseau (1755), Teil 2.

³ Vgl. Von Neumann/Morgenstern (1944).

⁴ Vgl. Nash (1950).

⁵ Vgl. Braithwaite (1955).

⁶ Vgl. im Folgenden: Verbeek/Morris (2004).

Der *Funktionalismus* untersucht, zu welchem Zweck Normen entstehen und was durch sie erreicht wird. So besteht im Beispiel von Braithwaite die Aufgabe einer Norm darin zu vereinbaren, welcher Musiker wann spielen darf.

Vertragstheoretiker verwenden die Spieltheorie, um Gesellschaftsverträge zu erstellen, in denen ein sozial erwünschtes Verhalten geregelt wird. Mithilfe der Spieltheorie wird dabei zunächst eine Prognose darüber erstellt, welches Verhalten sich voraussichtlich *ohne* eine Regelung einstellen wird, um dann die nötigen Änderungen durch eine Absprache aller beteiligten Personen (dem Vertrag) festzulegen. Außerdem können spieltheoretische Konzepte überprüfen, welche Verträge anreizverträglich sind, also auch tatsächlich von einem vernünftigen Menschen befolgt werden.

Die *Evolutionäre Ethik* schließlich sieht Moral eher als ein Zufallsprodukt an, das entsteht, wenn Menschen interagieren.⁷ Angenommen eine kleine Gruppe von Personen wird mehrmals mit einem ähnlichen ethischen Problem konfrontiert. Es ist zu erwarten, dass das Problem vielleicht nicht jedes Mal auf dieselbe Weise gelöst wird. Aber eine Lösung, die häufiger zur Anwendung gekommen ist, hat sich anscheinend bewährt und ist auf eine gewisse Weise *besser* als andere Lösungen. Vor allen Dingen diejenigen Verhaltensweisen, die auch nach wiederholtem Auftreten des Problems immer noch verwendet werden, sind daher offenbar von besonderer Güte. Wenn sich also moralische Normen mit der Zeit verändern, dann könnte die „Überlebenswahrscheinlichkeit“ einer Norm ein mögliches Auswahlkriterium darstellen. Wegen seiner objektiveren Betrachtungsweise von Moral und den geringeren Anforderungen an die intellektuellen Fähigkeiten der Akteure ist der evolutionstheoretische Ansatz für diese Arbeit von besonderem Interesse und wird daher später ausführlicher dargestellt.

1.3 Spieltheorie als Hilfswissenschaft

Der Gedanke, die Spieltheorie als eine Hilfswissenschaft der Ethik einzusetzen, wird beispielsweise von Gerhard Vollmer erwähnt. Er betrachtet diese Theorie als einen wichtigen Fortschritt auf der methodologisch-instrumentellen Seite der moralphilosophischen und ethischen Diskussion.⁸ Die Grenzen und Möglichkeiten der Spieltheorie fasst er ganz gut mit folgendem Zitat zusammen:

„Die Spieltheorie ist [...] nur ein Werkzeug, eine mathematisch durchgebildete Theorie, die uns hilft, Entscheidungssituationen zu *strukturieren* und Entscheidungen als rational auszuzeichnen, *wenn* bestimmte Bedingungen erfüllt sind. Zu diesen Bedingungen gehört,

⁷ So Verbeek/Morris (2004).

⁸ Vgl. Gerhard Vollmer (1993), S. 174.

dass die Beteiligten (die „Spieler, „Strategen“, „Entscheider“) gewisse *Interessen* haben, dass diese Interessen *vergleichbar* und möglichst quantitativ *bewertbar* sind, dass die Handlungsalternativen und wenigstens einige Handlungsfolgen bekannt sind, usw. *Ob* diese Bedingungen tatsächlich vorliegen, kann der Spieltheoretiker dagegen nicht entscheiden. Die Spieltheorie beschreibt nicht die Welt; sie ist keine erfahrungs-, sondern eine strukturwissenschaftliche Disziplin. Welche Interessen Menschen und gegebenenfalls andere Lebewesen *tatsächlich* haben oder verfolgen, das kann nur empirisch ermittelt werden.“⁹

Vollmer spricht mit der Bewertung und dem Vergleich von „Interessen“, also den unterschiedlichen Zielvorgaben der beteiligten Personen, einen wichtigen Punkt an, der genauer erläutert und geklärt werden muss, damit sich die Spieltheorie als Handwerkszeug zur Modellierung ethischer Probleme durchsetzen kann. Um aber auch nur annähernd die Bedeutung der Logik als Hilfswissenschaft in der Philosophie und damit auch der Ethik zu erreichen, müssen noch weitere Fragen beantwortet werden:

1. Ist die Berücksichtigung strategischer Vorgehensweisen tatsächlich unverzichtbar, um ethische Fragestellungen zu modellieren? Oder kann vielleicht auch eine andere Entscheidungstheorie diese Aufgabe ausreichend erfüllen, ohne die Handlungen aller beteiligten Personen zu berücksichtigen?
2. Ermöglicht die Spieltheorie eine *vollständige* Darstellung aller Aspekte, die für eine moralische Entscheidung relevant sein könnten? Oder ist sie in ihrer Anwendung auf einige wenige ethische Ansätze wie den Utilitarismus beschränkt?
3. Kann die Spieltheorie überhaupt neutrale Aussagen treffen, wenn sie doch unbestreitbar utilitaristische Wurzeln besitzt? Vielleicht entscheidet man sich schon mit der Anwendung spieltheoretischer Konzepte für den Utilitarismus und gegen jede andere Auffassung von Ethik.

1.4. Überblick der Arbeit

Im weiteren Verlauf dieser Arbeit sollen die oben genannten Punkte untersucht werden, um die Grenzen der Spieltheorie als Hilfswissenschaft einer Ethik festzustellen. In Abschnitt 2 wird zunächst der utilitaristische Hintergrund erläutert, um die Spieltheorie in der Philosophie zu verorten. Dabei spielt der Nutzenbegriff eine zentrale Rolle. Es werden Probleme bei der Nutzenmessung erörtert und mit dem Präferenzen-Konzept von Richard M. Hare¹⁰ ein Lösungsvorschlag vorgestellt.

⁹ Ebd. S. 176.

¹⁰ Siehe z. B. Hare (1981).

Abschnitt 3 führt dann nach und nach spieltheoretische Gleichgewichtskonzepte ein. An dieser Stelle soll auch gezeigt werden, warum es für eine Ethik sinnvoll ist, strategische Überlegungen der beteiligten Personen zu berücksichtigen. Außerdem soll in diesem Zusammenhang der Vorzug der Spieltheorie gegenüber anderen Entscheidungstheorien herausgearbeitet werden, nämlich dass sie bei gleichen Fähigkeiten bezüglich der Darstellung ethischer Fragestellungen auch mit verwickelten Situationen umgehen kann. Denn die strategische Vorgehensweise der Entscheider macht es leichter, ihre bisherigen Handlungen nachzuvollziehen und ihre zukünftigen Handlungen vorherzusagen, so dass die Komplexität einer Entscheidungssituation verringert werden kann.

Nachdem die Grundlagen der Spieltheorie dargestellt wurden, kann in Abschnitt 4 die Reichweite ihrer Möglichkeiten in der Ethik diskutiert werden. Den Anfang macht ein Beispiel, in welchem ein Utilitarist bei seiner Wahl der moralisch richtigen Handlung die Denkweise eines Anhängers einer nicht-utilitaristischen Ethik nachvollziehen muss. Anschließend werden Argumente dafür gebracht, dass die Spieltheorie in der Lage ist, auch nicht-utilitaristische Ansichten über moralisch richtiges Verhalten vollständig zu beschreiben. Das liegt daran, dass jedes Entscheidungsverhalten eine gewisse Rangordnung von Handlungen erkennen lässt, die unter bestimmten Voraussetzungen in eine Nutzenskala übertragen werden können. Leider stellt sich heraus, dass eine Ethik, die sich auf nur zwei Wertungsgrade für eine moralische Handlung beschränkt, die Grenzen der Spieltheorie übersteigt, da sie keine Kompromisse zulässt: eine Handlung ist nach einer derartigen Ethik entweder moralisch richtig oder moralisch falsch; eine gemischte Strategie, die diese beiden Handlungen kombiniert, ist nicht besser als die falsche Handlung, sondern genauso falsch. Die Art und Weise in der Kant seine Ethik begründet, kommt dieser Moralvorstellung wohl am nächsten, so dass zu bezweifeln ist, dass ein Kantianer viel mit spieltheoretischen Konzepten anfangen können wird.

Als Ergebnis bleibt jedoch festzuhalten, dass jede Ethik, die eine messbare Rangordnung von Handlungen erstellt (und sei es nur in der Art von „schlecht, besser, gut“), ausreichend von der Spieltheorie modelliert werden kann, damit diese Ethik spieltheoretische Konzepte verwenden kann. Möglicherweise lässt sich sogar zeigen, dass jede Ethik mit einer solchen Rangordnung letzten Endes konsequentialistisch in dem Sinne ist, dass sie den moralischen Wert einer Handlung nach ihren Folgen bestimmt, was z. B. für Bernard Williams Grund genug ist, eine Ethik dieser Art abzulehnen.¹¹ Aber für die Spieltheorie ist es nicht von wesentlicher Bedeutung, ob ihr Anwender den Zweck über die Mittel stellt. Sie bemüht sich

¹¹ Vgl. Williams (1973)

vorrangig um die ausreichende Darstellung von Entscheidungssituationen, gleichgültig welche Gründe die Wahl einer Entscheidung beeinflussen könnten.

Diese Ansicht ist das Motto des fünften Abschnittes dieser Arbeit, in dem es um die Frage geht, ob die Spieltheorie nicht vielleicht parteiisch ist und Anhänger einer utilitaristischen Ethik bevorzugt. Zunächst einmal sollen an spieltheoretischen Beispielen die Positionen von John C. Harsanyi und Franz von Kutschera in der Diskussion zwischen Akt- und Regel-Utilitaristen erläutert werden,¹² um einen ersten Hinweis darauf zu geben, dass man spieltheoretische Konzepte grundsätzlich auch zur Kritik utilitaristischer Ansätze einsetzen kann, so dass diese Ethik nicht unbedingt eine Sonderstellung in der spieltheoretischen Darstellung einnimmt. Anschließend wird am Konzept der „gemischten Strategie“ erläutert, wie die Spieltheorie zwischen Akt- und Regel-Utilitarismus vermitteln kann.

Um die Frage der Ansatzneutralität der Spieltheorie in Bezug auf nicht-utilitaristische Ethiken klären zu können, muss man jedoch eine Betrachtungsebene höher klettern und die Situation nicht mehr aus den Augen eines vernunftbegabten Entscheiders, sondern eines neutralen Außenstehenden beobachten. Nun kann man verschiedene ethische Ansätze miteinander vergleichen, indem man sich ein Kriterium überlegt,¹³ das für jeden dieser Ansätze erstrebenswert sein könnte. Eine Möglichkeit besteht darin, die Überlebensfähigkeit einer Norm in folgenden Generationen einzuschätzen. Vielleicht ist der Utilitarismus ja nicht die einzige Art von Moral, die sich behaupten kann. Das passende Analysekonzept liefert die Evolutionäre Spieltheorie: die *Evolutionär Stabile Strategie*. Ursprünglich von John Maynard Smith¹⁴ für die Evolutionsbiologie entwickelt, lässt sich damit auch eine kulturelle Evolution beschreiben, bei der nicht mehr die Weitergabe von Genen, sondern von moralischen Normen im Mittelpunkt steht. In dieser Arbeit soll dafür argumentiert werden, dass dieses spieltheoretische Konzept auf der deskriptiven Ebene möglicherweise sogar eine kantianische Ethik in den Griff bekommt, wenn auch nur aus der Sichtweise eines nicht-kantianischen Beobachters. Die Evolutionäre Spieltheorie fungiert also gewissermaßen als unparteiischer Schiedsrichter beim Rennen der unterschiedlichen moralischen Ansichten.

¹² Vgl. Harsanyi (1982) und von Kutschera (1982), S. 169ff.

¹³ Gerhard Vollmer spricht in diesem Zusammenhang von einer „Grundnorm“, „Basisnorm“ oder „Supernorm“, vgl. Vollmer (1993), S. 184.

¹⁴ Siehe z. B. Maynard Smith (1982).

2. Die utilitaristischen Wurzeln der Spieltheorie

Die Spieltheorie nimmt an, dass die Personen, deren Verhalten sie untersucht, mit ihren Handlungen ein Ziel verfolgen, das einen gewissen Wert für sie hat. Mit dem Prinzip der Nutzenmaximierung greift die Spieltheorie utilitaristisches Gedankengut auf, so dass sie auch anfällig für Kritik an dieser Ethik wird. Daher soll in diesem Abschnitt zunächst der Utilitarismus behandelt werden, bevor in Abschnitt 3 einige grundlegende Gleichgewichtskonzepte der Spieltheorie vorgestellt werden.

Die wichtigste Regel des Utilitarismus ist das Prinzip der Nutzenmaximierung: Eine Handlung ist genau dann moralisch richtig, wenn sie optimal für das Wohlergehen aller betroffenen Personen ist. Ein Utilitarist geht demnach davon aus, dass alle moralisch relevanten Handlungen tatsächlich Auswirkungen auf den Nutzen des Handelnden selbst und der von seiner Handlung betroffenen Personen haben. So äußert sich z. B. John Stuart Mill, der den Nutzen allerdings allein auf körperliche Empfindungen beschränkt:¹⁵

“[P]leasure, and freedom from pain, are the only things desirable as ends; and [...] all desirable things (which are as numerous in the utilitarian as in any other scheme) are desirable either for the pleasure inherent in themselves, or as means to the promotion of pleasure and the prevention of pain.”

Die wesentlichen Merkmale¹⁶ einer utilitaristischen Ethik sind a) die Auffassung, dass das Wohlergehen der Gesellschaft der einzige Zweck moralischen Verhaltens ist, b) dass verschiedene Personen bezüglich ihres Grades an Wohlergehen vergleichbar sind und c) dass Handlungen nicht danach beurteilt werden, ob sie richtig oder falsch sind, sondern nur nach ihrem Potential zu Maximierung gewisser Werte (insbesondere des Nutzens). Damit sind vor allen Dingen die Folgen einer Handlung von Bedeutung, weshalb der Utilitarismus als eine Art Konsequentialismus aufgefasst wird.

Wenn man sagt, das Wohlergehen der Gesellschaft sei der einzige Zweck, dann ist damit gemeint, dass jede Handlung, die man vollzieht, und jeder Gegenstand, den man besitzt, Auswirkung auf die Wohlergehen haben, weil sie zusammengenommen den Nutzen von Personen bestimmen. Das Wohlergehen der Gesellschaft, meistens *Wohlfahrt* genannt, berechnet sich aus den Nutzenniveaus aller Mitglieder der Gesellschaft durch eine Wohlfahrtsfunktion. Im klassischen Utilitarismus (z. B. bei Jeremy Bentham) ist dies einfach die Summe oder der Durchschnitt der Nutzenniveaus aller Personen. Zu dem Nutzenniveau einer bestimmten Person tragen wiederum viele einzelne Dinge bei (die Güter), von denen

¹⁵ Siehe Mill (1861), Kap. 2.

¹⁶ Vgl. Chappell/Crisp (1998).

jedes einen individuellen Wert für diese Person, also einen Nutzen, besitzt. Im nächsten Abschnitt wird der Begriff des „Nutzens“ genauer erläutert.

2.1. Der Nutzen

Der *Nutzen* eines Gutes ist eigentlich eine abstrakte Größe, eine Zahl, die einer Handlung oder einem Gegenstand (einem Gut) zugeordnet wird. Dieser Wert wird von verschiedenen Anhängern des Utilitarismus unterschiedlich interpretiert. Die ersten Utilitaristen, Jeremy Bentham und John Stuart Mill,¹⁷ vertreten einen Hedonismus und setzen Nutzen mit Empfindungen wie Glück oder Lust gleich. George Edward Moore¹⁸ erweitert mit dem idealen Utilitarismus den Nutzenbegriff auf Güter, die sich nicht direkt auf das Glück eines Menschen auswirken, aber trotzdem einen gewissen „intrinsischen Wert“ haben, wie z. B. Wissen oder Weisheit.

Inzwischen hat sich jedoch mehr und mehr ein Präferenzutilitarismus etabliert, der den Nutzenwert gemäß individueller Vorlieben festlegt, die nicht mehr zwingend auf körperliche Empfindungen zurückgeführt werden müssen. Als Vorreiter dieser Spielart des Utilitarismus ist Richard M. Hare zu nennen.¹⁹

Der Wert eines Gutes entspricht dem Nutzen, der bei der Konsumierung des Gutes entsteht. Der Wert einer Handlung entspricht hingegen dem Nutzen des *Zustandes* (*state of affairs*), der durch die Handlung realisiert wird. Der Begriff „Zustand“ ist für sich genommen sehr vage, weshalb J. J. C. Smart betont, dass im Utilitarismus nur die Folgen einer Handlung zusammengenommen einen Zustand bilden. Man könne zwar genauso von dem Zustand des „gerade so gehandelt haben“ sprechen (Beispiel von Smart: „the state of having just kept a promise“), auf diese Weise müsste aber *jede* Ethik als utilitaristisch bezeichnet werden.²⁰ Auch Bernard Williams sieht dieses Problem,²¹ aber er wendet ein, dass man manche Aktivitäten allein um ihrer selbst willen begehle, so dass eine gewisse Befriedigung (*satisfaction*) durch diese Handlung bestünde, die ein Utilitarist ebenfalls berücksichtigen müsse. Williams zufolge bewirke die Verknüpfung des Wertes einer Handlung mit dem realisierten Zustand daher entweder zuviel Freiheiten oder zu große Einschränkungen bei der Einstufung eines ethischen Ansatzes: entweder sei jede Ethik utilitaristisch oder gar keine.²²

¹⁷ Vgl. Bentham (1789) und Mill (1861).

¹⁸ Vgl. G. E. Moore (1903).

¹⁹ Siehe z. B. Hare (1981).

²⁰ Vgl. Smart (1973), S. 13.

²¹ Vgl. Williams (1973), S. 83ff.

²² Vgl. ebd., S. 85.

Trotzdem ist das Zustands-Konzept zur Bewertung von Handlungen interessant, gerade *weil* es offenbar auch nicht-utilitaristische Moralvorstellungen handhaben kann.

Williams löst für sich das Problem mit dem Begriff „Zustand“, indem er als Grundlage von Entscheidungen den *intrinsic* Wert eines Gutes (oder auch einer Handlung!) annimmt. Für die Ethik, die er als „Utilitarismus“ bezeichnet, hat nur das Glück (happiness) einen intrinsischen Wert, so wie bei den Hedonisten Bentham und Mill, auf die im nächsten Unterabschnitt eingegangen werden soll.

2.1.1. Der Hedonismus

Für Jeremy Bentham lautet der oberste Zweck moralischen Handelns „the greatest happiness of the greatest number“. Schon in der antiken griechischen Philosophie kommt der Gedanke des Strebens nach Glück auf, der Eudämonie (εὐδαιμονία).²³ Bentham greift diese Idee auf und weitet sie zu einem moralischen Prinzip aus, auf das er seine Ethik gründet und das „Nutzenprinzip“ nennt. Dazu Bentham:²⁴

“By the principle of utility is meant that principle which approves or disapproves of every action whatsoever, according to the tendency it appears to have to augment or diminish the happiness of the party whose interest is in question: or, what is the same thing in other words to promote or to oppose that happiness.”

Das Ausmaß des Glücks, das eine Handlung bewirkt, dient Bentham also als Maßstab für den moralischen Wert einer Handlung. Der Begriff des Nutzens hat für Bentham die folgende Bedeutung:²⁵

“By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered: if that party be the community in general, then the happiness of the community: if a particular individual, then the happiness of that individual.”

Bentham ist an einer quantitativen Bestimmung des Nutzen interessiert und nennt einige Kriterien, die eine zahlenmäßige Bewertung von Glücksempfindungen ermöglichen:²⁶ Intensität, Dauer, Sicherheit des Eintreffens, zeitliche Entfernung, Fruchtbarkeit und Reinheit der Glücksempfindung, sowie die Anzahl der an dem Glück teilhabenden Personen.

²³ So z. B. bei Aristoteles in der Nikomachischen Ethik.

²⁴ Siehe Bentham (1789), Kap. 1, II.

²⁵ Siehe Bentham (1789), Kap. 1, III.

²⁶ Vgl. Bentham (1789), Kap. 4, II.-IV.

Offensichtlich ist nicht nur von Bedeutung, wie stark das Ausmaß einer Glücksempfindung ist, sondern auch wie lange man dieses Glück erfahren darf. Auch liegt der Gedanke nahe, dass der moralische Wert einer Handlung umso größer ist, je mehr Menschen von den Folgen dieser Handlung profitieren können. Die anderen Kriterien bedürfen jedoch der genaueren Erläuterung.

Die Folgen einer Handlung treten selten direkt im Anschluss an die Handlung ein, sondern sind meistens zeitlich verzögert. Es ist aber wahrscheinlich besser, jetzt hundert Euro zu bekommen als erst in fünf Monaten. Außerdem ist es möglich, dass eine erwartete Konsequenz noch verhindert wird, sich also nur mit einer gewissen Wahrscheinlichkeit ereignet. Eine Hundert-Euro-Note in der Hand hat demnach einen größeren Wert als ein Lotterielos für einen nur potentiellen Gewinn von hundert Euro, selbst wenn die Gewinnwahrscheinlichkeit groß ist.

Wenn eine Handlung mehrere Dinge gleichzeitig bewirkt, kann es darüber hinaus zu gegenläufigen Effekten kommen: eine Auswirkung ist gutartig, eine andere hingegen schadet jemandem. Bentham gibt zunächst zu bedenken, dass die Folgen einer Handlung weitere Folgen nach sich ziehen können. Eine Glücksempfindung ist demnach umso *fruchtbarer*, je wahrscheinlicher es ist, dass sie von weiteren Glücksempfindungen begleitet wird. Die *Reinheit* einer Glücksempfindung schließlich hängt davon ab, ob mit der ursächlichen Handlung gleichzeitig anderen Menschen Schaden zugefügt wurde. Ein Glück ist offensichtlich umso wertvoller, je weniger Kompromisse man zu seiner Erlangung in Kauf nehmen musste.

Zwar mag es zutreffen, dass das eigene Glück in einem stark positiven Zusammenhang mit dem eigenen Nutzen steht. Es gibt jedoch auch Menschen, die offenbar daran gefallen finden, wenn sie Schmerzen erleiden, so z. B. ein religiöser Fanatiker, der über Glasscherben oder glühende Kohlen geht. Der dem Hedonismus verwandte Gedanke, Nutzen auf physiologische Daten zurückzuführen, wird dadurch aber nicht grundsätzlich ad absurdum geführt. Beispielsweise zieht die moderne psychologische Stressforschung²⁷ Verbindungen zum „subjektiven Wohlbefinden“ (subjective well-being) der Versuchspersonen. Es scheint allerdings schwierig, über die physiologische Reaktion auf eine Handlung den Nutzen dieser Handlung auf eine Weise festzustellen, dass dieser Wert auch für zukünftige Handlungen desselben Typs verwendet werden kann. Eine Stressreaktion tritt nämlich nur bei einer Veränderung eines Zustandes ein. Das Glück, das durch eine Handlung ausgelöst werden kann, nimmt also mit der Zeit ab, bis sich der Organismus an diesen Zustand gewöhnt hat.

²⁷ Siehe z. B. Lazarus (1991), S. 406ff., für einen Überblick.

Außerdem muss der betreffenden Person *bewusst* sein, dass ihr eine Handlung Nutzen bringt, damit sie Glück empfindet. Wie der Begriff „subjektiv“ aber bereits andeutet, ist es in den meisten Fällen immer noch unumgänglich, die Versuchspersonen direkt nach ihren Vorlieben zu fragen, um die körperliche Reaktion richtig einordnen zu können. Damit kommt man letztendlich doch nicht um eine Nutzenbewertung im Sinne des Präferenzutilitarismus herum, der im nächsten Abschnitt beschrieben wird.

2.1.2. Der Präferenzutilitarismus

Besonders wenn der Nutzen einer Handlung für andere beteiligte Personen eingeschätzt werden soll, führt eine Identifikation von Nutzen mit Glück zu Schwierigkeiten, da man schon bei sich selbst kaum sagen kann, wie stark eine Glücksempfindung ist. Sowieso scheint es nicht möglich zu sein, in jedem Fall einen absoluten Wert einer Handlung oder eines Gegenstandes anzugeben, weil unterschiedliche Personen verschiedene Vorlieben haben.

Deshalb geht der Präferenzutilitarismus einen eher ungewohnten Weg: Anstatt nach einer möglicherweise willkürlichen Skala zu beurteilen, welcher Nutzen einer Handlung oder einem Gegenstand zukommt, wird beobachtet, welche Handlungen oder Gegenstände eine bestimmte Person anderen Handlungen oder Gegenständen vorzieht, um dann aus den Beobachtungsdaten eine Rangfolge zu erstellen, die es ermöglicht, jedem Objekt einen bestimmten Nutzen zuzuordnen.

Bereits John Stuart Mill vergleicht die Qualität zweier Glücksempfindungen, indem er die betroffenen Personen fragt, welche Empfindung ihnen lieber sei:²⁸

“Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure.”

Dieses Zitat deutet allerdings einen möglichen naturalistischen Fehlschluss bei einer Nutzenmessung durch Präferenzen an: Wenn der Nutzen einer Handlung daraus abgeleitet wird, wie stark diese Handlung bevorzugt wird, kann eine Ethik nicht mehr diejenige Handlung mit dem größten Nutzen als moralisch richtige Tat auszeichnen. Der „Fehlschluss“ sähe dann wie folgt aus:

A wird B vorgezogen. => A hat einen größeren Nutzen als B. => Im Utilitarismus hat A einen größeren moralischen Wert als B. => A sollte B vorgezogen werden.

Aus der *Tatsache* „A wird B vorgezogen.“ wird also offenbar die *Norm* abgeleitet „A sollte B vorgezogen werden.“ Es handelt sich jedoch an dieser Stelle um keinen wirklichen

²⁸ Siehe Mill (1861), Kap.2.

Fehlschluss, da der Utilitarismus mit dem Prinzip der Nutzenmaximierung eine weitere Prämisse einbringt, nämlich: „Wenn A einen größeren Nutzen als B hat, dann sollte A gegenüber B vorgezogen werden.“

Trotzdem hat diese rein relationale Nutzenbestimmung nur einen begrenzten Verwendungszweck in einer Ethik wie dem Utilitarismus, insbesondere wenn verschiedene Handlungsalternativen, die sich in mehreren Punkten unterscheiden gegeneinander abgewogen werden sollen. Schon die Entscheidung zwischen einem Korb mit 8 Äpfeln und 5 Birnen und einem anderen Korb mit 3 Äpfeln und 7 Birnen kann nur dann auf rationale Weise getroffen werden, wenn der Wert von Äpfeln und Birnen durch *Zahlen* ausgedrückt wird. In Abschnitt 3.1 wird jedoch gezeigt, dass in den meisten Fällen aus Vergleichen der Bewertungen einer Person von verschiedenen Situationen tatsächlich eine Nutzenfunktion abgeleitet werden kann, die auch die Einschätzung des Nutzens von neuartigen Situationen für diese Person ermöglicht.

Ein weiteres Problem des Präferenzutilitarismus besteht darin, dass sich Vorlieben mit der Zeit ändern können, häufig sogar von einem Tag auf den anderen. Richard B. Brandt ist daher der Meinung, dass die Unbeständigkeit von Präferenzen eine konsistente Nutzenmessung unmöglich macht.²⁹ Man stelle sich einen Reisenden in der Wüste vor, dessen Wasservorrat zu Ende geht. Sicher wird dieser Reisende angesichts der Gefahr des Verdurstens schon eine einzige Flasche Wasser in Gold aufwiegen. Wenn er aber urplötzlich soviel Wasser bekommt, wie er trinken kann, wird sein Durst schnell gestillt sein, vielleicht sogar in dem Maße, dass er für einige Zeit kein Wasser mehr sehen kann. Falls aber nun eine moralische Handlung dieses Reisenden von seinem Bedürfnis nach Wasser abhängt (vielleicht bei der Frage, ob er den Wasservorrat eines Mitreisenden aufbrauchen soll), dann kommt der Präferenzutilitarismus unter Umständen zu unterschiedlichen Handlungsempfehlungen vor und nach der Veränderung der Vorlieben des Reisenden.

Auf den ersten Blick scheint die Wandelbarkeit von Präferenzen nicht schlimm zu sein, da man bei einer Person in der Regel jeweils vor und nach einer derartigen Präferenzumkehr konsistente Vorlieben feststellen kann. Wenn man die Ursache dieser Veränderung kennt, könnte man also zusätzlich die jeweiligen Umstände angeben, unter denen eine bestimmte Präferenzordnung Geltung besitzt. Falls die Präferenzumkehr allerdings während einer Entscheidungssituation geschieht, insbesondere nach Ausführung einer Handlung aber vor Eintreten der Folgen, so wird die betroffene Person wahrscheinlich im Nachhinein sagen, dass sie lieber eine andere Handlung gewählt hätte. Jedoch kann nur eine plötzlich auftretende,

²⁹ Vgl. Brandt (1979), S. 249f.

unvorhersagbare Änderung der Präferenzen verhindern, dass der Präferenzutilitarismus zu einer konsistenten Entscheidung verhilft. Bei wichtigen Entscheidungen dürften die Gründe für einen derartigen Stimmungsumschwung hingegen offenkundig sein, so dass es normalerweise keine Schwierigkeiten geben sollte. Zwar ist niemand davor gefeit eine bereits getroffenen Entscheidung angesichts neuer Informationen zu bereuen. Das ist aber noch kein Anzeichen für einen inkonsistenten Entscheidungsmechanismus.

2.1.3. Der Nutzen in der Spieltheorie

Die Spieltheorie übernimmt weitestgehend das Nutzenkonzept des Utilitarismus. Zur Bewertung von Handlungen werden ebenfalls die durch eine Handlung realisierten Zustände (states of affairs) herangezogen, weil in den meisten Fällen, *mehrere* Handlungen von unterschiedlichen Personen für ein bestimmtes Ergebnis verantwortlich sind. Dieses Ergebnis einer Kombination von Handlungen wird als *Auszahlung* (payoff) bezeichnet, da es bei spieltheoretischen Anwendungen in den Wirtschaftswissenschaften normalerweise um finanzielle Entscheidungen geht. Der Nutzen ist die subjektive Bewertung der Auszahlung, also der Zahlenwert, den die Nutzenfunktion einer Person für eine bestimmte Auszahlung liefert. So haben hundert Euro für einen Menschen vielleicht einen Nutzen von 5, für einen anderen einen Nutzen von 7.

Anders als im Utilitarismus sind die Zustände aber nicht auf eine Zusammenfassung der Folgen der Handlungen beschränkt. Der Nutzen eines Zustandes setzt sich vielmehr aus verschiedenen Bestandteilen zusammen:

1. Die Bewertung der (beobachtbaren) Folgen der Handlung
2. Die Bewertung der Handlung selbst, nach:
 - a) der Handlungsweise (Reihenfolge der Handlungen, sowie Abhängigkeit von den Handlungen anderer Personen und zufälligen Ereignissen),
 - b) der Absicht des Akteurs bei der Handlung.

Für die Bewertung der Folgen einer Handlung ist zunächst einmal der Wert der durch die Handlung umverteilten Güter von Bedeutung, wobei mit „Gut“ an dieser Stelle alles das gemeint ist, für das ein Preis angegeben werden kann. Tatsächlich wird der Nutzen in den empirischen Wirtschaftswissenschaften häufig bestimmt, indem die betreffenden Personen nach dem Geldbetrag gefragt werden, den sie für die Realisierung eines bestimmten Zustandes zahlen würden, insbesondere dann, wenn es sich um ein eher abstraktes Gut handelt, wie z. B. Informationen jeder Art oder die Verbesserung der Lebensqualität durch den Bau einer Umgehungsstraße, die zu einer Reduzierung der Lärmbelastung führen würde.

Selbst der Wert eines Menschenlebens wird in einigen Fällen in Geld gemessen, z. B. dann wenn eine Versicherung die finanziellen Schwierigkeiten einer Familie durch den Tod des Versorgers verhindern soll. Allerdings kann diese zweifelhafte Praxis in manchen ethischen Problemen umgangen werden, indem der Nutzen eines Menschenlebens einfach gleich unendlich gesetzt wird, also so hoch, dass die Bewahrung dieses Lebens auf jeden Fall vorzuziehen ist.

Der Teil des Nutzens, der auf Unterschieden in der Handlungsweise beruht, kann in einigen Fällen ebenfalls auf die Realisierung eines anderen Zustandes zurückgeführt werden und damit indirekt als Folge der Handlung aufgefasst werden. Beispielsweise kann man eine Wanderung machen, um einen bestimmten Ort zu erreichen, obwohl man dasselbe Ziel auch mit dem Auto hätte erreichen können. Die Handlung besteht in diesem Fall darin, eine Wegstrecke zurückzulegen, mit der Folge, dass man an dem besagten Ort ankommt. Die Entscheidung zwischen wandern und Auto fahren macht sich aber trotzdem im Ergebnis bemerkbar, sei es durch Benzinkosten, aufgewendete Zeit, körperliche Anstrengung und auch physiologische Anzeichen der Freude darüber, ein paar Stunden an der frischen Luft zu verbringen. Der Zustand, der hier bewertet wird, ist also nicht „das Erreichen des Zieles“, sondern „das Erreichen des Zieles durch eine Wanderung“ bzw. „das Erreichen des Zieles durch eine Autofahrt“.

Bei der Bestimmung des moralischen Wertes einer Handlung spielen – wenigstens in manchen Ethiken – auch die Gründe eine Rolle, aus denen diese Handlung ausgeführt wurde. Teilweise spiegelt sich diese Tatsache in den Anforderungen an die intellektuellen Fähigkeiten bzw. die Rationalität der Personen wider, die die entsprechende ethische Theorie für eine moralische Entscheidung heranziehen. Eine Handlung ist nur dann wirklich moralisch wertvoll, wenn sie und ihre Folgen auch so geplant werden, wie sie sich ereignen, also wenn es *gewollt* war, dass alles auf diese Weise passiert und nicht anders. Wenn jemand rein zufällig eine moralisch richtige Entscheidung trifft, wird das nicht auf sein pflichtgemäßes Verhalten zurückgeführt, sondern vielmehr auf ein glückliches Händchen oder eine gute Intuition. Eine Ethik unterstützt aber normalerweise nur Verhalten, das in ähnlichen Situationen wiederholt werden kann, so dass die Planbarkeit von Entscheidungen unverzichtbar ist und damit auch die Absicht hinter einer Handlung nachvollziehbar sein muss. Natürlich verfügt jemand, der sich Gedanken über Ethik macht, in der Regel auch über die geforderte Vernunft, um wenigstens in wichtigen Situationen „von selbst“, d. h. ohne Zufall, zur moralisch richtigen Entscheidung zu kommen. Es sieht aber vielleicht schon

anders aus, wenn man selbst nicht aktiv an der Handlung beteiligt ist, sondern vom Verhalten einer anderen Person abhängig ist.

Für die Bewertung der Folgen einer Handlung ist es meist weniger von Bedeutung, aus welchen Gründen man selbst, eine Entscheidung getroffen hat, sondern vielmehr welche Absichten andere beteiligte Personen bei ihren Handlungen hatten. Wenn ein gewisses Ereignis passiert, weil Menschen einander geholfen haben, so sorgt das möglicherweise für eine Verbindung zwischen diesen Menschen und hat deshalb einen größeren moralischen Wert, als ein identisches Ereignis, das nur als Nebeneffekt der Handlungen derselben Menschen entstanden ist, obwohl diese eigentlich andere Ziele verfolgten oder sich sogar schaden wollten.

Ein (zugegebenermaßen etwas konstruiertes) Beispiel: Person A besitzt 10 Euro, die sie mit Person B teilen will. Dabei kommt es zu zwei Szenarien:

- 1) Person A gibt Person B 3 Euro und behält den Rest.
- 2) Bevor Person A das Geld aufteilen kann, bekommt sie die Auflage, dass sie maximal 3 Euro abgeben darf. Danach reicht Person A wieder 3 Euro an Person B weiter und behält den Rest.

Die Frage ist nun: Wie bewertet Person B, die alle Details mitbekommen hat, die Entscheidung von Person A? Armin Falk und Urs Fischbacher³⁰, die dieses Beispiel in einem spieltheoretischen Modell darstellen, würden sagen, dass Szenario 1 ein schlechteres Licht auf Person A wirft als Szenario 2, obwohl es in beiden Fällen zu derselben Aufteilung kommt. Denn in Szenario 1 hätte Person A die Möglichkeit gehabt, das Geld gerecht aufzuteilen, so dass jeder die Hälfte bekommt, während sie in Szenario 2 mit 3 Euro zu 7 Euro die faireste erlaubte Aufteilung wählt. Wenn Person B Szenario 2 gegenüber Szenario 1 vorzieht, dann ist es ihr offenbar lieber, wenn andere Menschen ihr gegenüber bei Handlungen gute Absichten zeigen. Oder anders ausgedrückt: Person B erfährt einen negativen Nutzen durch die Absicht von Person A, eine unfaire Aufteilung vorzunehmen.

2.1.4. Interpersoneller Nutzenvergleich

Bei moralischen Entscheidungen ist es häufig erforderlich, einer Person zu schaden, damit andere Personen davon profitieren können. Wenn ein Nutzenkonzept sinnvoll angewendet werden soll, muss es daher möglich sein, die Präferenzen von verschiedenen Personen miteinander zu vergleichen, um sagen zu können, wann ein solches Opfer moralisch gerechtfertigt werden kann.

³⁰ Siehe Falk/Fischbacher (2000)

Der Vergleich der Nutzenniveaus unterschiedlicher Personen benötigt die Annahme, dass ähnliche Handlungen oder Gegenstände einen ähnlichen Nutzen erzeugen. Wenn Person A einen grünen Apfel isst und auch Person B einen grünen Apfel isst, dann sollte man mangels Hinweise zum Gegenteil davon ausgehen können, dass beide Personen durch diese Handlung einen ähnlichen Nutzen bekommen haben. Der Nutzen ist also eine objektive Größe, die den Vergleich zwischen unterschiedlichen Personen überhaupt erst möglich macht.

Es scheint nicht unplausibel zu sein, dass die meisten Menschen ähnlich denken, und deswegen auch zu ähnlichen Ergebnissen bei der Nutzenbewertung kommen. Schließlich ist es in den meisten Fällen möglich, eine Entscheidung nachzuvollziehen, wenn man sich in die Sichtweise des anderen hineinversetzt. Man kann also so tun, als besäße man in einer bestimmten Situation die Präferenzen einer anderen Person. John C. Harsanyi stellt daher folgendes „Ähnlichkeitspostulat“ (similarity postulate) auf:³¹

“[O]nce proper allowances have been made for the empirically given differences in taste, education, etc., between me and another person, then it is reasonable for me to assume that our basic psychological reactions to any given alternative will be otherwise much the same.”

Leider muss Harsanyi zugeben,³² dass diese Tatsache nicht durch direkte Beobachtungen überprüft werden kann, da man dazu wissen müsste, was in den betroffenen Personen tatsächlich vorgeht. Für ihn ist das Ähnlichkeitspostulat eine a priori gültige Aussage, die er damit rechtfertigt, dass jede Abweichung zweier verschiedener Personen in der Nutzenbewertung, die auf einen beobachtbaren Unterschied zwischen diesen Personen zurückgeführt werden könne, ja bereits vorher berücksichtigt worden sei. Warum solle man aber annehmen, dass es noch weitere Unterschiede gebe, die man *nicht* beobachten kann?

Ken Binmore nutzt die Wandelbarkeit von Vorlieben als Argument für die interpersonelle Vergleichbarkeit von Präferenzen.³³ Wenn sich die Präferenzen mit der Zeit ändern, wird sich ein gewisser Trend erkennen lassen, weil sich im Sinne einer kulturellen Evolution nur bestimmte Präferenzen bewähren werden. Anstatt diese Unbeständigkeit als Problem der Nutzenmessung anzusehen,³⁴ behauptet Binmore, dass durch die Veränderung von Präferenzen sogar erst ein Nutzenvergleich ermöglicht wird, weil sich über die Generationen hinweg ähnliche Präferenzstrukturen herausgebildet haben.³⁵ Menschen haben deshalb nicht zwangsläufig gleiche Vorlieben, aber sie sind sich ähnlich genug, um die Präferenzen ihrer Mitmenschen genügend einschätzen zu können. Im weiteren Verlauf der Arbeit werden

³¹ Siehe Harsanyi (1982), S. 50.

³² Vgl. ebd., S. 51.

³³ Vgl. Binmore (1994), S. 86ff. und S. 285ff.

³⁴ Siehe Abschnitt 2.1.2.

³⁵ Vgl. Binmore (1994), insbesondere S. 87.

Ergebnisse der Evolutionären Spieltheorie beschrieben, die Binmores Argument stützen können. Beispielsweise zeigen Helmut Bester und Werner Güth³⁶, dass altruistische Präferenzen einen evolutionären Vorteil in Situationen haben, in denen eine Zusammenarbeit allen Partnern einen größeren Nutzen beschert.

2.2 Akt- vs. Regel-Utilitarismus

Der Utilitarismus wird häufig kritisiert, weil er in der Praxis nur schwer umzusetzen ist:³⁷ bei jeder moralischen Entscheidung muss zunächst der Nutzen jeder Alternative berechnet werden, was in komplexeren Situationen sehr zeitaufwendig werden kann. Dabei ähneln sich viele moralische Probleme, so dass es nahe liegt, einmal erlangte Erkenntnisse auch in verwandten Entscheidungssituationen einzusetzen. Aus diesem Gedanken entspringt die Unterscheidung in Akt- oder Regel-Utilitarismus, je nachdem ob nur eine einzelne Handlung und ihre Folgen bewertet werden oder die Folgen einer allgemeinen Regel dafür, wie man sich in einer ähnlichen Situation verhalten sollte. Der Akt-Utilitarismus repräsentiert die klassische Sichtweise, die wegen ihres Problems der schwierigen Anwendbarkeit kritisiert wird. Hingegen bietet der Regel-Utilitarismus eine Möglichkeit der Vereinfachung, weil viele ethische Fragestellungen mit einer einzigen moralischen Norm entschieden werden können.

G. Scarre beschreibt drei Arten des Regel-Utilitarismus (RU):³⁸ den idealistischen RU, den „actual state“ RU und den konditionalen RU.

Der idealistische Regel-Utilitarismus ist die strengste Variante, da hier eine Regel unter keinen Umständen gebrochen werden darf, selbst wenn andere Personen diese Regel nicht befolgen. Eine Handlung ist demnach genau dann moralisch richtig, wenn diese Handlung in Übereinstimmung mit einem Bündel von Regeln geschieht, deren generelle Befolgung den Nutzen maximieren würde.

Etwas schwächer ist der „actual state“ Regel-Utilitarismus, weil die Handlung in dieser Variante auch tatsächlich mehrheitlich in der betreffenden gesellschaftlichen Gruppe akzeptiert sein muss. Nur wenn ein großer Teil der Gesellschaft bereit ist, diese Handlung zu unterstützen, sollte sie auch ausgeführt werden.

Wenn schließlich das Befolgen einer Regel an weitere Umstände geknüpft wird, spricht man von einem konditionalen Regel-Utilitarismus. In jedem Einzelfall wird dabei überprüft, ob die Anwendung der betreffenden moralischen Regel den Nutzen maximiert. Diese Variante ähnelt sehr dem Akt-Utilitarismus, weshalb David Lyons auch zu dem Ergebnis

³⁶ Vgl. Bester/Güth (1998).

³⁷ Vgl. z. B. Williams (1973), S. 137.

³⁸ Vgl. Scarre (1996).

kommt, dass der Regel-Utilitarismus mit dem Akt-Utilitarismus zusammenfällt.³⁹ Wenn in jeder einzelnen Situation eine spezielle Regel für das moralisch richtige Verhalten formuliert werden muss, dann geht man letztendlich genauso vor wie ein Akt-Utilitarist, der in jeder dieser Situationen *direkt* das richtige Verhalten bestimmt.

J. J. C. Smart⁴⁰ bevorzugt einen Akt-Utilitarismus, da seiner Ansicht nach nur eine einzige Regel eines Regel-Utilitarismus ohne Ausnahme gelten kann, nämlich: „Maximiere den wahrscheinlichen Nutzen (maximize probable benefit)!“ Nach dieser Regle entscheidet aber auch ein Akt-Utilitarist, wenn er eine Situation bewertet.

Richard M. Hare⁴¹ sieht hingegen Vorteile beim Regel-Utilitarismus, da es viel zu aufwendig sei, wenn man im Sinne des Akt-Utilitarismus jede moralische Entscheidungssituation neu analysieren müsse. Zur Bewältigung komplexer moralischer Probleme sei demnach ein gewisses Grundverständnis von Ethik nötig, das durch Lernprozesse verinnerlicht wird, genauso wie ein Autofahrer ohne größeres Nachdenken mit den Pedalen und der Gangschaltung umgeht und damit seine gesamte Aufmerksamkeit dem Straßenverkehr widmen kann. Eine Ethik muss nach Ansicht von Hare also Verhaltensvorschriften liefern, die nicht nur für Einzelfälle gelten, sondern auch zukünftig in ähnlichen Situationen angewendet werden können.

Der Akt-Utilitarismus ist nach Darstellung von Franz von Kutschera⁴² eine teleologische Ethik, weil Präferenzen für Handlungen und daraus resultierende Zustände gebildet werden, weshalb sich der Wert einer Handlung aus dem Wert ihrer Folgen ergibt. Ein Anhänger des Regel-Utilitarismus hat jedoch, so von Kutschera, Präferenzen für Handlungsweisen. Eine solche Ethik muss also als deontologisch bezeichnet werden.

Der Regel-Utilitarismus wird häufig so interpretiert, dass zwei Personen identisch handeln sollen, wenn sie sich in einer symmetrischen Situation begegnen, in welcher Handlungsalternativen und erreichbare Zustände bei beiden Personen übereinstimmen. Angenommen zwei Menschen begegnen sich auf der Straße und können sich entweder einen Guten Tag wünschen oder wortlos aneinander vorbeigehen. Existiert nun eine Regel für diese Situation, die da lautet „Begrüße andere Menschen!“, dann sagen beide Personen „Guten Tag“. Oder es kommt eine Regel zur Anwendung wie „Grüße niemals andere Menschen!“. Dann wird keiner der beiden einen Ton von sich geben. Ein anderes Verhalten kann aber nach von Kutschera in dieser Situation nicht durch eine Regel festgeschrieben werden. Entweder

³⁹ Vgl. Lyons (1965).

⁴⁰ Vgl. Smart (1973).

⁴¹ Vgl. Hare (1981), S. 36f.

⁴² Vgl. von Kutschera (1982).

befolgen alle beteiligten Personen eine Regel oder keiner. Die Probleme dieser Interpretation des Regel-Utilitarismus können mit Hilfe der Spieltheorie analysiert werden. Daher soll die weitere Diskussion des Regel-Utilitarismus in Abschnitt 5.1 weitergeführt werden, nachdem einige spieltheoretische Konzepte erläutert wurden.

2.3 Deontologismus vs. Teleologismus

Zum Abschluss dieses Abschnittes sollen noch einige Begriffe geklärt werden, die in der Utilitarismus-Diskussion eine Rolle spielen: Dem Utilitarismus wird häufig vorgeworfen, „teleologisch“ und „konsequenzialistisch“ zu sein; wohingegen eine „deontologische“ Ethik als Gegenspieler der utilitaristischen Ethik aufzufassen ist. Eine *teleologische* Ethik verfolgt mit ihrer Moral einen gewissen Zweck (griech. τέλος), so z. B. der Utilitarismus den Zweck der Nutzenmaximierung. Eine *konsequenzialistische* Ethik legt bei der moralischen Beurteilung einer Handlung einen besonderen Wert auf die Folgen dieser Handlung, während eine *deontologische* Ethik die Folgen von Handlungen ignoriert und als einzigen „Zweck“ von Moral ein Handeln aus Pflichtgefühl zulässt. Wie in Abschnitt 5 ausführlicher dargestellt werden soll, ist es für eine Hilfswissenschaft wie die Spieltheorie wichtig einen neutralen Standpunkt einzunehmen, wenn sie sich nicht schon von vornherein auf eine moralische Sichtweise beschränken lassen will. Daher sollte sie sich weder als konsequenzialistisch oder teleologisch, noch als deontologisch bezeichnen lassen müssen.

In der Literatur erfolgt keine einheitliche Abgrenzung von konsequenzialistischen und teleologischen Ethiken, auch wenn sie in der Regel als Gegenpositionen der deontologischen Ethik betrachtet werden. Julian Nida-Rümelin charakterisiert z. B. den Unterschied zwischen deontologischen und teleologischen Ansätzen wie folgt:⁴³

„Deontologische Theorien sind solche, die Handlungen (Handlungsregeln oder generischen Handlungen) einen intrinsischen Wert zuweisen, während für teleologische Theorien eine Handlung niemals intrinsisch (moralisch) wertvoll (oder wertlos) ist, sondern ihren Wert nur durch ihren Beitrag zum verpflichtungsunabhängigen Guten erhält.“

Auch William K. Frankena interpretiert den Begriff „teleologisch“ so, dass das moralische Entscheidungskriterium einer derartigen Ethik ein außermoralischer Wert ist, der mit einer Handlung geschaffen wird.⁴⁴ Allerdings zieht er zur Bestimmung dieses Wertes die Folgen der Handlung heran, so dass für ihn offenbar Konsequenzialismus und Teleologismus übereinstimmen. Dies wird insbesondere daran deutlich, dass Frankena eine deontologische

⁴³ Siehe Nida-Rümelin (1993), S. 65.

⁴⁴ Vgl. Frankena (1963), S. 13ff.

Theorie von einer teleologischen dadurch abgrenzt, dass erstere ein pflichtgemäßes Handeln eben *nicht* von den Folgen der Handlung abhängig macht.

John Rawls orientiert sich, wie er selber einräumt, an der Darstellung von Frankena, wenn er sagt, in einer teleologischen Theorie sei das Gute unabhängig von dem Rechten definiert, während das Rechte das Gute maximiere.⁴⁵ Genauso beschreibt jedoch Nida-Rümelin eine *konsequenzialistische* Theorie.⁴⁶

Bernard Williams macht allerdings deutlich, dass eine konsequenzialistische Ethik den moralischen Wert einer Handlung *immer* aus den Folgen dieser Handlung ableitet. Dies ist für ihn sogar ein Grund, den Konsequenzialismus abzulehnen, da jede Folge einer Handlung weitere Folgen nach sich zieht, so dass eine Nutzenbestimmung dieser Art nie zu einem Ende kommen würde.⁴⁷ Eine Handlung ist der konsequenzialistischen Sichtweise zufolge nur dann moralisch richtig, wenn sie einen Zustand (state of affairs) herbeiführt, der besser ist als alle anderen realisierbaren Zustände.⁴⁸

Der „intrinsic Wert“ einer Handlung resultiert also im Konsequenzialismus allein aus dem Zustand, der durch diese Handlung realisiert wird. Eine teleologische Ethik könnte aber grundsätzlich auch Zwecke verfolgen, die sich *nicht* in den Folgen einer Handlung bemerkbar. Während der Teleologismus damit den Zweck des moralischen Verhaltens über die eingesetzten Mittel stellt, bewertet eine deontologische Ethik hingegen *ausschließlich* die Mittel, mit denen ein bestimmtes Ergebnis erreicht wird. Wie in Abschnitt 4.2 dargestellt wird, kann die Spieltheorie allerdings einen Kompromiss herbeiführen, indem sie sowohl eine teleologische Komponente, als auch eine deontologische Komponente in ihren Modellen berücksichtigt.

Möglicherweise muss aber der Begriff „konsequenzialistisch“ viel weiter gefasst werden, weil die Definition des „Zustandes“ einen großen Spielraum lässt. In der großzügigsten Auslegung müsste dann selbst eine deontologische Ethik als „konsequenzialistisch“ bezeichnet werden. Bernard Williams sieht einen entscheidenden Punkt dieses Problems darin, ob Handlungen ebenfalls ein intrinsischer – d.h. nicht abgeleiteter – Wert zukommen kann. Denn: „To insist that what has intrinsic value are states of affairs and not actions seems to come near an important feature of consequentialism.“⁴⁹ Wenn ein Konsequenzialist einer Handlung einen Wert zuschreibt, dann nur, weil sie die „kausale Eigenschaft“ hat, einen bestimmten nützlichen Zustand zu erreichen. Der Wert einer Handlung leitet sich also aus

⁴⁵ Vgl. Rawls (1999), S. 21f. bzw. die Fußnote auf S. 22.

⁴⁶ Vgl. Nida-Rümelin (1993), S. 87.

⁴⁷ Vgl. Williams (1973), S. 82.

⁴⁸ Vgl. ebd. S. 88.

⁴⁹ Ebd. S. 84.

dem Wert des Zustandes ab, den sie ermöglicht. Wenn man gerne gute Taten vollbringt, dann liegt das also daran, dass man Freude über das Ergebnis dieser Handlung empfindet.

Aber andererseits gesteht Williams den Utilitaristen zu, dass sie manche Handlungen nur um ihrer selbst willen vollbringen.⁵⁰ Der Widerspruch, den er damit konstruiert, scheint jedoch gar keiner zu sein: Ein Konsequentialist darf Handlungen keinen intrinsischen Nutzen zukommen lassen; für einen Utilitaristen (nominell ebenfalls Konsequentialist) haben hingegen wenigstens einige Arten von Handlungen durchaus intrinsischen Wert. Hier besteht doch eigentlich nur dann ein Problem, wenn der Utilitarismus als eine Art Konsequentialismus aufgefasst wird. Leider zieht Williams nicht in Betracht zu überprüfen, ob der Utilitarismus tatsächlich unter seinen Begriff von Konsequentialismus fällt.

Mit der Anwendung des Prinzips der Nutzenmaximierung kommt der Ruf nach einer mathematischen Theorie zur Berechnung der Nutzenwerte: Der Wert jeder Handlung für jede beteiligte Person muss exakt bestimmt werden, um dann eine optimale Strategie zu bestimmen. Für einen Utilitaristen ist es also unumgänglich, eine mathematische Entscheidungstheorie zu Rate zu ziehen. Im nächsten Abschnitt soll dafür argumentiert werden, dass die Spieltheorie dafür am besten geeignet ist.

⁵⁰ Vgl. ebd. S. 84f.

3. Die Spieltheorie als strategische Entscheidungstheorie

Dieser Abschnitt der Arbeit erfüllt zwei Aufgaben. Zunächst einmal soll eine kurze Einführung in die Spieltheorie gegeben werden, indem dem Leser einige grundlegende Gleichgewichtskonzepte vorgestellt werden. Gleichzeitig soll erläutert werden, warum einer strategischen Herangehensweise an Entscheidungssituationen der Vorzug gegenüber anderen Arten von Entscheidungstheorien zu geben ist. Ein wichtiger Zweck dieses Abschnittes ist also die Begründung der folgenden These:

These 1: Weil eine vollständige Modellierung einer Ethik die Berücksichtigung strategisch motivierter Entscheidungen erfordert, ist die Spieltheorie anderen Entscheidungstheorien überlegen.

Die Aufgabe der Spieltheorie besteht nämlich darin, bei Entscheidungen zu helfen, deren Ergebnis von den Handlungen anderer Personen beeinflusst wird. Im Mittelpunkt steht dabei die Bewertung und Auswahl von Strategien. Die Spieltheorie ist dabei in erster Linie ein Mechanismus zur Lösung eines mathematischen Optimierungsproblems, dessen Ergebnis erst durch die Verknüpfung mit der Zielvorgabe einer normativen Theorie wie der Ethik oder den Wirtschaftswissenschaften einen tieferen Sinn bekommt. Was die Spieltheorie „Nutzen“ nennt ist nur eine Zahl, die den Wert eines Ergebnisses für einen bestimmten Menschen repräsentiert, ohne dass damit schon eine Festlegung getroffen wird, wie genau dieser Wert zu errechnen ist. Trotzdem macht auch ein Spieltheoretiker eine normative Aussage, sobald er sagt: „Wenn das dein Ziel ist, wähle die und die Handlung.“

Ein *Spiel* ist im Grunde ein Schema von Handlungskombinationen aller *Spieler*, d. h. der beteiligten Personen, und des Zustandes (state of affairs), der durch diese Handlungen realisiert wird. In der Regel beschränkt sich der realisierte Zustand jedoch auf die Konsequenzen der Handlungen. Bevor in Abschnitt 3.2 die Bestandteile des Spieles etwas genauer erläutert werden, soll in Abschnitt 3.1 kurz auf die Bayesianische Entscheidungstheorie eingegangen werden, die ebenfalls das Nutzenkonzept des Utilitarismus verwendet, sich aber nur auf eine einzige Person als aktiven Entscheider konzentriert. Mit der Evolutionären Spieltheorie wird danach in Abschnitt 3.3 ein Gleichgewichtskonzept dargestellt, das die Annahmen der klassischen Spieltheorie an die Vernunft der Spieler lockert. In Abschnitt 3.4 schließlich wird die Spieltheorie der Bayesianischen Entscheidungstheorie gegenübergestellt.

3.1 Die Bayesianische Entscheidungstheorie

Die Spieltheorie baut auf die Bayesianische Entscheidungstheorie auf und teilt sich mit ihr einige grundlegende Annahmen über die Bewertung von Strategien und die Rationalität der Spieler. Außerdem gibt es Situationen, in denen sich die zuerst offenbar komplexe Struktur eines Spieles auf ein einfacheres Entscheidungsproblem reduziert, bei welchem nur noch die optimale Handlung eines einzigen Entscheiders optimiert werden muss. In diesem Fall führt die Bayesianische Entscheidungstheorie die Arbeit der Spieltheorie zu Ende.

Die Anforderungen an die Rationalität der Spieler machen sich vor allen Dingen bei der Konstruktion der Nutzenfunktion bemerkbar. Damit nämlich die Vorlieben der Spieler in eine Nutzenskala übersetzt werden können, müssen die zum Nutzenvergleich verwendeten Präferenzrelationen *vollständig* und *transitiv* sein.⁵¹

Mit Vollständigkeit ist gemeint, dass alle durch Handlungen realisierbaren Zustände miteinander verglichen werden können. Ein Zustand A ist also immer besser, schlechter oder genauso gut wie ein Zustand B. Die Vergleichbarkeit von zwei Handlungsalternativen ist allerdings unter Umständen schon nicht mehr möglich, wenn zwei von einander unabhängige Größen für den Nutzen des erreichten Zustandes relevant sind. Angenommen eine Ethik verlangt, dass eine moralische Handlung den Nutzen jedes einzelnen Menschen getrennt maximiert.⁵² Dann ist eine Situation, in der Person A einen Nutzen von 5 und Person B einen Nutzen von 2 erhält, nicht vergleichbar mit einer anderen Situation, in der Person B einen Nutzen von 5 und Person A einen Nutzen von 2 bekommt. Denn aus Sicht von Person A wäre die erste Situation besser, während Person B die zweite Situation vorziehen würde. Das angenommene ethische Prinzip verlangt aber, dass *beiden* Personen zu einem höheren Nutzen verholfen wird. In der Praxis dürfte es jedoch recht einfach sein, die Vollständigkeit der Präferenzrelation einer Person zu erreichen, indem man die Person in unklaren Situationen fragt, welchen Zustand sie vorziehen würde. Falls auch die Person selbst zu keinem Vergleich in der Lage ist, dürfte es unproblematisch sein, wenn man annimmt, dass sie zwischen den beiden Zuständen indifferent ist und ihnen deshalb denselben Nutzenwert zuordnet.

Wenn drei oder mehr verschiedene Zustände miteinander verglichen werden, sichert die Forderung nach Transitivität, dass durch die Präferenzrelation keine Widersprüche entstehen. Angenommen Zustand A ist besser als Zustand B, während Zustand B wiederum besser als

⁵¹ Für eine formale Definition dieser Annahmen siehe z. B. Mas-Colell/Whinston/Green (1995), S.6.

⁵² Dies hat übrigens nichts mit dem üblichen utilitaristischen Prinzip der Nutzenmaximierung zutun, da weder der Gesamt- noch der Durchschnittsnutzen der beiden Personen für die Entscheidung von Bedeutung sind. Vielmehr wird nach der *pareto-besseren* Handlung gesucht, also einer Handlung, die keine Person schlechter und mindestens eine Person besser stellt. (Vgl. Mas-Colell/Whinston/Green (1995), S. 313)

ein dritter Zustand C ist. Dann muss ein Vergleich von A und C zu dem Ergebnis führen, dass A auch besser als C ist, wenn die Präferenzen konsistent bleiben sollen.

Andreu Mas-Colell, Michael D. Whinston und Jerry R. Green bezeichnen eine Präferenzrelation, die sowohl vollständig als auch transitiv ist, als *rational*. Eine derartige Präferenzrelation kann *immer* in eine Nutzenfunktion übersetzt werden, wenn die Menge der verglichenen Zustände endlich ist,⁵³ also wenn es nicht zu viele Alternativen gibt, die mit Nutzenwerten belegt werden müssen.

Ein bloßer qualitativer Vergleich von Alternativen – z. B. „B ist besser als C, aber schlechter als A“ – kann in einer Nutzenfunktion dargestellt werden, indem der besten Alternative eine hohe Zahl, der zweitbesten Alternative eine etwas niedrigere Zahl, usw. zugeordnet wird, bis hinunter zur schlechtesten Alternative, die die niedrigste Zahl bekommt. Der genaue Wert der Zahlen spielt dabei keine Rolle, solange die Reihenfolge erhalten bleibt. Wenn jemand zwischen zwei Alternativen indifferent ist, erhalten beide Alternativen denselben Nutzenwert. Die Stärke der Präferenzen, die für einen interpersonellen Nutzenvergleich von Bedeutung ist, kann wenigstens ansatzweise aus einer derartigen Nutzenfunktion abgelesen werden, indem man den Abstand zweier Alternativen, die in der Rangordnung direkt aufeinander folgen, als Vergleichsmaßstab festlegt. Aus „A ist besser als B“ und „B ist besser als C“ würde dann folgen: „A ist doppelt so gut wie C“.⁵⁴

Der Vorteil einer numerischen Darstellung der Präferenzen liegt in der Möglichkeit, durch das Mischen bereits verfügbarer Handlungen neue Handlungsalternativen zu erzeugen. Wenn man beispielsweise sein Geld an der Börse investieren will, wird man normalerweise nicht sein ganzes Geld auf ein einziges Unternehmen setzen, sondern über mehrere Aktien streuen, um das Risiko des Verlustes zu verringern, falls eine Firma einmal Insolvenz anmelden sollte. Zusätzlich zu einer begrenzten Anzahl von *reinen Strategien*, bei welchen das gesamte Geld in jeweils ein Unternehmen investiert wird, existieren eine Vielzahl weiterer *gemischter Strategien*, bei welchen auf mehrere Aktienwerte jeweils ein bestimmter Anteil der Geldes verteilt wird.

Eine *gemischte Strategie* ist besonders dann sinnvoll, wenn die Auswirkungen einer Handlung unsicher sind. Ein Kompromiss aus zwei Handlungen verringert das Risiko der Entscheidung, ähnlich wie eine Versicherungspolice für eine Absicherung im Schadensfall sorgt. Der Nutzen einer gemischten Strategie entspricht dem Erwartungsnutzen einer Lotterie, in welcher die Zustände, die durch die an der Mischung beteiligten Handlungsalternativen realisiert werden könnten, zu den Wahrscheinlichkeiten eintreten, mit denen sie in der

⁵³ Vgl. Mas-Colell/Whinston/Green (1995), S. 9.

⁵⁴ Vgl. Binmore (1994) S. 275 bzw. S 277f.

gemischten Strategie gespielt werden. Lautet eine gemischte Strategie „Spiele A zu $\frac{1}{4}$ und B zu $\frac{3}{4}$ “, dann erhält man für diese Strategie einen erwarteten Nutzen von $\frac{1}{4}$ des Nutzens von A plus $\frac{3}{4}$ des Nutzens von B.

3.2 Spieltheoretische Gleichgewichtskonzepte⁵⁵

3.2.1 Das Gefangenen-Dilemma

Das wohl bekannteste Beispiel für eine spieltheoretische Anwendung ist das Gefangenen-Dilemma. In diesem Spiel werden zwei Kriminelle vor Gericht wegen einer Straftat (z. B. einem Raubüberfall) angeklagt. Die Höhe ihrer Strafe hängt allein von ihrer Aussage ab, aber sie werden getrennt verhört und können sich nicht absprechen. Jeder der beiden Kriminellen hat nun zwei Möglichkeiten: er kann die Wahrheit sagen und die Tat gestehen, oder er kann lügen und alles abstreiten. Insgesamt können sich durch das Verhör also vier Situationen ergeben, die dann jeweils eine Auswirkung auf das Urteil haben:

- 1) Beide lügen und werden (um die Konstellation des Dilemmas zu erzeugen) wegen eines anderen Verbrechens zu einer Strafe von einem Jahr verurteilt.
- 2) Beide sagen die Wahrheit und bekommen eine milde Strafe von zwei Jahren.
- 3) Krimineller 1 sagt die Wahrheit, Krimineller 2 jedoch nicht. Krimineller 1 geht wegen einer Kronzeugenregelung straffrei aus, während Krimineller 2 die Höchststrafe von fünf Jahren bekommt.
- 4) Krimineller 2 sagt die Wahrheit und Krimineller 1 lügt. Analog zu 3) kommt jetzt Krimineller 2 frei, während Krimineller 1 für fünf Jahre ins Gefängnis geht.

Aus Sicht der Spieltheorie ist zu erwarten, dass beide Kriminelle die Wahrheit sagen und damit eine Strafe von zwei Jahren bekommen, obwohl sie in dem Fall, dass beide lügen, nur zu einem Jahr verurteilt werden würden. Das Gefangenen-Dilemma ist nur deshalb ein Dilemma, weil die meisten Menschen lieber kürzer als länger im Gefängnis sitzen. Bevor erklärt wird, wie es zu diesem Dilemma kommt (siehe Abschnitt 3.2.2), müssen aber zuerst die grundlegenden Bestandteile des Spiels genauer erläutert werden. Diese sind zunächst einmal die Spieler und die Handlungen, die sie ausführen können. Zu dem Ergebnis der Handlungen – der Auszahlung – wurde bereits in Abschnitt 2.1.3. einiges gesagt. Wie ein Spieler die möglichen Ergebnisse bei einem Spiel bewertet, das heißt sein Nutzen einer Kombination von Handlungen, hängt von seinen Präferenzen ab.

⁵⁵ Für eine ausführlichere Darstellung und weitere Referenzen siehe Gibbons (1992) oder Gintis (2000).

Für die Spieltheorie sind nur solche Handlungen von Bedeutung, die eine Auswirkung auf die Auszahlung der beteiligten Spieler haben können.⁵⁶ Im Gefangenen-Dilemma können die Spieler entweder ein lügen oder die Wahrheit sagen. Es hat für ihr Strafmaß jedoch keine Auswirkung, ob sie ihre Aussage auf Deutsch oder auf Englisch oder mit verbundenen Augen machen. Man könnte also sagen, dass es eine ganze Reihe von ähnlichen Handlungen gibt, die zu einem identischen Ablauf des Spiels führen. Es macht Sinn, diese unterschiedlichen Aktionen als eine einzige Handlung desselben Typs aufzufassen, selbst wenn es vielleicht doch minimale Nutzenveränderungen gibt (immerhin ist ein Geständnis in einer Fremdsprache wahrscheinlich ungleich anstrengender).

Wenn die Handlungen aller Spieler gleichzeitig erfolgen, kann das Spiel in der strategischen Form oder Normalform dargestellt werden (siehe Abbildung 3.1). Falls es nur zwei Spieler gibt, besteht die Normalform aus einer Matrix, in welcher die Zeilen und Spalten jeweils die Strategien der beiden Spieler enthalten. In den Zellen der Matrix stehen jeweils zwei Zahlen: die linke ist der Nutzen des Zeilenspielers, die rechte der des Spaltenspielers.

		Spieler 2	
		Wahrheit	Lüge
Spieler 1	Wahrheit	-2, -2	0, -5
	Lüge	-5, 0	-1, -1

Abb. 3.1 Das Gefangenen-Dilemma (vgl. Gibbons (1992), S. 3)

Spieler sind meistens einzelne menschliche Personen, aber im Prinzip können auch gesellschaftliche Gruppierungen wie der Staat oder sogar die künstliche Intelligenz eines Computers als Spieler aufgefasst werden. Die klassische Spieltheorie sieht Spieler als rationale Wesen, deren Entscheidungen widerspruchsfrei sind. In zwei identischen Situationen muss ein Spieler also jeweils dieselbe Strategie wählen. Außerdem müssen sich die Spieler in die Denkweise ihrer Konkurrenten hineinversetzen, um deren Handlungen voraussehen zu können.

⁵⁶ Eine relevante Handlung muss nicht zwangsläufig die Auszahlung verändern. In dem Spiel, das in Abbildung 3.5 dargestellt wird, macht die Handlung von Spieler 2 keinen Unterschied mehr bei der Auszahlung, wenn Spieler 1 mit „A“ bereits das Spiel beendet hat. Trotzdem ist es für die Entscheidung von Spieler 1 wichtig, ob Spieler 2 „C“ oder „D“ wählt.

3.2.2 Dominante Strategien

Im Gefangenendilemma ist die Entscheidung über die optimale Strategie eines Spielers unabhängig von der Strategie des anderen Spielers. Dies liegt daran, dass „Wahrheit“ für eine gegebene Handlung eines anderen Spielers immer die geringere Strafe zur Folge hat. Wenn Spieler 2 lügt, kann Spieler 1, indem er ebenfalls lügt, eine Strafe von nur einem Jahr erreichen. Sagt Spieler 1 jedoch in dieser Situation die Wahrheit, so geht er sogar straffrei aus. „Wahrheit“ ist in diesem Fall also eine bessere Strategie als „Lüge“, wenn das Ziel eine geringere Strafe ist. Falls Spieler 2 die Wahrheit sagt, erhält Spieler 1 bei einer Lüge die Höchststrafe von fünf Jahren. Nur durch eine korrekte Aussage kann er sich noch eine geringere Strafe von zwei Jahren sichern. Auch in dieser Situation ist die Strategie „Wahrheit“ also optimal für Spieler 1. Aus Symmetriegründen sollte Spieler 2 ebenfalls die Wahrheit sagen, so dass (Wahrheit, Wahrheit) bei rationalen Spielern als Ausgang des Spiels zu erwarten ist. Der Spieltheoretiker bezeichnet eine Strategie wie „Wahrheit“, die immer optimal ist, egal was der oder die anderen Spieler machen, als *dominant*.

Die Untersuchung von möglichen Strategien auf Dominanz stellt geringere Anforderungen an die Rationalität der Spieler als andere spieltheoretische Konzepte zur Bewertung von Strategien. Ein Spieler muss nämlich keine Vermutungen über die wahrscheinlichen Handlungen anderer Spieler anstellen, da diese Handlungen ja sowieso nichts daran ändern, welche Strategie für ihn optimal ist. Er kann also alle Auszahlungen außer seiner eigenen ignorieren und das Spiel wie ein simples Ein-Spieler-Entscheidungsproblem behandeln (siehe Abbildung 3.2).

		Spieler 2	
		X	Y
Spieler 1	X	2, -	3, -
	Y	4, -	5, -

Abb. 3.2 Beispiel für eine dominante Strategie

3.2.3 Common Knowledge

Die optimale Wahl der Strategie beruht auf der richtigen Einschätzung des Verhaltens der anderen Spieler. Dazu müssen die möglichen Handlungen der anderen Spieler, sowie deren Auswirkung auf die Auszahlungen bekannt sein. Weil jeder Spieler eine Entscheidung trifft, muss auch jeder Spieler all diese Informationen besitzen. Aber das ist noch nicht genug:

Damit alle Spieler von einem rationalen Verhalten ihrer Konkurrenten ausgehen können, muss jeder von ihnen wissen, dass alle Spieler diese Informationen besitzen. Außerdem muss jeder von ihnen wissen, dass alle wissen, dass die Informationen bekannt sind, usw. Man sagt: Die grundlegenden Bestandteile des Spiels (mögliche Strategien und Auszahlungen) sind *common knowledge*.

Der spieltheoretische Entscheidungsmechanismus setzt also einen gewissen Grad an Intentionalität voraus. Natürlich kann man auch allein auf Grundlage der eigenen zu erwartenden Auszahlungen handeln. Die Auswirkungen, die die Handlungen anderer Spieler auf die Auszahlungen haben, können dabei wie Zufallsereignisse behandelt und mit Wahrscheinlichkeiten belegt werden, so als würde man den Kauf eines Regenschirms vom Wetterbericht für die nächsten Tage abhängig machen. Aber ein Spieler, der die Handlungen seiner Konkurrenten *vorhersagen* kann, kann so ein Verhalten zu seinem Vorteil nutzen. Möglicherweise kommt er zu dem Schluss, dass sein Gegenüber eine Strategie vernünftigerweise nicht spielen sollte, weil sie von einer anderen Strategie dominiert wird. Wenn er dies berücksichtigt, verringert sich damit auch *für ihn* die Anzahl der möglichen Ergebnisse.

		Spieler 2	
		L	R
Spieler 1	O	3, 7	2, 6
	U	1, -500	0, 1

Abb. 3.3 a) Einschätzung des Verhaltens eines Mitspielers

Spieler 2 könnte sich im obigen Beispiel bestenfalls eine Auszahlung von 7 sichern, wenn er „L“ spielt und Spieler 1 gleichzeitig „O“ wählt. Wenn es aber aus Sicht von Spieler 2 nur eine winzige Möglichkeit gibt, dass Spieler 1 „U“ spielt, sollte er um „L“ einen weiten Bogen machen, da er sonst mit einem großen Verlust rechnen muss.⁵⁷ Handelt Spieler 1 jedoch rational, so wird er niemals „U“ spielen, da diese Strategie von „O“ dominiert wird ($3 > 1$, $2 > 0$). Wenn Spieler 2 aber weiß, dass Spieler 1 in jedem Fall „O“ wählen wird, so vereinfacht sich aus seiner Sicht das Spiel wie folgt:

⁵⁷ Ein ähnliches Spiel beschreibt Bicchieri (1998).

	Spieler 2		
		L	R
Spieler 1	O	3, 7	2, 6

Abb. 3.3 b) Einschätzung des Verhaltens eines Mitspielers

Damit steht Spieler 2 jedoch vor einem einfachen Entscheidungsproblem zwischen den beiden Alternativen „L“ mit der Auszahlung 7 und „R“ mit der Auszahlung 6, so dass der Spieler vernünftigerweise „L“ wählen sollte.

Leider ist eine solche Vereinfachung eines Spieles nur in Ausnahmefällen möglich. Im nächsten Abschnitt wird eine Möglichkeit beschrieben, um für jedes Spiel in Normalformdarstellung die optimalen Entscheidungen bestimmen zu können: das Nash-Gleichgewicht.

3.2.4 Das Nash-Gleichgewicht

Wenn beide Spieler rational entscheiden und diese Tatsache *common knowledge* ist, dann wird sich in dem Spiel ein *Nash-Gleichgewicht* ergeben. Dies ist eine Situation, in der die Strategien aller Spieler *wechselseitig beste Antworten* sind, d. h. kein Spieler kann seinen Nutzen durch Wahl einer anderen Strategie verbessern, ohne dass der andere Spieler ebenfalls seine Strategie ändert (was ohne Absprache nicht wahrscheinlich ist). Im folgenden Beispiel sind (O, L) und (U, R) Gleichgewichte in reinen Strategien:

	Spieler 2		
		L	R
Spieler 1	O	5, 4	0, 1
	U	2, 2	5, 3

Abb. 3.4 Spiel mit zwei Nash-Gleichgewichten in reinen Strategien

Um zu sehen, dass (O, L) ein Gleichgewicht ist, muss man sich nur ansehen, was passiert, wenn einer der beiden Spieler abweicht: Spieler 1 bekäme durch „U“ nur einen Nutzen von 2 – im Vergleich zu 5, wenn er bei „O“ bleibt. Spieler 2 würde sich durch einen Wechsel von „L“ auf „R“ von 4 auf 1 verschlechtern und deshalb auch nicht seine Strategie ändern. Die Begründung des Gleichgewichtes (U, R) verläuft analog.

Im Gefangenen-Dilemma⁵⁸ ist (Wahrheit, Wahrheit) das einzige Nash-Gleichgewicht, denn würde ein Spieler von seiner Strategie abweichen und lügen, dann bekäme er wegen des noch bestehenden Geständnisses seines Komplizen die Höchststrafe. Manche Spiele besitzen aber auch Gleichgewichte in *gemischten Strategien*. In Abschnitt 3.1 wurde gesagt, dass eine Mischung von Strategien sinnvoll sein kann, um das Risiko in einer Entscheidungssituation zu verringern. In der Spieltheorie besteht dieses Risiko in der Ungewissheit über die Entscheidung der anderen Spieler, so dass es nahe liegt, in der optimalen Strategie mehrere Handlungen mit gegenläufigem Ergebnis zu verbinden.

Im obigen Beispiel lautet das Gleichgewicht in gemischten Strategien: Spieler 1 spielt zu $1/4$ „O“ und zu $3/4$ „U“, während Spieler 2 zu $5/8$ „L“ und zu $3/8$ „R“ spielt. Bei gemischten Gleichgewichten ist die Begründung, warum kein Spieler von der Gleichgewichtsstrategie abweichen will, etwas trickreicher. Wenn Spieler 1 die besagte gemischte Strategie spielt, ist Spieler 2 indifferent zwischen seinen beiden Strategien, weil er sowohl für „L“ als auch für „R“ einen erwarteten Nutzen von 2,5 bekommt. *Jede* der Strategien von Spieler 2, auch andere Mischungen, führt in diesem Fall zum selben erwarteten Nutzen, also ist es aus seiner Sicht eigentlich gleichgültig, welche Strategie er spielt. Aber ein Gleichgewicht stellt sich nur dann ein, wenn Spieler 1 *ebenfalls* indifferent zwischen seinen möglichen Strategien ist, und aus diesem Grunde ergibt sich auch nur eine einzige Gleichgewichtsstrategie für Spieler 2.

3.2.5 Dynamische Spiele

Die wenigsten Handlungen passieren gleichzeitig oder so zeitnah, dass keine Reaktion auf die Aktionen anderer Personen mehr möglich ist. Daher soll nun auf die Modellierung der so genannten „dynamischen Spiele“ eingegangen werden, die diesem Fakt Rechnung tragen.

Wenn die Spieler ihre Handlungen nacheinander ausführen, wird das Spiel in der Extensivform dargestellt (siehe Abbildung 3.5). Die Extensivform ist ein Baum, der überall dort Knotenpunkte besitzt, wo ein Spieler eine Entscheidung trifft. Jede Handlung, die bei dieser Entscheidung zur Wahl steht, wird durch einen Ast repräsentiert. Der Nutzen aller Spieler wird am Ende jedes Astes aufgelistet.

Der erste Spieler ist bei Spielen dieses Typs im Vorteil, da er seinen Konkurrenten mit seiner Handlung vor vollendete Tatsachen stellt. Im Beispiel kann Spieler 1 durch seine Entscheidung das Spiel vorzeitig beenden, indem er „A“ spielt. Nur durch die Handlung „B“ erhält Spieler 2 überhaupt die Gelegenheit, in das Spiel einzugreifen. Spieler 2 hat also keine

⁵⁸ Siehe oben, Abschnitt 3.2.1.

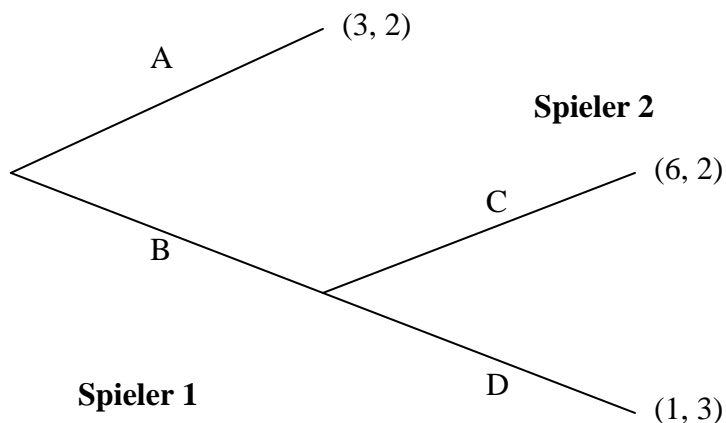


Abb. 3.5 Beispiel für ein dynamisches Spiel

Möglichkeit mehr, das Verhalten von Spieler 1 zu beeinflussen, und kann nur noch das Beste aus der Situation machen. Er wird die Entscheidung zwischen „C“ und „D“ davon abhängig machen, welche der beiden Handlungen ihm den größtmöglichen Nutzen verspricht. Als rationaler Spieler kann Spieler 1 diese Entscheidung voraussehen. Er wird also seine Strategie auf die optimale Handlung für Spieler 2 abstimmen. Im obigen Beispiel sollte Spieler 2 „D“ „C“ vorziehen ($3 > 2$), so dass Spieler 1 bei einer Wahl von „B“ mit der Handlung „D“ rechnen muss. In diesem Fall erhält er für „A“ einen Nutzen von 3, für „B“ aber nur einen Nutzen von 1, so dass „A“ für ihn optimal ist. Diese Vorgehensweise zur Bestimmung der optimalen Strategie nennt man *Rückwärtsinduktion*, weil die Entscheidungen des Spiels in umgekehrter zeitlicher Reihenfolge analysiert werden. Das Gleichgewicht dieses Spiels lautet also: Spieler 1 spielt „A“, Spieler 2 spielt „D“ und das Spiel endet nach dem ersten Zug.

Es mag etwas merkwürdig scheinen, dass die Handlung von Spieler 2 für das Gleichgewicht noch eine Rolle spielt. Jedoch ist die Überzeugung von Spieler 1, dass Spieler 2 „D“ spielen wird, der ausschlaggebende Faktor für seine Entscheidung, das Spiel mit der Handlung „A“ vorzeitig zu beenden. Wäre er unsicher, wie sein Konkurrent entscheiden würde, dann würde er möglicherweise „B“ wählen, in der Hoffnung sich eine Auszahlung von 6 sichern zu können.

In dynamischen Spielen zeigt sich der Vorteil einer strategischen Denkweise, der auch für die Ethik von Bedeutung sein könnte. In diesen Situationen kann die optimale Handlung nur bestimmt werden, wenn später erfolgende Entscheidungen richtig vorhergesagt werden. Dazu muss man sich aber in die Lage der anderen beteiligten Personen hineinversetzen können, um ihre Gedanken nachzuvollziehen. Wie in Abschnitt 4.1 erläutert werden soll, muss ein

Utilitarist unter Umständen sogar die moralische Denkweise eines Anhängers einer kantischen Ethik verstehen können, wenn er moralisch richtig handeln will.

3.2.6 Spiele mit unvollständigen Informationen

Bisher wurde immer davon ausgegangen, dass den Spielern bekannt ist, welchen Nutzen welche Handlungen bringen, oder welche Strategien den Konkurrenten zur Verfügung stehen. Das muss aber natürlich nicht so sein, sondern dürfte im Alltag wohl nur in Ausnahmefällen zutreffen. Die eigenen Informationen kann man vielleicht noch ganz gut in Erfahrung bringen, indem man sich die möglichen Folgen einer Handlung überlegt und deren Ausmaß einschätzt. Für eine strategische Überlegung muss man aber auch eine gewisse Vorstellung davon haben, was *andere* Spieler wissen und was nicht.

Unvollständige Informationen bedeuten Unsicherheit über die Folgen der eigenen Handlung und werden in der Spieltheorie deshalb wie Entscheidungen unter Risiko behandelt. Es wird dabei angenommen, dass die Spieler nicht vollkommen ahnungslos sind, sondern über mögliche Ausprägungen eines unbekanntem Parameters und deren Wahrscheinlichkeitsverteilungen Bescheid wissen. Sollte es für das Spiel von Bedeutung sein, wissen sie zum Beispiel, dass ein ansonsten anonym Mitspieler entweder männlich oder weiblich ist und dass eines von beiden mit einer Wahrscheinlichkeit von etwa $\frac{1}{2}$ zutrifft. Sie finden aber frühestens nach ihrer Entscheidung durch den Ausgang des Spieles heraus, ob sie es mit einem Mann oder einer Frau zutun hatten.

Im Bayesianischen Gleichgewichtskonzept wird die fehlende Information durch einen „Naturzug“ modelliert, der vor allen anderen Handlungen passiert. Es wird so getan, als lege eine höhere Macht als weiterer Spieler zu Beginn alle Parameter fest, über die bei den Spielern Unklarheit besteht, indem diese Macht jeweils eine mögliche Ausprägung eines Parameters auswählt. Wenn eine Entscheidung also beispielsweise wieder davon abhängt, ob ein anderer Spieler männlich oder weiblich ist, wird zunächst per Zufall der *Typ* des Spielers bestimmt, bevor das eigentliche Spiel abläuft. Jeder Spieler kennt in der Regel die ganze Zeit über seinen eigenen Typ, die Informationen über die anderen Spieler werden jedoch natürlich erst *nach* der Entscheidung über die Handlungen aufgedeckt.

3.2.7 Probleme der Spieltheorie im Alltag

Leider stellt das Bayesianische Gleichgewichtskonzept sehr hohe Anforderungen an die Vernunft der Spieler, weil sie nicht nur selbst die Häufigkeit einschätzen müssen, mit der die

verschiedenen Typen von Situationen eintreten können, sondern außerdem eine Vorstellung darüber entwickeln müssen, wie *andere* Spieler diese Wahrscheinlichkeiten einschätzen.

Auch gemischte Strategien haben in der Praxis keinen besonders guten Ruf, da sie nur schwer mit dem gesunden Menschenverstand in Einklang gebracht werden können. Man stelle sich einen Mann vor, der vor jeder seiner Handlungen erst eine Münze wirft. Er wird mit diesem Verhalten bei seinen Mitmenschen bestimmt nicht den Eindruck erwecken, besonders vernünftig zu sein. Sie werden eher denken, er sei abergläubisch, weil er seine Entscheidungen dem „Schicksal“ überlässt. Dazu kommt noch, dass viele alltägliche Entscheidungen so schnell gehen müssen, dass keine Zeit mehr bleibt, um einen Zufallsmechanismus zu Rate zu ziehen. Spontane Handlungen in Alltagsproblemen sind aber ein generelles Problem einer mathematischen Entscheidungstheorie. Der Aufwand eines komplizierten Entscheidungsmechanismus lässt sich nur dann vertreten, wenn die Tragweite der Entscheidung genügend groß ist.

Viele alltägliche Handlungen geschehen darüber hinaus aus Gewohnheit oder aus einer Laune heraus. Man kann deshalb nicht mehr von Rationalität sprechen, da dem Entscheidungsträger doch wohl bewusst sein sollte, wie die Entscheidung für eine bestimmte Handlung zustande gekommen ist. Die herkömmliche Spieltheorie versagt an dieser Stelle, weil sie nur das Verhalten *rationaler* Spieler beschreiben kann.

Mit einem kleinen Kunstgriff kann dieses Problem jedoch gelöst werden: Wenn die Spieler nicht über Rationalität verfügen, um Entscheidungen zu treffen, dann müssen ihre Strategien eben von außerhalb festgelegt werden, z. B. indem sie per Zufall bestimmt werden. So als wäre es ein Spiel mit unvollständigen Informationen wird ein „Naturzug“ eingefügt, der vor allen anderen Handlungen stattfindet. Die „Natur“, also sozusagen eine höhere Macht, tritt hypothetisch gesprochen als weiterer Spieler auf und legt die Strategien der normalen Spieler fest. Der Charakter des Spiels ändert sich dadurch ein wenig, da nun Spieler und Strategie zusammenfallen. Ein Spieler wird gewissermaßen dazu „geboren“, eine bestimmte Strategie zu spielen. Natürlich können die Spieler nun nicht mehr ihren Nutzen maximieren, so dass ein anderer Weg gefunden werden muss, um Strategien miteinander zu vergleichen. Ein mögliches Bewertungskriterium – den Prozentsatz der Spieler, die eine bestimmte Strategie noch spielen, wenn ein Spiel über mehrere Generationen betrachtet wird – greift die Evolutionäre Spieltheorie auf.

3.3 Die Evolutionäre Spieltheorie

Ursprünglich von John Maynard Smith⁵⁹ entwickelt, ist die Evolutionäre Spieltheorie wohl besonders für eine deskriptive Ethik geeignet, da sie Menschen nicht mehr als aktive Entscheider in moralischen Fragestellungen ansieht, sondern nur noch als Spielfiguren miteinander im Wettstreit stehender moralischer Vorstellungen. In der Biologie konkurriert die Evolutionstheorie durch ihre Beschreibung der Entstehung der Arten mit religiösen Schöpfungstheorien, in der Soziologie könnte eine evolutionstheoretische Beschreibung der Entstehung moralischer Normen eine Alternative zu herkömmlichen ethischen Ansätzen bieten. In beiden Fällen ist ein *Replikator* von wesentlicher Bedeutung. Dieser Träger der Erbmasse – Gen in der darwinistischen Evolution, Mem in der kulturellen Evolution á la Richard Dawkins⁶⁰ – sichert sein Fortbestehen, indem er von Generation zu Generation nahezu unverändert weitergegeben wird.

3.3.1 Das Taube-Falke-Spiel

In einer typischen Spielsituation wird der Revierkampf zweier Individuen simuliert, deren Verhalten jeweils von einem Replikator vorgegeben wird. Das Taube-Falke-Spiel⁶¹ ist wohl das bekannteste Beispiel für die Modellierung eines solchen Konkurrenzkampfes zweier Gene durch die Evolutionäre Spieltheorie. In diesem Beispiel teilen sich Falken und Tauben ein Gebiet mit einem begrenzten Nahrungsangebot V . Zur Vereinfachung werde angenommen, dass Falken und Tauben eigentlich identische Vögel sind, die sich nur in ihrem Verhalten unterscheiden: Tauben sind friedlich, Falken aggressiv. Bei einem direkten Aufeinandertreffen zweier Vögel gibt es drei Möglichkeiten:

1. Zwei Tauben teilen die Nahrung friedlich (Auszahlung: $V/2$).
2. Ein Falke verjagt eine Taube und kann die ganze Nahrung für sich beanspruchen, während die Taube leer ausgeht (Auszahlung: V).
3. Zwei Falken kämpfen miteinander. Der Sieger erhält V , der Verlierer erleidet Verletzungen in Höhe von $-C$. Da beide Tiere identisch sind, beträgt die Wahrscheinlichkeit eines Sieges 50%, so dass die erwartete Auszahlung wie folgt aussieht: $V/2 - C/2 = \frac{1}{2}(V - C)$.

⁵⁹ Siehe z. B. Maynard Smith (1982).

⁶⁰ Vgl. Dawkins (1976).

⁶¹ Vgl. Maynard Smith (1982), S. 11ff.

	Spieler 2	
	Taubе	Falke
Spieler 1	Taubе	V/2, V/2
	Falke	V, 0
		$\frac{1}{2}(V - C), \frac{1}{2}(V - C)$

Abb. 3.6 Das Taube-Falke-Spiel (vgl. Maynard Smith (1982), S. 12)

Ob sich Tauben oder Falken über mehrere Generationen behaupten können, hängt von den genauen Werten für V und C ab. Tauben haben nur dann eine Chance auf ein dauerhaftes Überleben, wenn die Kosten eines Kampfes der Falken groß sind, also wenn $C > V$. Andernfalls ist „Falke“ eine dominante Strategie, also die optimale Wahl eines rationalen Spielers. Ein Wechsel der Strategie ist in diesem Spiel jedoch nicht möglich. Eine Taube ist ihr ganzes Leben lang eine Taube, auch wenn sie als Falke besser gestellt wäre. Erst in der nächsten Generation macht sich eine Veränderung bemerkbar, denn es konnten sich nur die Tauben fortpflanzen, die nicht von Falken verjagt worden sind. Im Fall $V > C$ geben aber nicht nur die Falken ihre Gene weiter, die Tauben vertreiben konnten, sondern auch diejenigen, die sich gegen andere Falken durchgesetzt haben. Es wird also mehr neue Falken als neue Tauben geben, so dass nach mehreren Generationen fast nur noch Falken vorhanden sein werden.

3.3.2 Nutzen und fitness

Der Nutzen einer Auszahlung besteht in der Evolutionären Spieltheorie in dem Beitrag dieser Auszahlung zur *fitness* einer Strategie, die als die *erwartete Zahl der Nachkommen pro Generation* definiert wird. Diese Definition von *fitness* ist nicht unproblematisch, da dadurch für die Bewertung von Strategien bereits Informationen darüber vorhanden sein müssen, wie diese Strategien im direkten Vergleich miteinander abschneiden. Ähnlich wie beim Präferenzutilitarismus⁶² besteht also die Gefahr, dass die Aussagen der Evolutionären Spieltheorie keine neuen Erkenntnisse bringen, weil ihre Schlussfolgerungen mit den Prämissen übereinstimmen.

Dieser Tautologie-Vorwurf ist jedoch ein generelles Problem der Evolutionstheorie, und wird unter Anderem von Elliott Sober durch den Einwand widerlegt, dass mathematische Wahrheiten, obwohl sie a priori gültige Sätze sind, trotzdem unerwartete Ergebnisse

⁶² Siehe oben, Abschnitt 2.1.2.

beschreiben können.⁶³ Beispielsweise folgt die Tatsache, dass im Taube-Falke-Spiel unter Umständen Tauben und Falken gemeinsam in einer Population existieren können, allein aus den Annahmen über die Auszahlungen in den jeweiligen Situationen. Ohne eine spieltheoretische Analyse hätte man dieses Ergebnis aber wahrscheinlich nicht erwartet. Und so kann man, wie Ken Binmore bemerkt,⁶⁴ selbst in diesem einfachen Spiel etwas Neues lernen, obwohl man sich auf eine tautologische Aussage stützt.

3.3.3 ESS und Replikatorodynamik

Solange sich die Evolutionäre Spieltheorie auf deskriptive Aussagen beschränkt und beschreibt, wie sich bereits bekannte Verhaltensweisen unter gegebenen Umständen über einen längeren Zeitraum hinweg in der Gesellschaft entwickeln, gibt es keine Schwierigkeiten mit der Bestimmung der *fitness* der Strategien, weil ja alle nötigen Informationen schon vorhanden sind. Wie soll man aber neue Verhaltensweisen bewerten, die gerade erst durch Mutation entstanden sind? Bevor diese Frage beantwortet wird, soll zunächst dargestellt werden, wie die Evolutionäre Spieltheorie die Prozesse der Mutation auf der einen Seite und der Selektion auf der anderen Seite simulieren kann.

Jörgen Weibull ordnet dem Mutationsmechanismus das Konzept der *evolutionär stabilen Strategie* (ESS) zu, während die Selektion von Strategien seiner Ansicht nach von dem Konzept der *Replikatorodynamik* übernommen wird.⁶⁵ Evolutionäre Stabilität vermittelt eine gewisse Resistenz gegenüber Neueindringlingen in eine Bevölkerung und wird durch die folgende Definition mathematisch umschrieben, wobei u den Nutzen bzw. die *fitness* repräsentiert:⁶⁶

Defition ESS: Eine Strategie x aus der Menge aller verfügbaren Strategien ist *evolutionär stabil*, wenn für jede Strategie $y \neq x$ eine Zahl $\bar{\varepsilon}_y \in (0,1)$ existiert, so dass für alle $\varepsilon \in (0, \bar{\varepsilon}_y)$ gilt:

$$u[x, \varepsilon y + (1 - \varepsilon)x] > u[y, \varepsilon y + (1 - \varepsilon)x]$$

Die Strategie y stellt einen Mutanten dar, der in eine bestehende Population eindringt und die Strategie x herausfordert. Damit sich x behaupten kann, muss ihre *fitness* größer sein als die des Mutanten, also die obige Ungleichung für einen gewissen Schwellenwert $\bar{\varepsilon}_y \in (0,1)$

⁶³ Vgl. Sober (2000), S. 74.

⁶⁴ Vgl. Binmore (1994), S. 100.

⁶⁵ Vgl. Weibull (1995), S. 69

⁶⁶ Ebd. S. 36.

erfüllt sein. Das bedeutet, dass in der Generation, in welcher die Mutation geschieht, nur ein (normalerweise kleiner) Teil der Gesellschaft tatsächlich die neue Verhaltensweise übernimmt. Die Ungleichung selbst vergleicht die beiden Strategien x und y hinsichtlich ihrer *fitness* in einer Population, die zu einem Anteil von ε die neue Verhaltensweise befolgt und zum übrigen Teil bei dem bisherigen Verhalten bleibt.

Die Entwicklung der Anteile bestimmter Strategien oder Gene an einer Gesamtpopulation lässt sich am besten über Wachstumsraten ausdrücken. Man betrachtet also die Veränderung von einer Generation zur nächsten und versucht, einen langfristigen Trend auszumachen. Die *Replikatorodynamik*,⁶⁷ ein System gewöhnlicher Differentialgleichungen, befolgt genau diese Vorgehensweise. Dieses Gleichungssystem beschreibt aber nur das Wachstum bereits bestehender Strategien; die Neubildung von Mutanten wird zunächst nicht berücksichtigt. Wenn eine Strategie dominant ist, so wie der Falke im obigen Beispiel, wächst der Bevölkerungsanteil dieser Strategie immer weiter an. Irgendwann besteht die Bevölkerung nur noch aus Falken, so dass ein Punkt des maximalen Wachstums erreicht wird. Danach befindet sich die Bevölkerung im Gleichgewicht, d. h. der Bevölkerungsanteil der Strategien ändert sich nicht mehr. Die Stabilität dieses Zustandes besteht jedoch nur so lang, bis ein Mutant in die Population eindringt. Danach verändern sich die Bevölkerungsanteile aller Strategien wieder, bis ein neues Gleichgewicht erreicht wird. Üblicherweise werden daher alle denkbaren Strategien, also auch alle zu erwartenden Mutanten, in die Replikatorodynamik aufgenommen, damit das entstehende Gleichgewicht nur aus Strategien gebildet wird, die *evolutionär stabil* im Sinne der oben stehenden Definition sind. Dass manche Mutanten „erwartet“ werden können, liegt daran, dass es sich dabei um Strategien handelt, deren Verhalten eine Kombination von bereits bekannten Strategien darstellt. Damit ist dieses Mal keine gemischte Strategie gemeint, sondern eine gewisse Steuerung des Verhaltens, die es dem Individuum, das die Strategie ausführt, erlaubt, auf unterschiedliche Konkurrenten mit verschiedenen Verhaltensweisen zu reagieren. So könnte eine derartige Strategie lauten: „Verhalte dich wie eine Taube, wenn du auf eine Taube triffst, und wie ein Falke, wenn du auf einen Falken triffst.“⁶⁸ Es lässt sich leicht bestimmen, wie diese Strategie gegen die bisherigen Strategien abschneiden wird. Leider gibt es aber bis jetzt offenbar keine Möglichkeit, um die *fitness* völlig neuartiger Mutanten einzuschätzen, außer vielleicht ein Spiel mit unvollständigen Informationen, in welchem allerdings wenigstens die Wahrscheinlichkeitsverteilung der unbekanntenen *fitness* des Mutanten bekannt sein muss.

⁶⁷ Vgl. ebd. S. 69 ff.

⁶⁸ Vgl. beispielsweise Binmore (1994), S. 196f.

Ein weiteres mögliches Problem besteht darin, dass es nicht in allen Spielen eine evolutionär stabile Strategie gibt, so dass in manchen Situationen kein Verhalten aus evolutionärer Sicht ausgezeichnet ist. Im „Stein-Schere-Papier“-Spiel⁶⁹ ist beispielsweise jede einzelne Strategie jeweils einer weiteren Strategie überlegen und einer dritten unterlegen: „Stein“ schlägt „Schere“, „Schere“ schlägt „Papier“ und „Papier“ schlägt wiederum „Stein“. Wenn man außerdem annimmt, dass sich beide Spieler bei Wahl derselben Strategie auf ein Unentschieden einigen, ergibt sich ein Spiel ähnlich dem in Abbildung 3.7.

		Spieler 2		
		Stein	Schere	Papier
Spieler 1	Stein	0, 0	1, -1	-1, 1
	Schere	-1, 1	0, 0	1, -1
	Papier	1, -1	-1, 1	0, 0

Abb. 3.7 Stein-Schere-Papier (vgl. Maynard Smith (1982), S. 19)

Es lässt sich zeigen, dass das einzige Nash-Gleichgewicht in diesem Spiel (ein gemischtes Gleichgewicht, in welchem jede reine Strategie zu $1/3$ gespielt wird), *nicht* evolutionär stabil ist. Damit wird also keine der drei reinen Strategien dauerhaft eine Mehrheit in einer Population behalten können. Wenn z. B. ein großer Teil der Bevölkerung „Stein“ spielt, werden „Scheren“ dadurch zurückgedrängt. Gleichzeitig werden „Papiere“ aber gestärkt, weil sie seltener auf die ihnen überlegene Strategie „Schere“ treffen. Mit der Zeit wird also ein immer größerer Teil der Bevölkerung „Papier“ spielen, bis „Stein“ seine Vormachtstellung eingebüßt hat. So entsteht ein Kreislauf, bei dem jede Strategie im Wechsel die Oberhand gewinnt. Wie in Abschnitt 5.2.2 weiter ausgeführt werden soll, kann diese Situation grundsätzlich auch auf moralischer Ebene entstehen, falls sich die Stärken verschiedener ethische Ansätze auf genau den Gebieten zeigen, auf denen ihre Konkurrenten Schwächen besitzen.

3.4 Die Bedeutung von strategischem Denken für die Entscheidungsfindung

Es gibt viele Methoden, um bei einem ethischen Problem herauszufinden, welche Entscheidung moralisch richtig ist. Die Spieltheorie in Verbindung mit der Bayesianischen Entscheidungstheorie ist nur eine Möglichkeit. Die Tatsache, dass die Spieltheorie die

⁶⁹ Vgl. Maynard Smith (1982), S. 19f.

Berücksichtigung strategischer Überlegungen zulässt, spricht jedoch dafür, die spieltheoretische Art der Entscheidungsfindung vorzuziehen.

Wie bereits in Abschnitt 3.1 angedeutet wurde, kann so gut wie jede Rangordnung von Handlungen in eine Nutzenfunktion übersetzt werden. Damit ist aber jede Entscheidungstheorie, deren Aussagen auf einer derartigen Rangordnung basieren, gleichwertig zu der Bayesianischen Entscheidungstheorie und auch zu der Spieltheorie, die dasselbe Nutzenkonzept verwendet.

Nur weil die Bayesianische Entscheidungstheorie und die Spieltheorie sich als mathematische Theorien zur Beschreibung von Entscheidungssituationen bewährt haben, heißt das aber noch nicht, dass es nicht auch andere Arten von Entscheidungstheorien geben könnte. Man könnte eine Handlung in einem ethischen Kontext beispielsweise nach der Anzahl der durch sie verletzten moralischen Normen bewerten.⁷⁰ Es sollte dann aus ethischen Gesichtspunkten diejenige Handlung gewählt werden, welche die wenigsten Normen verletzt. Dieser Entscheidungsmechanismus verlangt von seinem Anwender nicht, sich in andere beteiligte Personen hineinzusetzen, um die Bewertung einer Handlung vornehmen zu können. Deshalb kann ein Computer damit programmiert werden, um moralisches Verhalten zu simulieren. Aber auch dieses Entscheidungskriterium erstellt eine Rangfolge von Handlungen, aus der dann wieder eine Nutzenfunktion abgeleitet werden kann.

Selbst für einen Anhänger der kantischen Ethik, der seine Handlungen in „moralisch richtige“ und „moralisch falsche“ unterteilt, kann eine Nutzenfunktion angegeben werden. Wenn er moralisch richtige Handlungen moralisch falschen vorzieht, dann könnten erstere mit dem Wert „1“ und letztere mit dem Wert „0“ belegt werden. Leider ist damit nicht viel gewonnen, da der eigentliche Vorteil der Nutzenfunktion gegenüber einer einfachen Präferenzordnung in ihrer Stetigkeit liegt, so dass eine gemischte Strategie auch einen Nutzen sichert, der zwischen dem Nutzen der beiden kombinierten Strategien liegt. Der Kantianer würde aber auch eine gemischte Strategie entweder als „moralisch richtig“ oder als „moralisch falsch“ einstufen, so dass sich immer noch entweder ein Nutzen von 1 oder von 0 ergibt. Damit wird er sich wahrscheinlich nicht auf derartige Kompromisse aus mehreren möglichen Handlungen einlassen, aber immerhin kann sein Verhalten durch die Spieltheorie beschrieben werden und sei es, indem seine Entscheidungen wie in der Bayesianischen Entscheidungstheorie als Zufallsereignisse behandelt werden.

Im Folgenden soll gezeigt werden, dass die Spieltheorie der Bayesianischen Entscheidungstheorie aber doch in einigen Bereichen überlegen ist. Die Bayesianische

⁷⁰ Vgl. Boella/van der Torre (2003).

Entscheidungstheorie vergleicht unterschiedliche Handlungsalternativen hinsichtlich des erwarteten Nutzens. Diese Theorie geht dabei nur von einem einzigen aktiven Entscheider aus, nämlich ihrem Anwender. Alle anderen Umstände, die sich auf den Ausgang einer Entscheidungssituation auswirken können, insbesondere die Handlungen anderer Personen, werden wie Zufallsereignisse behandelt und mit Wahrscheinlichkeiten belegt. Auf diese Weise sichert diese Entscheidungstheorie, dass alles, was für eine Entscheidung relevant ist, auch modelliert werden kann.

Es entstehen jedoch Probleme, wenn *zu viele* Einflussfaktoren in einer bestimmten Situation von Bedeutung sind. Wenn jeder dieser Umstände auch nur zwei mögliche Folgen hat, die zu einem unterschiedlichen Ergebnis führen, so kann bereits eine Situation, in der vier verschiedene Faktoren relevant sind, auf sechzehn (zwei hoch vier) unterschiedliche Weisen enden. Möglicherweise tritt jedes dieser Ergebnisse dann aber nur mit einer geringen Wahrscheinlichkeit ein, so dass keine weitere Hilfe bei der Entscheidung gegeben werden kann. Wenn eine Entscheidung allerdings von den Handlungen anderer Menschen abhängt, denen man vermutlich einen gewissen Grad an Vernunft unterstellen kann, dann ist die Spieltheorie in der Lage, die Anzahl der zu bedenkenden Folgen einer Handlung zu verringern, indem sie unvernünftige Handlungen anderer Menschen ausschließt.⁷¹

Außerdem gibt es Situationen, in denen es notwendig ist, die eigene Handlung mit der einer anderen Person abzustimmen. Schon wenn sich zwei Personen zu einem Treffen verabredet haben, aber nicht mehr wissen, an welchem Ort die Verabredung stattfinden sollte, stößt die Bayesianische Entscheidungstheorie an ihre Grenzen.

		Spieler 2	
		A	B
Spieler 1	A	1, 1	0, 0
	B	0, 0	1, 1

Abb. 3.8 Koordinationsproblem

Diese Situation, die in Abbildung 3.8 als ein Koordinationsproblem dargestellt wird, ist auch von der Spieltheorie nicht direkt auflösbar, weil die zwei Nash-Gleichgewichte (A, A) und (B, B) für beide Spieler zu identischen Nutzenwerten führen. Je mehr derartige Gleichgewichte ein Spiel besitzt, umso unwahrscheinlicher ist es, dass man, ohne eine Vorstellung über das Verhalten der anderen Spieler zu haben, zu einer richtigen Entscheidung

⁷¹ Siehe Abschnitt 3.2.3.

gelangt. Die Spieltheorie lässt aber in wiederholten Spielen eine Festlegung auf eines dieser Gleichgewichte zu.⁷² Die Berücksichtigung der Entscheidungen aus früheren vergleichbaren Situationen, erlaubt eine Einschätzung des Verhaltens der anderen Spieler, die in der Bayesianischen Entscheidungstheorie nicht möglich ist. Die Spieltheorie führt diesen Ansatz jedoch sogar noch weiter, indem sie eine Erklärung für die Entstehung von Absprachen zulässt und damit den Grundstein für die Etablierung eines gesellschaftlichen Vertrages setzt.

⁷² Vgl. Gibbons (1992), S. 82ff.

4. Die Vollständigkeit der spieltheoretischen Darstellung

Ein Werkzeug muss in der Lage sein, die ihm zugeordnete Aufgabe zu erfüllen, sonst wird es nicht zum Einsatz kommen. Je vielseitiger jedoch ein Werkzeug ist, umso häufiger wird es gebraucht werden. Genauso sieht es auch bei einer Hilfswissenschaft aus: wenn sie nur unter sehr restriktiven Annahmen verwendet werden kann, wird ihr nur wenig Beachtung geschenkt; wenn sie hingegen mit neuen Aufgaben wächst und ihr Spektrum erweitert, dann kann sie sogar für Vermittlung zwischen eigentlich gegenläufigen Positionen sorgen.

Die Spieltheorie wird wegen der Verwendung des Nutzenkonzepts in erster Linie mit dem Utilitarismus in Verbindung gebracht. Da die Spieltheorie in den Wirtschaftswissenschaften, die ebenfalls stark unter utilitaristischem Einfluss stehen, immer wieder durch neue Anwendungsprobleme gefordert wird, dürfte außer Frage stehen, dass eine utilitaristische Ethik ausreichend durch spieltheoretische Modelle beschrieben werden kann. Tatsächlich liegt der einzige wesentliche Unterschied zwischen der Ökonomik und dem Utilitarismus in dem Zweck, der mit einer Verhaltensvorschrift erreicht werden soll: ein Wirtschaftler maximiert in der Regel nur seinen individuellen Nutzen, während ein Utilitarist das Wohlergehen der Gesellschaft, also den Gesamtnutzen aller Menschen optimieren will. Aus Sicht der Spieltheorie unterscheiden sich Utilitarist und Wirtschaftler damit aber nur in ihrer persönlichen Nutzenfunktion, in die beim Utilitaristen zusätzlich die Nutzenwerte aller Mitglieder der Gesellschaft eingehen. Die Bestimmung der optimalen Handlung verläuft in beiden Fällen jedoch analog.

Wie bereits gesagt, ist der Nutzen aber nur eine mathematische Größe, welche die Präferenzen von Personen in Zahlen ausdrückt. Allerdings haben *alle* Menschen Gründe, aus denen sie gewisse Handlungen anderen vorziehen, so dass die Frage aufkommt, warum man nicht auch nicht-utilitaristische Präferenzen in Nutzenwerten „messen“ soll. Wenn man dann aber bereits für jeden Menschen eine Nutzenfunktion bestimmt hat, ist es doch eigentlich auch nicht verwerflich, wenn man ihn seinen „Nutzen“ maximieren lässt, da sein Verhalten doch vernünftigerweise seinen Vorlieben entsprechen sollte. Dieser Gedankengang führt zu folgender These, die in diesem Abschnitt diskutiert werden soll:

These 2: Jeder ethische Ansatz kann in der Spieltheorie modelliert werden.

Ein Beweis dieser These dürfte schwierig, wenn nicht sogar unmöglich sein. Das liegt nicht unbedingt daran, dass die Spieltheorie dies nicht leisten kann, sondern daran, dass eine Entwicklung fundamental neuartiger Ethiken nicht grundsätzlich ausgeschlossen werden

kann. Man kann jedoch nicht vorhersehen, wie eine solche Ethik moralische Urteile begründet, weshalb immer die Möglichkeit besteht, dass sich eine moralische Entscheidungssituation konstruieren lässt, die nicht von der Spieltheorie modelliert werden kann, obwohl eine Ethik eine zufrieden stellende Entscheidungshilfe bereitstellt.

Die Behauptung der Vollständigkeit der Spieltheorie bei der Darstellung moralischer Fragen kann natürlich widerlegt werden, wenn sich ein entsprechendes Gegenbeispiel finden lässt. Eine Möglichkeit, dies zu erreichen, wäre, ein Entscheidungskriterium zwischen zwei Handlungen anzugeben, das sich nicht in unterschiedlichen Folgen der Handlungen niederschlägt. Wenn man jedoch davon ausgeht, dass nur das wirklich ist, was auch Wirkungen in der Welt hat, dann scheint es unmöglich, zwei unterschiedliche Handlungen in einer Situation anzugeben, die sich *nicht* in beobachtbaren und damit messbaren Auswirkungen unterscheiden.

Man könnte allerdings auch etwas anders argumentieren: Zugegeben, jeder Umstand in einer Entscheidungssituation, der beobachtet werden kann, kann auch in einem Spiel modelliert werden. Trotzdem zeigen manche Menschen in experimentellen Situationen ein Verhalten, das von spieltheoretischen Konzepten nicht vorhergesagt wurde, wie zum Beispiel das Spielen einer dominierten Strategie im Gefangenendilemma. Eine Kritik, die auf experimentellen Daten beruht, trifft natürlich in erster Linie die Wirtschaftswissenschaften, deren Hypothesen hier unter Verwendung der Spieltheorie überprüft werden. Aber nicht selten ergeben sich nur deshalb Ungereimtheiten, weil die Methoden der Spieltheorie noch nicht ausgereift genug sind, um jede Situation hundertprozentig korrekt darzustellen. Für die Spieltheorie besteht also immer der Ansporn, durch Modellerweiterungen widersprüchliche Ergebnisse in den Griff zu bekommen.⁷³ In Abschnitt 4.2 sollen daher einige Möglichkeiten aufgezeigt werden, wie die Spieltheorie mit nicht-utilitaristischen Moralvorstellungen umgehen könnte. Zuvor jedoch soll an einem Beispiel geschildert werden, wie ein Utilitarist durch die Verwendung der Spieltheorie auf das Verhalten eines nicht-utilitaristischen Entscheiders eingehen kann.

⁷³ Aus wissenschaftstheoretischer Sicht könnte hier die Frage aufkommen, ob diese Modellerweiterungen nicht vielleicht nur die Widerlegung der Spieltheorie hinauszögern, weil sie „ad hoc“ eine Antwort auf Kritik entwickeln, ohne die fundamentalen Probleme der Theorie anzugehen. Allerdings kann man dieselben Vorwürfe an die Wirtschaftswissenschaften richten, da die experimentell widerlegten Hypothesen ja erst aus der Kombination von Wirtschaftswissenschaften und Spieltheorie entstanden sind. Die ausführliche Erörterung, die dieses Problem sicher verdient hätte, würde jedoch den Rahmen dieser Arbeit sprengen.

4.1 Vermeidung eines moralischen Dilemmas

Der erste Schritt zur Modellierung unterschiedlicher ethischer Ansätze besteht darin, den Spielern ein gewisses Verständnis für fremde Moralvorstellungen zu erlauben, so dass die Spieltheorie auch in Situationen angewendet werden kann, in denen Vertreter verschiedener moralischer Sichtweisen miteinander interagieren.

Angenommen ein Anhänger der kantischen Ethik befindet sich in einer Dilemma-Situation, in welcher jede seiner Handlungen zu der Verletzung eines kategorischen Imperativs führen würde. Wäre es für einen Utilitaristen nicht moralisch geboten, diese Dilemma-Situation schon im Vorfeld zu vermeiden, wenn er es könnte? Aus Sicht einer utilitaristischen Ethik verringert jede Handlung des Kantianers seinen Nutzen und damit auch den gesellschaftlichen Gesamtnutzen, da er in jedem Fall gegen seine „Vorlieben“ handeln und ein moralisches Gebot verletzen muss. Wenn der Kantianer aber gar nicht erst in diese Zwickmühle gerät, wird er seiner Verantwortung enthoben und kann unbeschwert seiner Wege gehen. Auch wenn die Entscheidung des Kantianers in diesem Fall nur aus Sicht eines Utilitaristen beschrieben wird und noch nicht spieltheoretisch erklärt werden kann, spricht die Spieltheorie eine Handlungsempfehlung für den Utilitaristen aus, für die von wesentlicher Bedeutung ist, *dass der Utilitarist eine Vorstellung darüber hat, wie der Anhänger der kantischen Ethik handeln wird.*

Man stelle sich eine Situation vor, die auch häufig benutzt wird, um einem Utilitaristen vorzuwerfen, dass für ihn das Schicksal eines Menschen nicht mehr als eine Zahl in einer Statistik ist:⁷⁴ Bei einer Katastrophe in einem Kernkraftwerk scheint der Tod von tausenden Menschen unvermeidlich, wenn nicht ein kleines Rettungsteam geopfert wird, um das Kraftwerk rechtzeitig unter Kontrolle zu bringen. Der Erfolg dieses Teams ist unbestritten, aber es ist zu erwarten, dass die dabei eingesetzten Personen an der hohen radioaktiven Strahlung sterben werden. Für die Entscheidung über den Einsatz des Teams sind zwei Personen verantwortlich – der Anführer des Teams und der Kommandeur des örtlichen Katastrophenschutzes, dem das Team samt Anführer unterstellt sind. Darüber hinaus ist bekannt, dass der Kommandeur ein Anhänger des Utilitarismus ist, während der Teamführer eine kantische Moralvorstellung besitzt. Normalerweise liegt in derartigen Fällen die letzte Entscheidungsgewalt vermutlich beim Anführer des Teams, der für das Leben seiner Leute verantwortlich ist. In diesem Beispiel soll nun aber der Kommandeur des Katastrophenschutzes die Möglichkeit haben, der Entscheidung des Teamführers vorzugreifen und das Team selber in das Kernkraftwerk zu schicken.

⁷⁴ Vgl. zu diesem Vorwurf Smart (1973), S. 69 ff.

Aus Sicht der Spieltheorie lässt sich diese Entscheidungssituation wie folgt darstellen: Der Kommandeur hat den ersten Spielzug und entscheidet, ob er in die Situation eingreift oder nicht. Falls er eingreift, nimmt er die Entscheidung des Teamführers vorweg und beendet damit das Spiel. Falls er nicht eingreift, ist der Teamführer am Zug und entscheidet, ob er sein Team in das Kernkraftwerk schickt oder nicht.

Bei der Bestimmung des Nutzens der verschiedenen Ausgänge des Spiels für die Entscheider sind nun die unterschiedlichen Moralvorstellungen von Bedeutung. Der Teamführer wird aus moralischer Sicht keine seiner beiden Alternativen vorziehen, da er in jedem Fall das Leben von Menschen beenden muss. Er steht vor einem Dilemma, weil er es einerseits als seine Pflicht betrachtet, andere Menschen zu retten, aber andererseits dafür jemanden dem sicheren Tode ausliefern muss. Vielleicht kommt er durch eine längere Überlegung zu dem Schluss, dass er sein Team doch in das Kraftwerk schicken sollte, vielleicht denkt er aber auch solange über die Entscheidung nach, dass es für eine Rettungsaktion zu spät wird. Für dieses Beispiel soll angenommen werden, dass der Teamführer tatsächlich mit der Entscheidung überfordert ist und den letztmöglichen Zeitpunkt für einen Einsatz verstreichen lässt. Er „entscheidet“ sich also dafür, dass Team nicht in das Kraftwerk zu schicken.

Der Kommandeur wird es mit seiner Bewertung der Lage nicht viel einfacher haben, da auch er weiß, dass in jedem Fall Menschen ihr Leben verlieren. Seine utilitaristische Denkweise lässt ihn aber dann doch zu der Entscheidung kommen, dass es besser ist, wenn ein Dutzend Menschen geopfert werden, damit tausende Menschen weiterleben können. Er würde es also vorziehen, wenn der Teamführer seine Leute in das Kraftwerk schicken würde. Sollte der Kommandeur selbst den Einsatzbefehl geben, ändert sich damit nichts an seiner Bewertung der Situation. In diesem Fall ist der Teamführer jedoch seiner Verantwortung enthoben, so dass er aus moralischer Sicht nicht mehr am Ausgang des Geschehens interessiert ist. Allerdings wird er höchstwahrscheinlich erleichtert sein, dass er selbst die Entscheidung nicht treffen muss.

Abbildung 4.1 zeigt das Spiel in Extensivform. Die gewählten Zahlen sollen nur grob die Präferenzen der Spieler repräsentieren und nicht den genauen Nutzenwert wiedergeben, der für den kantianischen Anführer des Teams möglicherweise ins negativ Unendliche geht. Die folgende Argumentation wird durch die Höhe der Zahlen nicht beeinflusst, aber vielleicht etwas durchsichtiger. So kann nämlich die Tendenz des Teamführers, seine Leute nicht in das Kraftwerk zu schicken, durch einen für ihn minimal größeren Nutzen dieser Handlung (-9 statt -10) dargestellt werden. Der Nutzen des Teamführers ist geringer als der des

Kommandeurs, da ersterer durch die Last der Entscheidung zusätzliche „Kosten“ zu tragen hat. Außerdem wurden die Nutzenwerte „normiert“ in dem Sinne, dass der beste erreichbare Zustand in dieser Situation, nämlich die Verhinderung einer Katastrophe, mit einem Nutzen von 0 bewertet wird, obwohl selbstverständlich jeder Ausgang des Geschehens ein Unglück bedeutet.

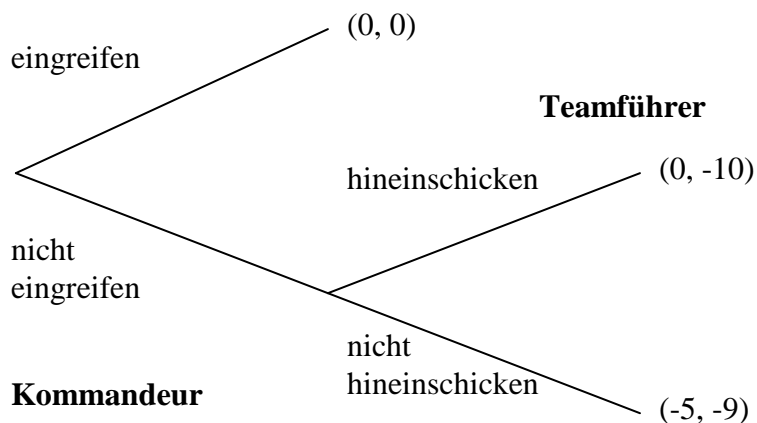


Abb. 4.1 Kernkraftwerk-Beispiel

Für die Entscheidung des Kommandanten ist es nun von Bedeutung, dass er die Handlung des Teamführers richtig einschätzen kann. Denn sollte es nur den geringsten Zweifel daran geben, dass der Anführer des Teams seine Leute in das Kraftwerk schickt, wäre ein Eingreifen des Kommandeurs optimal – gleichgültig ob nur sein eigener Nutzen maximiert werden soll oder die Summe der Nutzenwerte beider Spieler.⁷⁵

Um jedoch überhaupt erkennen zu können, dass der kantisch geprägte Anführer des Teams in einer Dilemma-Situation steckt, muss der utilitaristisch denkende Kommandeur dessen Handlungen vorherahnen, und beurteilen können, wie seine eigenen Handlungen die des Teamführers beeinflussen werden. Er muss also ein Verständnis für die Denkweise des Teamführers haben und daher auch dessen moralischen Ansichten in ein Verhaltensmodell übertragen können. Wenn nun aber ein Utilitarist die Spieltheorie verwendet, um in dieser Situation die optimale Entscheidung zu treffen, dann muss die Spieltheorie auch die nicht-utilitaristischen Moralvorstellungen eines Kantianers darstellen können, wenigstens so wie sie ein Utilitarist interpretieren würde.

⁷⁵ Ein egoistischer Kommandeur würde für „eingreifen“ einen Nutzen von 0 erhalten, während bei „nicht eingreifen“ die Gefahr bestünde, dass er nur einen Nutzen von -5 erlangt. Betrachtet man die Summe der Nutzenwerte, wie es ein Utilitarist machen würde, so wäre „eingreifen“ sogar unabhängig von der Strategie des Teamführers optimal, da dieser durch die Last der Entscheidung einen stark negativen Nutzen erleidet.

Andersherum könnte es aber auch ein Utilitarist sein, der eine schwierige Entscheidung zu treffen hat, während ein Anhänger der kantischen Ethik die Möglichkeit zum Eingreifen besitzt. Wenn nun der Kantianer überlegt, ob eine Intervention moralisch richtig ist, ist *er* es, der die Handlungen seines Mitmenschen vorausahnen können muss, um sagen zu können, wie sich sein Verhalten auf die Handlungen des Utilitaristen auswirkt. Damit kommt aber selbst der Kantianer nicht mehr daran vorbei, sich mit einer Theorie des strategischen Handelns wie der Spieltheorie zu beschäftigen. Er wird sie bei ethischen Fragestellungen jedoch anders einsetzen. Für den Utilitaristen hat die Spieltheorie eine normative Funktion und sagt ihm, welches Verhalten optimal hinsichtlich der Nutzenmaximierung ist. Für den Kantianer ist nur von Bedeutung, auf das Verhalten seiner Mitmenschen reagieren zu können. Die Spieltheorie hilft ihm dabei, dieses Verhalten zu beschreiben und zu erklären, aber nicht ein optimales moralisches Verhalten auszuwählen. Seine moralischen Entscheidungen muss der Kantianer letztendlich also immer noch alleine treffen.

4.2 Erweiterungsmöglichkeiten der Spieltheorie

Im Kernkraftwerk-Beispiel aus dem vorigen Abschnitt wird die Entscheidung auf Grundlage einer nicht-utilitaristischen Ethik einfach aus den angenommenen Präferenzen des Teamführers abgeleitet und damit praktisch von außerhalb vorgegeben. Dadurch entstanden keine weiteren Probleme, da für diesen Spieler keine normative Vorgabe gemacht wurde. Es war der utilitaristische Kommandeur, der die Spieltheorie für seine Entscheidung verwendet hat. Der Umgang mit nicht-utilitaristischen Fragestellungen beschränkte sich also auf die deskriptive Ebene.

In diesem Abschnitt soll weiter dafür argumentiert werden, dass die Spieltheorie auch von ethischen Ansätzen verwendet werden kann, die sich nicht direkt als „utilitaristisch“ bezeichnen oder bezeichnen lassen wollen. Die Spieltheorie muss genau wie die Entscheidungstheorie in den Wirtschaftswissenschaften mit den unterschiedlichsten Darstellungsproblemen fertig werden. Die wichtigsten Faktoren – der Informationsstand der Spieler (vollständig oder eingeschränkt) bzw. der Zeitpunkt der Handlungen (gleichzeitig oder nacheinander) – und ihre Modellierung wurden bereits angesprochen. Viele weitere Gleichgewichtskonzepte, die so genannten „Nash Refinements“⁷⁶ versuchen mehr oder weniger erfolgreich, auch andere Umstände einer gegebenen Situation ausreichend zu beschreiben. Dazu gehören beispielsweise das Erinnerungsvermögen der Spieler, das in Spielen mit mehreren nacheinander erfolgenden Handlungen von Bedeutung sein könnte, oder

⁷⁶ Siehe z. B. Gintis (2000).

auch scheinbar irrationale Verhaltensweisen, bei denen manche Spieler nur noch eingeschränkt das Prinzip der Nutzenmaximierung verfolgen.

Besonders die experimentellen Untersuchungen zum Gefangenen-Dilemma haben die Entwicklung neuer spieltheoretischer Modelle vorangetrieben, weil sich viele Menschen in diesem Spiel *eben nicht* so verhalten, wie es die klassische Spieltheorie vorhersagt, also offenbar andere Vorstellungen über eine optimale Verhaltensweise in derartigen Situation besitzen.

Julian Nida-Rümelin⁷⁷ nennt verschiedene Gründe, warum ein Spieler im Gefangenen-Dilemma eine Strategie wählen könnte, die eigentlich von einer anderen dominiert wird:

1. Der Spieler hat *altruistische Motive*, die ihn die Situation anders bewerten lassen.
2. Er fürchtet eine *Bestrafung* (auf welche Weise auch immer) durch seinen Komplizen, die den Nutzen eines Geständnisses verringern würde.
3. Die Zusammenarbeit mit dem Komplizen, die durch das Spielen der dominierten Strategie erreicht wird, stellt ein *erlerntes Verhalten* dar.
4. Der Spieler verhält sich *tatsächlich irrational*, z. B. aufgrund unzureichender Informationen über die Folgen seiner Handlungen, oder weil er die Struktur des Spieles nicht verstanden hat.
5. Er befolgt eine deontologische Norm (z. B. eine kantische Maxime) und wählt die Handlung, die er für moralisch richtig hält.

Die ersten beiden Punkte stellen keine wirkliche Herausforderung für die Spieltheorie dar. In beiden Fällen ist nämlich die Beschreibung des Spieles, das analysiert werden sollte, unvollständig gewesen: Altruistisches Verhalten kann durch eine Abwandlung der Nutzenfunktion bewältigt werden; im zweiten Fall muss eine weitere Spielrunde hinzugefügt werden, in der die Bestrafung erfolgt.

Eine Strategie, die aus Gewohnheit gespielt wird, so wie die gesellschaftliche Tradition im obigen Beispiel, fällt in den Aufgabenbereich der Evolutionären Spieltheorie. Ein Verhalten, das es Wert ist, gelernt zu werden, hat sich auf irgendeine Art und Weise in ähnlichen Situationen bewährt und als vorteilhaft herausgestellt. Da strikt dominierte Strategien aber niemals *evolutionär stabil* sind,⁷⁸ bleibt auch hier die Frage nach dem tatsächlichen Aufbau des Spieles zurück.

Wenn ein merkwürdiges Verhalten eines Spielers auf Irrationalität zurückgeführt wird, hängt es von der Art und Weise ab, in der sich dieses Verhalten zeigt, ob und wie die

⁷⁷ Vgl. Nida-Rümelin (1993), S. 152ff. Dort finden sich auch Referenzen zu den Vertretern der verschiedenen Varianten von Erklärungen.

⁷⁸ Das liegt daran, dass sie von der Strategie, die ihnen überlegen ist, aus der Population vertrieben werden.

Spieltheorie reagieren kann. Fehlende Informationen sind kein Problem, solange der Spieler die vorhandenen Informationen richtig zu verwenden weiß. Andererseits kann eine Handlung, die keinen rationalen Grund erkennen lässt, immer noch auf *evolutionäre Stabilität* überprüft werden, da die Spieler bei diesem Gleichgewichtskonzept nicht über Vernunft verfügen müssen. Und wenn alle Stricke reißen, kann die Entscheidung eines Spielers als Zufallsereignis behandelt werden, da jede mögliche Handlung wenigstens mit einer gewissen Wahrscheinlichkeit erfolgt, die jedoch von den anderen Spielern nur subjektiv eingeschätzt werden kann.

Der letzte Punkt, eine Handlung aufgrund einer deontologischen Norm, ist der interessanteste und soll deshalb ausführlicher behandelt werden. Der klassische Utilitarismus ist teleologisch in dem Sinne, dass das moralisch Gute unabhängig von dem moralisch Richtigen ist und das Richtige das Gute maximiert.⁷⁹ Das bedeutet, dass alle Handlungen nach ihrem Grad an moralischer Güte sortiert werden können, bevor eine Norm gesucht wird, mit der die bestmögliche Handlung festgeschrieben wird. In einer teleologischen Ethik existieren nur Präferenzen hinsichtlich des Prädikats „moralisch gut“, also des Zweckes einer Handlung. Es ist nicht von Bedeutung, wie dieser Zustand erreicht wird. Anhänger einer deontologischen Ethik andererseits besitzen nur Präferenzen hinsichtlich des Prädikats „moralisch richtig“ einer Handlung, da sie nur interessiert, ob diese Handlung gerecht oder ungerecht ist. Im extremsten Fall ist nur eine einzige Handlung moralisch richtig, jede andere Handlung ist moralisch falsch, ohne dass es dabei zu weiteren Abstufungen der „Falschheit“ kommt.

An dem Beispiel der Aufteilung einer Geldmenge sollen die Unterschiede zwischen teleologischen und deontologischen Ethiken noch einmal erläutert werden: Zwischen zwei Personen A und B soll ein Betrag von 10 Euro aufgeteilt werden; welche Aufteilung ist gerecht, wenn über die Personen nur bekannt ist, dass sie lieber mehr als weniger Geld besitzen?

Vom teleologischen Standpunkt aus gesehen, macht es (unter diesen Annahmen) keinen Unterschied, wie das Geld aufgeteilt wird, da die Gesamtsumme jedes Mal 10 Euro beträgt, so dass auch der Gesamtnutzen dieser „Gesellschaft“ gleich bleibt. Eine deontologische Ethik würde hingegen argumentieren: beide Personen sind identisch, also müssen sie auch denselben Betrag erhalten. Die gerechte Aufteilung lautet dann: Jeder bekommt 5 Euro, also die Hälfte des Betrages. Wie werden aber zwei andere Aufteilungen bewertet, in denen 1. Person A 2 Euro und Person B 8 Euro bekommt oder 2. Person A 10 Euro erhält und Person B leer ausgeht? Eine deontologische Ethik würde vermutlich sagen, dass beide

⁷⁹ So formuliert es John Rawls, der sich dabei wiederum auf Frankena (1963) bezieht. Vgl. Rawls (1999), S. 22.

Aufteilungen gleich ungerecht sind, da sich vernünftigerweise nur die egalitäre Aufteilung als allgemeingültige Regel festschreiben lässt. Damit wäre dann aber selbst eine Aufteilung ungerecht, durch welche Person A 4,99 Euro und Person B 5,01 Euro erhält, obwohl es hier fast keinen Unterschied in den Geldbeträgen mehr gibt.

Warum sollte es aber nicht auch möglich sein, verschiedene Abstufungen von Gerechtigkeit in einer Ethik zu berücksichtigen? Diese müsste immer noch als „deontologisch“ bezeichnet werden, da der Begriff „gerecht“ mit dem Begriff „pflichtgemäß“ verwandt ist. Interessanterweise lässt sich nun aber eine stetige Präferenzordnung über moralisch richtige Handlungen erstellen, so dass diesen Handlungen ein Nutzen zugeordnet werden kann. Der letzte Schritt zum Utilitarismus wird erreicht, wenn eine derartige Ethik das folgende Handlungsprinzip vorgibt: „Wähle diejenige Handlung, die zu der gerechtesten Aufteilung führt.“ Denn dies würde dann ja nichts anderes bedeuten als „Maximiere den Nutzen, den dir deine Gerechtigkeits-Präferenzen vorgeben.“ Das Gute ist dann nicht mehr unabhängig von dem Rechten, sondern das Rechte ist selbst ein Gut.

Wenn man eine Ethik dieser Art dem Utilitarismus zuordnet, scheint es allerdings nicht mehr möglich zu sein, die utilitaristische Ethik als eine Form des Konsequentialismus aufzufassen, da dieser genau wie eine teleologische Ethik als Gegenüber des Deontologismus angesehen wird. Vielleicht kann man zeigen, dass auch der Nutzen, der durch Gerechtigkeitspräferenzen beschrieben wird, letztendlich auf die Folgen der zur Auswahl stehenden Handlungen zurückgeführt werden kann. Dadurch würde jedoch ein fließender Übergang zwischen Deontologismus und Konsequentialismus entstehen, da die unstetige Nutzenfunktion, die aus den Gerechtigkeitspräferenzen eines Anhängers einer deontologischen Ethik gebildet werden kann, wenigstens im Grenzwert durch eine stetige Nutzenfunktion angenähert werden kann (siehe Abb. 4.2).

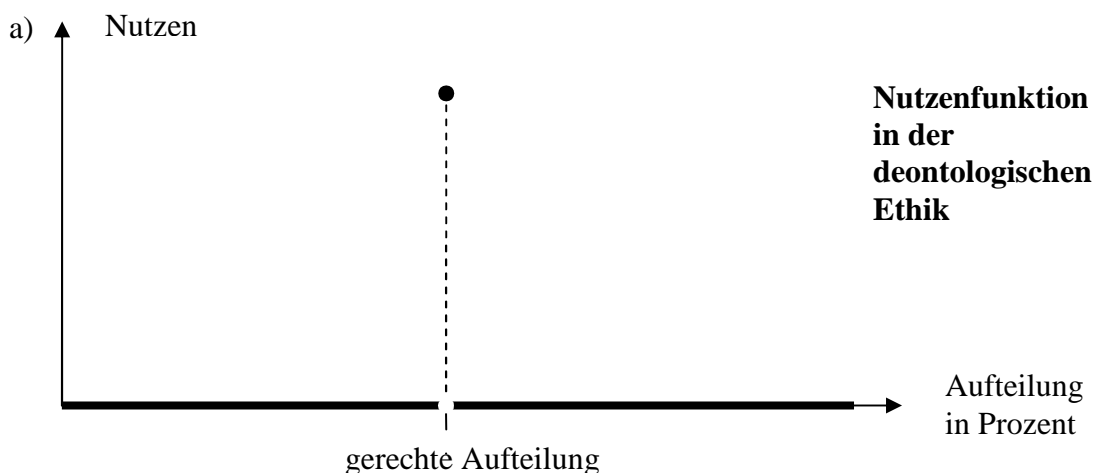


Abb. 4.2 a) Nutzenfunktion in der deontologischen Ethik

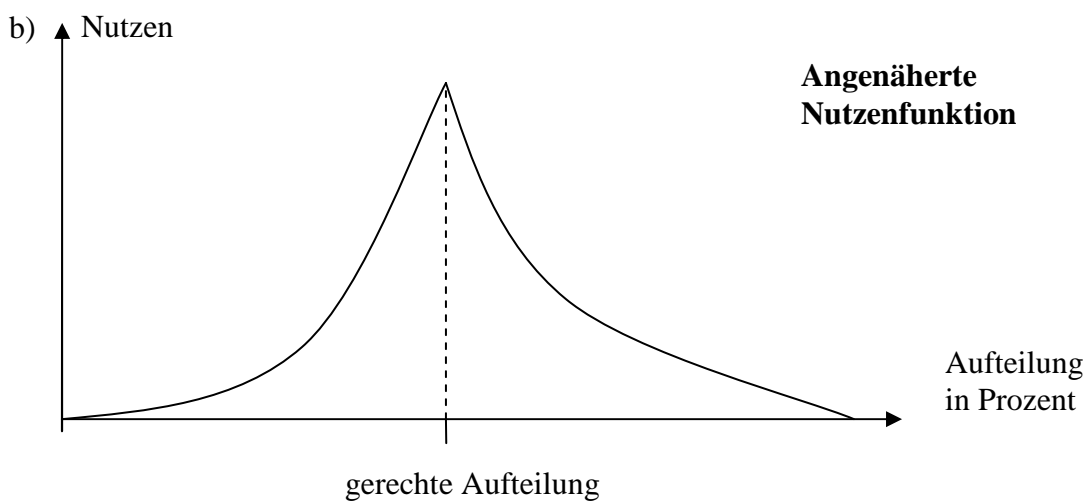


Abb. 4.2 b) Angenäherte Nutzenfunktion

Für die Spieltheorie ist jedoch nur von Bedeutung, dass Entscheidungen aufgrund von Gerechtigkeitsüberlegungen modelliert werden können, wenn die Entscheider mehrere Abstufungen des Begriffes „gerecht“ vornehmen. Wie so eine Modellierung aussehen kann, soll am Beispiel des „Ultimatum-Spiels“ von W. Güth, R. Schmittberger und B. Schwarze⁸⁰ gezeigt werden. Auch in diesem Spiel geht es um die Aufteilung eines Geldbetrages, nur das jetzt einer der Spieler, z. B. Person A das Geld verteilt. Person B hat danach die Möglichkeit, die Aufteilung zu akzeptieren oder abzulehnen, mit der Konsequenz, dass keiner von beiden etwas von dem Geld bekommt. Beispielsweise soll Person A 10 Euro aufteilen. Sie schlägt eine Aufteilung von 6 Euro für sich selbst und 4 Euro für Person B vor. Wenn Person B die Entscheidung akzeptiert, dann erhält A 6 Euro und B 4 Euro. Wenn Person B die Entscheidung jedoch ablehnt, erhalten beide jeweils 0 Euro; das Geld wird also an niemanden ausbezahlt.

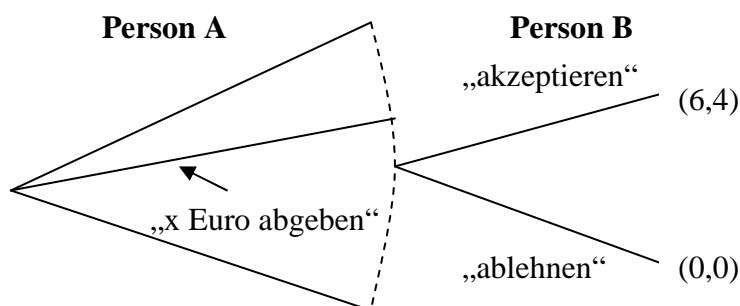


Abb. 4.3 Das Ultimatum-Spiel (Güth, Schmittberger, Schwarze (1982))

⁸⁰ Vgl. Güth/Schmittberger/Schwarze (1982)

Dieses Spiel sorgte für Aufsehen in der experimentellen Wirtschaftsforschung, da die Vorhersage der klassischen Spieltheorie – Person A behält das ganze Geld bis auf einen Eurocent (einen marginal kleinen Betrag) und Person B akzeptiert – nicht bestätigt werden konnte. Vielmehr lehnten Versuchspersonen, die den Standpunkt von Person B einnehmen sollten, sogar Aufteilungen ab, bei denen sie 20% des Gesamtbetrages bekommen hätten, *obwohl sie sich durch die Ablehnung schlechter stellten*. Wenn diese Spieler einer moralischen Norm folgten, dann war sie also ganz bestimmt nicht utilitaristisch, denn der Gesamtnutzen der Spieler wurde ja durch diese Handlung offensichtlich verringert.

Trotzdem gab es einige Versuche von Spieltheoretikern,⁸¹ diese Ergebnisse zu erklären und in ein Modell zu integrieren. Eine Möglichkeit ist eine Bestrafungstendenz bei benachteiligten Spielern, die sogar bereit sind, Geld zu opfern, um einem unfairen Konkurrenten eine Lektion zu erteilen (so genannte *punishment games*). Eine andere Möglichkeit, die *Ungleichheitsaversion*, greift die Idee der Gerechtigkeitspräferenzen auf: eine ungerechte Aufteilung löst einen negativen Nutzen aus, der umso stärker negativ wird, je weiter die tatsächliche Aufteilung von der vorher als gerecht angenommen Aufteilung abweicht. Die Nutzenfunktion eines ungleichheitsaversen Spielers hat eine Form wie in Abbildung 4.2 b).

Der Anwendungsbereich der Spieltheorie in der Ethik ist also nicht nur auf den Utilitarismus beschränkt. Insbesondere wenn davon ausgegangen werden kann, dass Vertreter anderer Ethiken ein konsistentes moralisches Verhalten zeigen, das Rückschlüsse auf die Gründe für Entscheidungen liefert, steht einer deskriptiven Analyse dieses Verhaltens in spieltheoretischen Modellen nichts im Wege. Damit ist in jedem Fall gesichert, dass ein Anwender der Spieltheorie eine für ihn optimale Verhaltensempfehlung erhält. Ob sich Vertreter nicht-utilitaristischer Ethiken jedoch tatsächlich einen Gefallen tun, wenn sie die Spieltheorie verwenden, hängt davon ab, ob die Spieltheorie bei ihren normativen Vorgaben neutral bleibt und nicht vielleicht eher eine utilitaristische Denkweise fördert. Dieses Problem soll im nächsten Abschnitt untersucht werden.

⁸¹ Für einen Überblick siehe z. B. Falk/Fischbacher (2000) oder Gintis (2000).

5. Die Unparteilichkeit der Spieltheorie bei der Beschreibung ethischer Ansätze

Durch den Einsatz einer Hilfswissenschaft dürfen die grundlegenden Aussagen einer Theorie nicht verfälscht werden. Weil die Spieltheorie aber stark utilitaristisch geprägt ist, besteht die Gefahr, dass eine Ethik, die spieltheoretische Konzepte verwendet, ihre Eigenständigkeit aufgibt und zu einer bloßen Variante des Utilitarismus wird. In einem spieltheoretischen Modell müssen die Spieler mit ihrem Verhalten ein höheres Ziel verfolgen, das im Prinzip der Nutzenmaximierung abgebildet werden kann. Es stellt sich jedoch die Frage, ob dabei nicht der Zweck einen höheren Stellenwert bekommt als die Mittel, mit denen er realisiert wird. Ken Binmore äußert sich dazu, wie folgt:⁸²

“[F]ar from game theory demanding that the ends be taken to justify the means, the discipline of setting up a problem as a [noncooperative] game makes it impossible for the model-builder to neglect to take into account the attitude of his players to such questions. The *definition* of an outcome in a game includes not only the physical consequences for each player at the end of its play, but also the manner in which the physical consequences were achieved. One is therefore provided with total flexibility in deciding how the ends are to be balanced against the means.”

Die Spieltheorie ist also als Ableger des Utilitarismus nicht zwangsläufig ebenfalls eine teleologische Theorie. Sowohl die Folgen einer Handlung, als auch der Weg, auf dem diese Folgen zustande gekommen sind, sind in dem Nutzenwert enthalten, den der Spieler bei Ausübung dieser Handlung erhalten kann. Wie im vorigen Abschnitt angedeutet wurde, liefert die empirische Wirtschaftsforschung häufig genug Beispiele dafür, dass einige Menschen in experimentellen Situationen Entscheidungen treffen, bei denen sie einen schlechteren Gesamtzustand einem besseren vorziehen, z. B. wenn sie Geld opfern, damit ein Mitspieler ebenfalls etwas von seinem Geld verliert. Der Grund dafür kann eine Bestrafung für ungerechtes Handeln im Verlaufe des Spieles sein oder einfach pure Boshaftigkeit. Der Punkt ist jedoch, dass die Spieltheorie Modelle konstruieren muss, in denen diesen Handlungen ein Nutzen zukommt, damit das Verhalten von Spielern dieses Typs vorhergesagt werden kann.

Möglicherweise verfälscht also die Spieltheorie selbst die Bewertung ethischer Ansätze, weil ihre Konzepte von dem Anwender eine gewisse Denkrichtung erfordern, die ein konsistentes, strategisch motiviertes Verhalten hochhält, mit der zusätzlichen Vorgabe, sich so gut wie möglich über eine Entscheidungssituation zu informieren. Indem die Spieltheorie

⁸² Binmore (1994), S. 109.

dieses Verhalten vorschreibt, trifft sie eine normative Aussage und lässt ihre Rolle als Hilfswissenschaft hinter sich. Damit ist sie aber auch nicht mehr neutral gegenüber einer Ethik, die den Entscheidern einen größeren Spielraum gibt und auch einmal erlaubt, eine Entscheidung weniger gut vorbereitet anzugehen, wenn dadurch ein ähnlich gutes Ergebnis erzielt wird.

Daher muss die Frage der Unparteilichkeit der Spieltheorie auf normativer Ebene wohl negativ beantwortet werden, denn eine spieltheoretisch erzeugte Verhaltensvorschrift maximiert *immer* den Nutzen der betroffenen Person. Auch wenn der Begriff „Nutzen“, wie bereits beschrieben, sehr weit ausgedehnt werden kann, wird doch nicht mehr getan, als bestimmte moralische Zielvorgaben in die utilitaristische Vorgehensweise der Nutzenmaximierung zu übertragen.

Auf deskriptiver Ebene sieht das Ganze schon anders aus: die Evolutionäre Spieltheorie kann die Entstehung unterschiedlicher Moralvorstellungen erklären, *ohne* ein nutzenmaximierendes Verhalten der Vertreter dieser ethischen Ansätze zu unterstellen. Stattdessen sind es die moralischen Normen selbst, die um ihr Überleben „kämpfen“ und nach der größten *fitness* streben. Ziel dieses Abschnittes ist deshalb die Bestätigung der folgenden These:

These 3: Die (evolutionäre) Spieltheorie ermöglicht eine neutrale Beschreibung einer moralischen Entscheidung aus nicht-utilitaristischen Motiven.

Der Ausweitung dieses evolutionstheoretischen Ansatzes auf normative Aufgaben einer Hilfswissenschaft in der Ethik steht jedoch das Sein-Sollen-Problem der Evolutionären Ethik im Wege. Die Tatsache, dass sich eine bestimmte Verhaltensweise in einem evolutionären Prozess behauptet hat, ist außerdem für sich genommen noch kein Grund, dieses Verhalten als Norm festzusetzen.

Die Probleme der Evolutionären Ethik werden ausführlicher gegen Ende des Abschnittes beschrieben. Zunächst sollen aber die Grenzen der nicht-evolutionären Spieltheorie ausgelotet werden. Immerhin ist die klassische Spieltheorie zumindest in der Lage, gewisse utilitaristische Varianten kritisch zu betrachten. Besonders bei der Gegenüberstellung von Akt- und Regel-Utilitarismus kann die Spieltheorie eine vermittelnde Rolle übernehmen, da sie Argumente für und gegen *beide* Ansichten liefert. So benutzt z. B. Franz von Kutschera⁸³ ein spieltheoretisches Beispiel um den Regel-Utilitarismus zu kritisieren, wohingegen John C.

⁸³ Vgl. von Kutschera (1982) S. 169ff.

Harsanyi⁸⁴ strategische Argumente verwendet, um diese utilitaristische Variante zu verteidigen. Im nächsten Abschnitt sollen beide Positionen dargestellt werden, um zu zeigen, dass auch die klassische Spieltheorie nicht zwangsläufig eine gewisse Art von Moral mit sich bringt. Ein Beispiel von J. J. C. Smart⁸⁵ soll schließlich mit der Verwendung von gemischten Strategien eine mögliche Lösung dieses Konfliktes aufzeigen.

5.1 Akt- vs. Regel-Utilitarismus aus Sicht der Spieltheorie

Um den Akt-Utilitarismus in einem spieltheoretischen Modell mit dem Regel-Utilitarismus vergleichen zu können, müssen die Unterschiede in den Auffassungen beider Ethiken in eine mathematische Sprache übersetzt werden. Harsanyi äußert sich dazu wie folgt:⁸⁶

„An act-utilitarian moral agent assumes that the strategies of all other moral agents (including those of all other utilitarian agents) are given and that his task is merely to choose his own strategy so as to maximize social utility when all other strategies are kept constant. In contrast, a rule-utilitarian moral agent will regard not only his own strategy but also the strategies of all other rule-utilitarian agents as variables to be determined during the maximization process so as to maximize social utility.”

Der Akt-Utilitarist verhält sich also in einem spieltheoretischen Modell wie ein normaler rationaler Spieler, mit der einzigen Ausnahme, dass seine Nutzenfunktion von dem Gesamtnutzen aller beteiligten Spieler abhängt. Er wählt seine eigene Strategie als beste Antwort auf jede gegebene Strategie seiner Mitspieler.

Anders der Regel-Utilitarist: er bestimmt die Handlungen aller Spieler gleichzeitig, indem er eine Regel vorgibt, an die sich alle halten müssen. Im Ergebnis trifft also jeder Spieler in einer identischen Situation dieselbe Entscheidung, so dass in symmetrischen Spielen, in denen jeder Spieler über dieselben Strategien verfügt und für gleiche Kombinationen von Handlungen den gleichen Nutzen erhält, alle Spieler gleich handeln werden.

5.1.1 Die Positionen von Harsanyi und von Kutschera

Harsanyi bevorzugt die regel-utilitaristische Variante, weil sie offenbar dabei hilft, gewisse Spielsituationen zu erreichen, die ansonsten eher unwahrscheinlich wären. Als Beispiel nennt er eine Abstimmung in der über die Umsetzung einer politischen Maßnahme entschieden werden soll. Bei der Vielzahl der Wähler hat ein einzelner Befürworter der Maßnahme nur einen geringen Anreiz, seine Stimme abzugeben, weil sie allein für den Ausgang der Wahl

⁸⁴ Vgl. Harsanyi (1982).

⁸⁵ Vgl. Smart (1973), S. 57ff.

⁸⁶ Harsanyi (1982), S. 57.

kaum entscheidend sein dürfte. Es ist nämlich äußerst unwahrscheinlich, dass ein Wahlergebnis so knapp ausfällt, dass eine einzige Stimme den Ausschlag geben könnte. Im Einzelfall sind die Kosten des Wählens also möglicherweise größer als der Nutzen der Stimmabgabe, so dass nach Harsanyi ein Akt-Utilitarist nicht wählen wird. Falls aber jeder Wähler diese Sichtweise teilt, wird am Ende niemand zur Wahl gehen, so dass die eigentlich sinnvolle politische Maßnahme nicht umgesetzt werden kann.

Anders sieht das Ganze aus, wenn die Befürworter der Maßnahme Regel-Utilitaristen sind. Denn für sie gibt es nach Ansicht von Harsanyi nur zwei Alternativen, zwischen denen sie sich entscheiden müssen: Entweder alle sollten ihre Stimme abgeben, oder niemand sollte wählen gehen. Da die politische Maßnahme aber offenbar für die Gesellschaft von Vorteil ist, wird sich ein Regel-Utilitarist also in jedem Fall für die Stimmabgabe entschließen.

In manchen Situationen ist es jedoch von Nachteil, wenn sich alle Personen identisch verhalten, wie Franz von Kutschera mit einem einfachen Beispiel deutlich macht.⁸⁷

Gegeben sei folgendes Spiel:

		Spieler 2	
		A	B
Spieler 1	A	3, 3	5, 5
	B	5, 5	0, 0

Abb. 5.1 Gegenbeispiel zum Regel-Utilitarismus (von Kutschera (1982))

Ein Regel-Utilitarist lässt in dieser Situation nur zwei Verhaltensweisen zu: entweder beide Spieler spielen „A“ oder beide spielen „B“. Zu diesem Schluss kommt wenigstens von Kutschera, da ja beide Spieler in dieser symmetrischen Situation *dieselbe* Regel befolgen müssen. Jedoch sind weder (A, A) noch (B, B) Nash-Gleichgewichte in diesem Spiel.

Wenn man aber den Gesamtnutzen der beiden Spieler maximieren will, wäre es optimal, wenn beide Spieler unterschiedliche Strategien wählen würden. Es gibt also Situationen, in denen es unvorteilhaft ist, eine Verhaltensweise als allgemeingültige Regel festzulegen. Nämlich immer dann, wenn nur einige Spieler eine Handlung ausführen sollen (d. h. nicht alle, aber auch nicht keiner).

Diese Kritik des Regel-Utilitarismus erweist sich jedoch als relativ zahnlos, wenn man die folgenden Punkte beachtet: Zunächst einmal sind die wenigsten Situationen wirklich

⁸⁷ Vgl. von Kutschera (1982) S. 169ff.

symmetrisch. Wenn allerdings beide Spieler zwischen unterschiedlichen Handlungen wählen können, dann sind auch asymmetrische Nash-Gleichgewichte erreichbar. Aber auch in symmetrischen Spielen gibt es noch einen Ausweg: das gemischte Gleichgewicht. Warum sollte nicht auch eine gemischte Strategie als allgemeingültige Regel festgelegt werden können?

5.1.2 Gemischte Strategien als allgemeingültige Regel

Theoretisch sollte nichts dagegensprechen, ein Zufallsereignis zwischen zwei Strategien entscheiden zu lassen. Im obigen Beispiel wäre es optimal, wenn beide Spieler in durchschnittlich sieben von neun Fällen A und in den restlichen beiden Fällen B spielten. Natürlich ist es nicht möglich, mit dieser Strategie in *allen* Fällen eine Handlungskombination (A, B) oder (B, A) zu erzielen. In etwa 60% ($7/9$ mal $7/9$) der Fälle wird vielmehr (A, A) gespielt und zu 5% sogar (B, B). Der erwartete Nutzen dieser Strategie ist mit ca. 3,5 aber größer als der einer Regel, die eine *reine* Strategie vorschreibt. In Abschnitt 3.2.7 wurde jedoch bereits erläutert, dass das Befolgen einer gemischten Strategie im Alltag möglicherweise den Eindruck der Unvernünftigkeit erweckt, weil in ähnlichen Situationen unter Umständen ganz unterschiedliche Handlungen ausgeführt werden. Im Folgenden soll dargestellt werden, dass das Konzept der gemischten Strategie aber dennoch in der Ethik von Bedeutung sein könnte, weil es einen Kompromiss zwischen unterschiedlichen ethischen Standpunkten ermöglicht, so dass die Spieltheorie in Diskussionen mit verhärteten Fronten vermitteln könnte.

Indem er die Idee der gemischten Strategie aufgreift, will J. J. C. Smart zeigen, dass eine spieltheoretische Darstellung insbesondere einem Akt-Utilitarismus entgegen kommt.⁸⁸ Ausgangspunkt ist ein Beispiel von R. B. Brandt, mit dem dieser beweisen will, dass der Utilitarismus inkonsistent ist, weil er zu Ergebnissen führt, die ein Utilitarist selbst eigentlich gar nicht für vernünftig hält, da ihm sein gesunder Menschenverstand zu einer anderen Handlung rät:⁸⁹

England befindet sich im Krieg, und die Regierung ordnet an, dass Wohnungen nur noch auf etwa 15°C geheizt werden sollen, um Strom und Gas zu sparen. Ein französischer Utilitarist, der gerade in England lebt, macht sich nun darüber Gedanken, welches Verhalten in dieser Situation moralisch richtig ist: Strom zu sparen und zu frieren, oder die Anordnung der Regierung zu ignorieren, um es zu Hause schön warm zu haben. Er überlegt sich, dass die sich meisten Engländer wohl an die Anordnung halten werden. Wenn einige wenige jedoch

⁸⁸ Vgl. Smart (1973), S. 57ff.

⁸⁹ Vgl. Brandt (1959), S. 389.

trotzdem wie gewohnt heizen, wird das im Gesamtergebnis nicht auffallen, während es diesen Abweichlern aber dadurch viel besser geht. Deshalb kommt der Utilitarist zu dem Schluss, dass er selbst auch nicht an Strom und Gas sparen sollte, da er damit den Gesamtnutzen der Gesellschaft erhöht.

Die Situation ist ähnlich der eines Gefangenendilemmas: Für die Gesellschaft wäre es optimal, wenn alle Bürger Energie sparen würden, aber jeder einzelne Bürger hat einen größeren Anreiz, *nicht* zu sparen. Der gesunde Menschenverstand vermittelt in dieser Situation offenbar den besseren Durchblick als der Kalkül des rationalen Utilitaristen. Zwar mag es nicht auffallen, wenn einige wenige Bürger nicht mitmachen, **falls aber alle so denken**, ist die Anordnung sinnlos gewesen. Man muss also dem Herzen folgen und nicht dem Verstand, um hier das Richtige zu tun. Was nützt dem Utilitaristen aber seine ausgefeilte moralische Theorie, wenn er am Ende gar nicht auf sie zurückgreifen will?

Diesen Einwand kontert Smart mit der Behauptung, dass der Franzose bei seiner Überlegung ja schon davon ausgegangen sei, dass nur sehr wenige Leute so denken wie er und deshalb auch brav den Anordnungen der Regierung Folge leisten. Falls eine Gesellschaft aber wirklich ausschließlich aus rationalen Akt-Utilitaristen bestünde, dann könne diese Situation durch die Spieltheorie aufgelöst werden. In diesem Fall wäre eine optimale Strategie nämlich eine Mischung aus den Optionen „Heizen“ oder „Sparen“. Wenn jeder Engländer nur zu einer gewissen Wahrscheinlichkeit die Anordnung der Regierung befolgt, dann werden möglicherweise im Ergebnis noch genug Engländer sparen, dass sich der Appell lohnt.

Wenn sich ein Regel-Utilitarist dieses Ergebnis zunutze machen will, muss er glaubhaft vermitteln, dass *jeder einzelne* Engländer zu einem gewissen Prozentsatz „Heizen“ und andernfalls „Sparen“ wird, also gewissermaßen vor seiner Entscheidung eine Münze wirft. Für einen außen stehenden Betrachter macht es tatsächlich keinen Unterschied, ob die gemischte Strategie als zufällige Entscheidung einer Einzelperson interpretiert wird oder angenommen wird, dass es verschiedene Typen von Spielern gibt, die zu gewissen Anteilen in einer Bevölkerung vertreten sind. Trotzdem dürfte es schwer zu erklären sein, warum eine moralische Regel von allgemeiner Gültigkeit mit einem Zufallsereignis verbunden werden sollte. Daher muss wohl als Fazit festgehalten werden, dass der Akt-Utilitarismus mehr Unterstützung durch die Spieltheorie erhält als der Regel-Utilitarismus.

Es ist also durchaus möglich, mithilfe der Spieltheorie eine utilitaristische Ethik zu kritisieren. Die Umsetzung dieser Kritik führt jedoch vermutlich wieder auf eine andere Variante des Utilitarismus. Denn die moralischen Vorgaben, die ein Anwender der Spieltheorie errechnet, maximieren in jedem Fall seinen Nutzen und haben deshalb immer

einen utilitaristischen Beigeschmack. Zwar wurde in Abschnitt 4 nahe gelegt, dass auch eine Reihe von nicht-utilitaristischen Moralvorstellungen in spieltheoretischen Modellen erfasst werden können, aber nur indem sie in einen Mechanismus der Nutzenmaximierung übersetzt werden. Damit fügt die Spieltheorie einer moralischen Verhaltensvorschrift in jedem Fall eine zusätzliche Zielvorgabe hinzu, nämlich den eigenen Nutzen mit einer moralisch richtigen Handlung zu maximieren. Solange aber nicht gezeigt wird, dass der Begriff „Nutzen“ so weit gefasst werden kann, dass tatsächlich *jede* Grundlage von Entscheidungen in einer Nutzenfunktion erfasst werden kann, so dass entweder jede Ethik dem Utilitarismus zugerechnet werden muss oder sich der Nutzen als neutrale mathematische Größe etablieren kann; solange können wohl nur Utilitaristen die Spieltheorie gefahrlos verwenden, um in einer ethischen Entscheidungssituation die moralisch richtige Handlung zu bestimmen.

Auf normativer Ebene kann die Spieltheorie auch deshalb keinen vollständig unparteiischen Standpunkt einnehmen, weil bereits die rationale Denkweise, die sie von ihrem Anwender verlangt, ihr Einsatzgebiet in der Ethik einschränkt. Zwar bemerkt John C. Harsanyi, dass sowohl Ethik als auch Spieltheorie nur Bestandteile einer übergeordneten Theorie des rationalen Verhaltens sind und „im wesentlichen dieselbe Methode verwenden“.⁹⁰ Das bedeutet jedoch nicht, dass nicht auch „unvernünftige“ Wesen, wie z. B. manche Tierarten, eine gewisse Art von Moral entwickeln können, die ihr Verhalten steuert, ohne dass sie vielleicht selbst begründen können, warum sie dieses Verhalten ausüben. Wie in Abschnitt 3.3 dargestellt wurde, kann in diesem Fall aber ein evolutionärer Ansatz wenigstens beschreiben, wie diese Wesen miteinander interagieren. Im nächsten Abschnitt soll deshalb erörtert werden, wie die Evolutionäre Spieltheorie eine deskriptive Ethik unterstützen kann.

5.2 Evolutionäre Spieltheorie und deskriptive Ethik

Selbst wenn sich zeigen lässt, dass ein rationaler Anwender der Spieltheorie ein Utilitarist sein muss, bleibt also immer noch die Evolutionäre Spieltheorie, die Entscheidungen zwischen Strategien nicht mehr von der Vernunft der Spieler, sondern von einem natürlichen Prozess abhängig macht. Ihre Unparteilichkeit wird dadurch nahe gelegt, dass sich grundsätzlich auch eine Konstellation von Strategientypen ergeben könnte, in der ein Utilitarist eine geringere *fitness* besitzt als Anhänger anderer moralischer Strömungen, so dass eine Kritik der Evolutionären Spieltheorie am Utilitarismus vielleicht nicht wahrscheinlich, aber durchaus möglich ist.

⁹⁰ Vgl. Harsanyi (1982), S. 42ff.

5.2.1 Evolutionäre Ethik und Spieltheorie

Die Verbindung ethischer Konzepte mit der Evolutionstheorie gründet sich auf die Beobachtung, dass sich soziales Verhalten offenbar über einen längeren Zeitraum hinweg entwickelt hat. Laut Gerhard Vollmer, einem wichtigen Vertreter der Evolutionären Ethik, kann die Evolutionstheorie die Entstehung eines solchen Verhaltens nicht nur beschreiben, sondern auch erklären. Spätestens mit der Einsicht, „dass die unserem Verhalten zugrunde liegenden Maßstäbe und Normen nicht unverrückbar festliegen, sondern ihrerseits dauerndem Wandel unterworfen sind,“⁹¹ kommt aber auch die Frage nach der Verbindung zwischen den Tatsachenaussagen der naturwissenschaftlichen Evolutionstheorie und den normativen Aussagen der geisteswissenschaftlichen Ethik.

Die Behauptung, dass ein gewisses Verhalten in einer Gesellschaft als moralisch wertvoll angesehen wird, kann überprüft und durch Beobachtungsdaten widerlegt oder bestätigt werden. Die Richtigkeit eines moralischen Gebotes, das dieses Verhalten vorschreibt, ist zwar nicht von wissenschaftlichen Beobachtungen abhängig, Normen sind aber – so Vollmer – *kritisierbar*. Unter gewissen Umständen kann es unvernünftig sein, an einer Norm festzuhalten, weil sie möglicherweise „nicht anwendbar, nicht durchführbar oder nicht durchsetzbar ist oder zu anderen Normen in Widerspruch steht“.⁹² Wenn eine Norm auf diese Weise kritisiert werden kann, ist sie aber noch nicht *widerlegt*, da es eben auch Umstände geben kann, unter denen das Befolgen dieser Norm durchaus vernünftig ist. Beim Gebrauch von Normen stehen also offenbar Ziele im Hintergrund, die gewissermaßen eine Kosten-Nutzen-Entscheidung für oder gegen diese Norm erlauben. Vollmer bietet hier die Rückführung auf „Basis- oder Supernormen“ an, wie zum Beispiel „Die Menschheit soll überleben“, oder „Die Evolution soll weitergehen“.⁹³ Diese Supernormen sind genau die Ziele, die bei einer kritischen Betrachtung einer Norm berücksichtigt werden und die eine Bewertung einer Norm erst möglich machen.

Eine *evolutionär stabile Strategie* kann auch dann entstehen, wenn die Spieler *keine* Vorstellung davon haben, was sie durch ihr Verhalten erreichen wollen. Die Evolutionäre Spieltheorie untersucht, wie gesagt, die Entwicklung einer bestimmten Population, in der gewisse Typen von Strategien verfolgt werden. Beobachtet man diese Population über einen Zeitraum von mehreren Generationen, so kann man sehen, welche Strategien „erfolgreicher“ in dem Sinne sind, dass sie zu größeren Anteilen in der Population vertreten sind.

⁹¹ Siehe Vollmer (1993), S. 166.

⁹² Siehe Vollmer (1993) S. 169.

⁹³ Vgl. Vollmer (1993) S. 184f.

Mithilfe der Evolutionären Spieltheorie kann man erklären, warum manche Normen besser zurechtkommen als andere. Es ist jedoch offenbar nicht möglich, im Voraus zu bestimmen, wie sich eine neue Verhaltensstrategie in einer Population behaupten wird. Und erst *nachdem* man diese Beobachtung gemacht hat, kann man das Zusammenspiel dieser neuen Strategie mit anderen Strategien bestimmen. Dies bedeutet auch, dass das Entstehen neuer Strategien nicht vorhersagbar ist, sondern nur im Nachhinein erklärt werden kann. Sowieso macht die Evolutionäre Spieltheorie keine normativen Aussagen, sondern Aussagen über das *Bestehen* von Normen.

5.2.2 Welches moralische Verhalten bevorzugt die Evolutionäre Spieltheorie?

Durch den Rückzug auf die Position des Beobachters könnte es die Evolutionäre Spieltheorie schaffen, sich von dem utilitaristischen Einfluss zu lösen. Denn warum sollte sich ausgerechnet eine utilitaristische Ethik in einem evolutionären Prozess behaupten können? Für die *fitness* einer Verhaltensstrategie ist schließlich nicht von Bedeutung, welche Ziele ein Individuum mit seinen Handlungen verfolgt, also ob es selbst seinen Nutzen maximiert.

Warum sollte aber überhaupt irgendeine Art von Moral evolutionär stabil sein? Vielleicht entsteht vielmehr eine Art Zyklus, ähnlich wie im „Stein-Schere-Papier“-Spiel,⁹⁴ in welchem verschiedene Moralvorstellungen abwechselnd in einer Population vorherrschend sind. So könnte eine überwiegend kantisch geprägte Gesellschaft wegen ihres uneigennütigen Verhaltens anfällig für egoistische Menschen sein, die dieses Verhalten auszunutzen wissen. Auf der anderen Seite könnten Utilitaristen möglicherweise in eine vornehmlich egoistisch denkende Population eindringen, weil sie durch Zusammenarbeit gewisse Vorteile erzielen.⁹⁵ Und abschließend wäre es denkbar, dass die Anhänger der kantischen Ethik Vorteile in einer größtenteils utilitaristischen Gesellschaft besitzen, weil sie ohne großen Rechenaufwand und damit schneller zu Entscheidungen gelangen. Demnach würden sich Kantianer, Egoisten und Utilitaristen über mehrere Generationen hinweg abwechseln, ohne dass eine Verhaltensstrategie dauerhaft die Oberhand gewinnt.

Wahrscheinlicher ist jedoch, dass sich ein gemischtes Gleichgewicht bildet, so dass mehrere moralische Ansichten gleichzeitig in einer Gesellschaft vertreten sind. Ein gewisser Prozentsatz der Bevölkerung verfolgt eine kantische Ethik und handelt aus Pflichtgefühl, ein weiterer Anteil bemüht sich, im utilitaristischen Sinne das Wohlergehen der Mitmenschen zu maximieren, und der Rest denkt ausschließlich an sich selbst.

⁹⁴ Siehe oben, Abschnitt 3.3.3.

⁹⁵ So ähnlich argumentieren Bester und Güth (1998).

Offenbar ermöglicht die Evolutionäre Spieltheorie also tatsächlich einen unparteiischen Vergleich unterschiedlicher Moralvorstellungen. Diese Neutralität entsteht aber genau genommen nur dadurch, dass der moralübergreifende Maßstab des evolutionären Erfolges aufgrund des Sein-Sollen-Problems der Evolutionären Ethik bisher selbst nicht die Grundlage einer Ethik bilden konnte. Die Beschränkung auf die deskriptive Ebene ist also vielleicht gar keine Schwäche des evolutionären Ansatzes sondern seine eigentliche Stärke.

6. Abschließendes Fazit und Ausblick

Diese Arbeit sollte gezeigt haben, dass spieltheoretische Konzepte für den Utilitarismus unverzichtbar sind, weil wegen der ständigen Interaktion unterschiedlicher Menschen eine strategische Herangehensweise an ethische Fragestellungen nötig ist. Vertreter anderer Ethiken dürften jedoch skeptisch sein, dass die Spieltheorie ihre Moralvorstellungen gut genug repräsentieren kann, um ihr normative Aufgaben anzuvertrauen und mit ihrer Hilfe neue moralische Vorschriften zu formulieren. Allerdings steckt die Spieltheorie auf diesem Gebiet noch in den Kinderschuhen, so dass es nicht unvorstellbar ist, dass neue spieltheoretische Konzepte auch in Ethiken zum Einsatz kommen können, die dem Utilitarismus ähnlich sind (im Sinne einer messbaren Rangordnung von Präferenzen).

Auf der deskriptiven Ebene kann die Spieltheorie da schon mehr leisten. Evolutionäre Gleichgewichtskonzepte abstrahieren von den intellektuellen Fähigkeiten der betrachteten Individuen, so dass es nicht mehr von Bedeutung ist, ob eine Verhaltensweise aus Pflichtgefühl oder dem utilitaristischen Drang nach Verbesserung des Wohlergehens der Gesellschaft erfolgt: eine Norm kann sich entweder in einem evolutionären Prozess behaupten oder nicht. Damit existiert nun auch eine Vergleichsmöglichkeit unterschiedlicher ethischer Ansätze hinsichtlich eines gemeinsamen Zieles, eines Zweckes den jede Ethik als Motivation nimmt, überhaupt bestimmte Verhaltensweisen vorzugeben. Man mag darüber streiten, ob die evolutionäre *fitness* tatsächlich für jede Ethik erstrebenswert ist, aber es spricht grundsätzlich nichts dagegen, auch einen anderen Maßstab zu nehmen, um den Erfolg einer Norm daran zu messen.

Welche Rolle die Spieltheorie zukünftig in der Ethik spielen wird, hängt jedoch von verschiedenen Punkten ab, bei denen die Ergebnisse der weiteren Forschung entscheidend sind. Einerseits könnte an einem Beispiel gezeigt werden, dass tatsächlich manche Ethiken nicht in der Spieltheorie dargestellt werden können, was ihre Bedeutung sicherlich schwächen würde. Vielleicht gelingt aber auch ein Vollständigkeitsbeweis, der belegt, dass die Spieltheorie in der Ethik universell einsetzbar ist. Indirekt könnte daraus dann möglicherweise sogar abgeleitet werden, dass im Prinzip *jede* Ethik utilitaristisch ist (auch wenn ein derartiges Ergebnis noch pure Spekulation ist).

Von besonderem Interesse für die Evolutionäre Spieltheorie wäre eine Lösung des Sein-Sollen-Problems, also eine Möglichkeit moralische Normen durch wissenschaftliche Tatsachen zu stützen und zu begründen, weil dadurch ihr Einsatzgebiet von der deskriptiven auf die normative Ebene ausgedehnt werden könnte, ohne einen naturalistischen Fehlschluss befürchten zu müssen. Dazu muss vor allen Dingen die Definition der *fitness* so in einer

naturwissenschaftlichen Theorie verankert werden, dass die Überlebenswahrscheinlichkeit einer bisher unbekanntem Verhaltensweise bestimmt werden kann, *bevor* sie als Mutation in eine bestehende Bevölkerung eindringt.

Festzuhalten bleibt jedoch, dass Spieltheorie und Ethik ähnliche Anforderungen an die Personen stellen, mit deren Verhalten sie sich beschäftigen. Um Spieltheorie anzuwenden muss man

- alle möglichen Handlungen in Erwägung ziehen,
- die Folgen dieser Handlungen bedenken,
- die Vorlieben und Verhaltensweisen anderer Spieler berücksichtigen,

und auf Grundlage dieser Informationen ein vernünftiger Verhalten zeigen, das andere Menschen nachvollziehen können. Das einzige, was der Spieltheorie fehlt, um den Rang einer Ethik einzunehmen, ist eine *Richtung*, also ein Maßstab des moralisch Guten.

Literaturverzeichnis

- Aritoteles (1972): „Nikomachische Ethik“, herausgegeben von G. Bien, Hamburg
- Bentham, J. (1789, 1970): „An Introduction to the Principles of Morals and Legislation“, herausgegeben von J. H. Burns und H. L. A. Hart, London
- Bester, H. / Güth, W. (1998): „Is Altruism Evolutionarily Stable“, Journal of Economic Behavior and Organization
- Bicchieri, C. (1998): „Decision and Game Theory“, in: Craig, E. (1998)
- Binmore, K (1994): „Game Theory and the Social Contract Volume I: Playing Fair“, The MIT Press, Cambridge, London
- Boella, G. / van der Torre, L. (2003): „Obligations as Social Constructs“, in: A. Cappelli, F. Turini (Hrsg.) (2003): AI*IA 2003: Advances in Artificial Intelligence, 8th Congress of the Italian Association for Artificial Intelligence, Pisa, Italien
- Braithwaite, R. B. (1955): „The Theory of Games as a Tool for the Moral Philosopher“, Cambridge
- Brandt, R. B. (1959): „Ethical Theory“, Englewood Cliffs, New Jersey
- Brandt, R. B. (1979): „A Theory of the Good and the Right“, Oxford
- Chappell, T. / Crisp, R. (1998): „Utilitarianism“, in: Craig, E. (1998)
- Craig, E. (Hrsg.) (1998): „Routledge Encyclopedia of Philosophy“, London, New York
- Dawkins, R. (1994): „Das egoistische Gen“, 2. Auflage, München, u. a.
- Falk, A. / Fischbacher, U. (2000): „A Theory of Reciprocity“, Working Paper No. 6, überarbeitete Fassung, Institut für Empirische Wirtschaftsforschung, Universität Zürich
- Frankena, W. K. (1963): „Ethics“, Englewood Cliffs, New Jersey
- Gibbons, R. (1992): „Game Theory for Applied Economists“, Princeton, New Jersey
- Gintis, H. (2000): „Game Theory Evolving“, Princeton University Press
- Güth, W. / Schmittberger, R. / Schwarze, B. (1982): „An Experimental Analysis of Ultimatum Bargaining“, Journal of Economic Behavior & Organization, Vol. 3, S. 367-388, zitiert nach: Gintis, H. (2000)
- Hare, R. M. (1981): „Moral Thinking: Its Methods Levels and Point“, Oxford
- Harsanyi, J. C. (1982): „Morality and the theory of rational behaviour“, in: Sen, A. / Williams, B. (1982)
- Hume, D. (1739, 1973): „Ein Traktat über die menschliche Natur“, übersetzt von T. Lipps, herausgegeben von R. Brandt, Hamburg
- Kant, I. (1785, 2004): „Grundlegung zur Metaphysik der Sitten“, herausgegeben von J. Timmermann, Göttingen

- Lazarus, R. S. (1991): „Emotion and Adaptation“, New York
- Lyons, D. (1965): „Forms and Limits of Utilitarianism“, Oxford
- Mas-Colell, A. / Whinston, M. D. / Green, J. R. (1995): „Microeconomic Theory“, New York, Oxford
- Maynard Smith, J. (1982): „Evolution and the Theory of Games“, Cambridge
- Moore, G. E. (1903): „Principia Ethica“, Cambridge
- Mill, J. S. (1861, 1910): „Utilitarianism“, herausgegeben von A. D. Lindsay, London
- Nash, J. (1950): „The Bargaining Problem“, *Econometrica*, Vol. 18
- Nida-Rümelin, J. (1993): „Kritik des Konsequentialismus“, München
- Rawls, J. (1999): „A Theory of Justice“, überarbeitete Fassung, Cambridge, Massachusetts
- Rousseau, J.-J. (1755): „Discours sur l’origine et les fondements de l’inégalité parmi les hommes“
- Sen, A. / Williams, B. (Hrsg.) (1982): „Utilitarianism and Beyond“, Cambridge
- Scarre, G. (1996): „Utilitarianism“, London
- Smart, J. J. C. (1973): „An outline of a system of utilitarian ethics“,
in: Smart/Williams (1973)
- Smart, J. J. C. / Williams, B. (1973): „Utilitarianism – for and against“, Cambridge
- Sober, E. (2000): „Philosophy of Biology“, 2. Auflage, Boulder, Colorado
- Verbeek, B. / Morris, C. (2004): „Game Theory and Ethics“, *The Stanford Encyclopedia of Philosophy* (Winter 2004 Edition), Edward N. Salta (Hrsg.).
- Vollmer, G. (1993): „Sein und Sollen: Möglichkeiten und Grenzen einer Evolutionären Ethik“, in: Vollmer, G. (1995)
- Vollmer, G. (1995): „Biophilosophie“, Stuttgart
- Von Kutschera, F. (1982): „Grundlagen der Ethik“, Berlin
- Von Neumann, J. / Morgenstern, O. (1944): „Theory of Games and Economic Behavior“, Princeton
- Weibull, J. (1995): „Evolutionary Game Theory“, MIT Press
- Williams, B. (1973): „A critique of utilitarianism“, in: Smart/Williams (1973)